



OPEN

## Sequence characterisation and novel insights into bovine mastitis-associated *Streptococcus uberis* in dairy herds

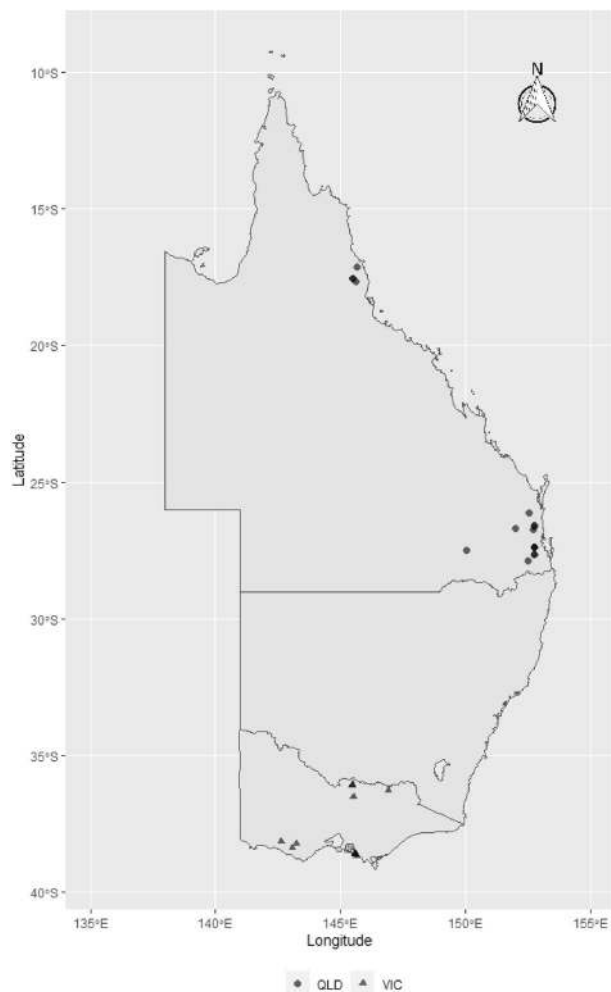
Ben Vezina<sup>1,3,8</sup>, Hulayyil Al-harbi<sup>2</sup>, Hena R. Ramay<sup>4</sup>, Martin Soust<sup>5</sup>, Robert J. Moore<sup>6</sup>, Timothy W. J. Olchowy<sup>1,7</sup> & John I. Alawneh<sup>1,2,8</sup>✉

*Streptococcus uberis* is one of the most frequent mastitis-causing pathogens isolated from dairy cows. Further understanding of *S. uberis* genetics may help elucidate the disease pathogenesis. We compared the genomes of *S. uberis* isolates cultured from dairy cows located in distinctly different geographic regions of Australia. All isolates had novel multi locus sequence types (MLST) indicating a highly diverse population of *S. uberis*. Global clonal complexes (GCC) were more conserved. GCC ST86 and GCC ST143 represented 30% of the total isolates ( $n = 27$ ) and were clustered within different geographic regions. Core genome phylogeny revealed low phylogenetic clustering by region, isolation source, and MLST. Identification of putative sortase (*srtA*) substrates and generation of a custom putative virulence factor database revealed genes which may explain the affinity of *S. uberis* for mammary tissue, evasion of antimicrobial efforts and disease pathogenesis. Of 27 isolates, four contained antibiotic resistance genes including an antimicrobial resistance cluster containing *mel/mef(A)*, *mrsE*, *vatD*, *lnuD*, and transposon-mediated *lnuC* was also identified. These are novel genes for *S. uberis*, which suggests interspecies lateral gene transfer. The presence of resistance genes across the two geographic regions tested within one country supports the need for a careful, tailored, implementation and monitoring of antimicrobial stewardship.

Mastitis poses a risk to public health, animal welfare, and farm profitability worldwide<sup>1–5</sup>. Increased costs related to clinical mastitis are the result of reduced milk production and price, discarded milk, animal culling, mortality, labour and herd veterinary costs<sup>5</sup>. Both contagious and environmental pathogens have been implicated in intramammary infections and bovine mastitis. *Streptococcus uberis* is one of the most frequently identified environmental pathogens responsible for mastitis in dairy herds, a trend which appears to be increasing worldwide<sup>4–6</sup>. Mastitis control programs have been effective in decreasing the prevalence of contagious pathogens in well managed dairy herds<sup>7</sup>, but *S. uberis* remains responsible for a significant proportion of both clinical and subclinical intramammary infections<sup>8</sup>. In Australia, *S. uberis* associated mastitis accounts for one of every three cases of clinical mastitis<sup>9</sup>. This pathogen can be found in a variety of sites in the dairy cow environment such as bedding materials and milking equipment, as well as on other items within the dairy cows' daily living spaces<sup>10</sup>. However, *S. uberis* is not only an environmental risk, this pathogen has also been associated with cow-to-cow transmission consistent with a contagious pathogen role for *S. uberis*<sup>10–12</sup> and indicative of the complexity of the bacteria, environment and dairy cow relationship.

A number of different DNA-based techniques have been used to investigate the genetic diversity of *S. uberis* isolated from cases of bovine mastitis, including MLST<sup>11, 13–16</sup>. While MLST has provided insights into the population diversity of *S. uberis*<sup>11, 16, 17</sup>, it is limited in its ability to delve deeply into genetic variation of isolates. The recent availability of low-cost Whole Genome Sequencing (WGS) technology and advances in computational

<sup>1</sup>Good Clinical Practice Research Group (GCPRG), The University of Queensland, School of Veterinary Science, Gatton, QLD 4343, Australia. <sup>2</sup>The University of Queensland, School of Veterinary Science, Gatton, QLD 4343, Australia. <sup>3</sup>Centre for Cell Factories and Biopolymers, Griffith Institute for Drug Discovery, Griffith University, Nathan, Australia. <sup>4</sup>International Microbiome Centre, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada. <sup>5</sup>Terragen Biotech Pty Ltd., Coolumb Beach, QLD 4573, Australia. <sup>6</sup>School of Science, RMIT University, Bundoora, Melbourne 3083, Australia. <sup>7</sup>Faculty of Veterinary Medicine, University of Calgary, Calgary, AB T3R 1J3, Canada. <sup>8</sup>These authors contributed equally: Ben Vezina and John I. Alawneh. ✉email: j.alawneh@uq.edu.au



**Figure 1.** Map showing the geographical distribution of *S. uberis* isolates sourced from farms in climatically distinct regions across Australia: Queensland (QLD) farms (North QLD (circles), South QLD (circles)), and Victorian farms (triangles). The symbols represent the townships centroids where the farms are located. The map was generated using ozmap package version 0.3.6 [<https://CRAN.R-project.org/package=ozmaps>]<sup>99</sup> implemented within R version 4.0.2 [<https://www.R-project.org/>]<sup>100</sup>.

biology has made it possible to compare the entirety of the *S. uberis* genome and completely characterise the levels of heterogeneity between strains<sup>18</sup>.

WGS analysis of the genomic sequence of *S. uberis* strains has revealed both a capacity for nutritional flexibility and a large variety of metabolic capabilities that promote bacterial survival in different environments<sup>18</sup>, potentially responsible for the high degree of genetic diversity of *S. uberis* within the dairy herd environments<sup>16, 19, 20</sup>. Variations in genetic profile among *S. uberis* strains were not only observed in isolates originating from different dairy herds, but also in the pathogens present within the same herd<sup>7, 21</sup>. This inherently high genetic diversity renders the attribution of a specific set of virulence factors (and associated pathogenesis) to a given *S. uberis* strain/s challenging<sup>12, 22, 23</sup>. The development of strategies to control this important mastitis-causing organism necessitates monitoring of the genetic diversity and determination of the significance of any detected diversity to the disease pathogenesis<sup>22, 24</sup>. Therefore, the main objectives of this study were to compare *S. uberis* isolates cultured from dairy cows with clinical mastitis in distinctly different geographic and climatic regions of Australia. We then identified antibiotic resistance genes, virulence factors, and mobile genetic elements which could then inform antimicrobial stewardship programs. A secondary objective was to examine the diversity of the isolates, in particular their specific antimicrobial resistance profiles from these different regions to identify geographic or climatic associations that may exist as a potential basis for development of more effective and sustainable mastitis treatments for dairy cows.

## Results

***S. uberis* mastitis isolates, sequence types and phylogenetic relationships.** Eleven isolates from 10 locations in Victoria (VIC) and 16 isolates from 8 sites in Queensland (QLD) for a total of 27 isolates were collected and analysed in this study (Fig. 1). The 27 *S. uberis* genomes ranged in size from 1,832,555 to 2,019,389 bp

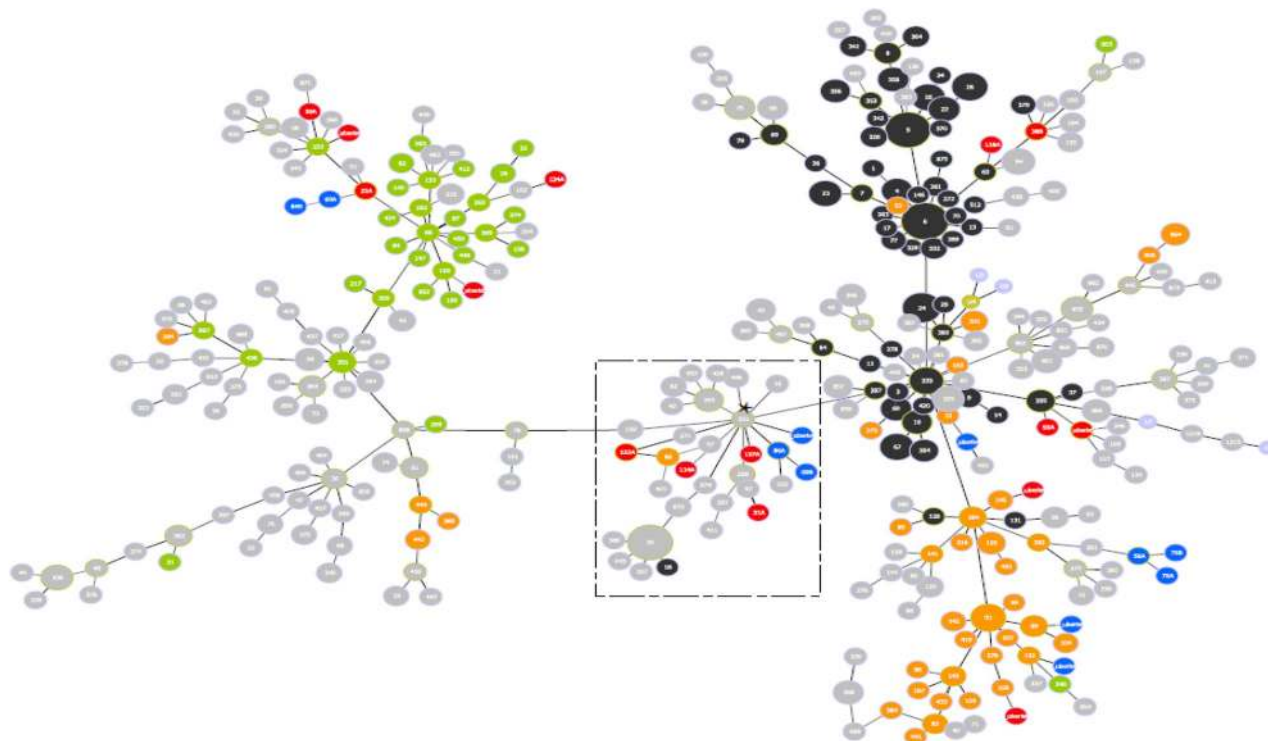
Isolate	Genome size (bp)	Source	State	MLST ST	GCC ST
<i>S. uberis</i> 22A	1,938,913	QLD1	Queensland	~222	86
<i>S. uberis</i> 31A	1,900,760	QLD2	Queensland	~97	–
<i>S. uberis</i> 36A	1,914,697	QLD3	Queensland	~222	86
<i>S. uberis</i> 38B	1,958,229	QLD3	Queensland	~253	5
<i>S. uberis</i> 42	1,883,527	VIC1	Victoria	~271	143
<i>S. uberis</i> 43	1,925,541	VIC2	Victoria	~402	5
<i>S. uberis</i> 44	1,879,665	VIC3	Victoria	~89	143
<i>S. uberis</i> 45	1,961,274	VIC4	Victoria	~190	–
<i>S. uberis</i> 46	1,899,378	VIC5	Victoria	~132	143
<i>S. uberis</i> 47	1,971,757	QLD4	Queensland	~27	143
<i>S. uberis</i> 48	1,927,709	QLD5	Queensland	~222	86
<i>S. uberis</i> 49	2,019,389	QLD6	Queensland	~145	143
<i>S. uberis</i> 50	1,938,127	QLD7	Queensland	~425	86
<i>S. uberis</i> 51	1,984,097	QLD8	Queensland	~56	86
<i>S. uberis</i> 56A	1,906,743	VIC6	Victoria	~261	143
<i>S. uberis</i> 60A	1,967,409	VIC6	Victoria	~412	86
<i>S. uberis</i> 69B	1,926,363	VIC7	Victoria	~271	–
<i>S. uberis</i> 70A	1,919,457	VIC8	Victoria	~256	143
<i>S. uberis</i> 75B	1,844,929	VIC9	Victoria	~156	143
<i>S. uberis</i> 86A	1,977,432	VIC10	Victoria	~90	–
<i>S. uberis</i> 93A	1,856,167	QLD9	Queensland	~67	5
<i>S. uberis</i> 106A	1,832,555	QLD6	Queensland	~1223	86
<i>S. uberis</i> 107A	1,832,592	QLD6	Queensland	~1223	–
<i>S. uberis</i> 114A	1,927,632	QLD7	Queensland	~1223	–
<i>S. uberis</i> 118A	1,894,094	QLD8	Queensland	~850	5
<i>S. uberis</i> 122A	1,886,974	QLD8	Queensland	~80	–
<i>S. uberis</i> 124A	1,909,330	QLD8	Queensland	~1223	86

**Table 1.** Summary of *S. uberis* strains identified in this study, isolated from cows affected by bovine mastitis. MLST STs are also shown. ~ indicates closest ST. All STs were considered novel. The isolation sources have been given codes for privacy reasons.

(Table 1). There were 21 MLST<sup>16</sup> sequence types (STs) represented within this collection. All sequence types (STs) were considered novel as they matched no known sequence types in the *S. uberis* MLST database (pubmlst.org/suberis)<sup>25</sup>. The closest ST matches were ST ~ 1223 (4/27) followed by ST ~ 222 (3/27). These were only present in QLD farms (Table 1). Analysis of GCC STs showed the most abundant STs to be GCC ST86 (8/27) and GCC ST143 (8/27) with GCC ST5 representing 4/27 isolates. All GCC ST86 isolates were from herds with high bulk milk somatic cell counts ( $270\text{--}400 \times 10^3$  cells/mL). GCC ST143 isolates were predominantly ( $p=0.048$ ) found in Victoria (75%) while GCC ST86 isolates were predominantly found in Queensland (88%). Seven isolates (27%) did not belong to a defined GCC. Based on the goeBURST analysis, a possible founder ST for those 7 isolates was ST251. The relationship between the STs of this study and those reported globally in the *S. uberis* MLST database (pubmlst.org/suberis) is depicted in Fig. 2.

**Pangenome analysis.** Pangenome analysis showed a total gene pool of 3764 genes (including RNA encoding), with the core genome making up 1509 genes (Table 2). The soft-core genome, found in  $\geq 26/27$  isolates, was made up of 1542 genes. Of these, 1380 genes (89.49%) could be assigned a Gene Ontology (GO) identity and descriptors (Data S5). The most abundant class of genes, based on predicted localisation of the encoded proteins was as ‘integral member of the membrane’ followed by ‘cytoplasm location’. Most genes appeared to be housekeeping. In addition, 64 encoded proteins predicted to be associated with proteolysis, 5 associated with virulence, 4 with adhesion, 1 with capsule polysaccharide biosynthetic process, and 1 was pillus-associated. Of the 5 virulence-associated genes, 3 encoded proteins that were predicted to be secreted: pauA, LytR, and the YSIRK-type signal peptide-containing protein. Several proteins associated with nutrients commonly found in abundance within the mammary tissue of dairy cows were also identified including fibrinogen binding proteins (WP\_046389274.1, WP\_012658173.1) and lactose catabolism proteins (*lacA*, WP\_000215993.1; *lacB*, WP\_000686149.1).

**Virulence factors.** To address the lack of traditional *S. uberis* virulence factors detected from the Virulence Factor Database (VFDB), a custom *S. uberis* virulence database was constructed, named the *S. uberis* Putative Virulence Database; *su*PVDB (Data S2). This was built based on three main sources. Firstly, identification of translated homologues from the VFDB using relaxed alignment statistics ( $>70\%$  query cover and  $>60\%$  similarity)



**Figure 2.** Multilocus sequence type (MLST)-based minimal spanning tree of 1220 *Streptococcus uberis* isolated from bovine mastitis milk as per NCBI database<sup>103</sup> (as of 8 June 2020) to date. Coloured circles are Australian isolates from bovine origin ( $n = 240$ ). The tree was calculated and generated using the goeBURST version 1.2.1 (<http://www.phyloviz.net/goeburst/>)<sup>113</sup> full MST algorithm in Phyloviz version 2.0 ([www.phyloviz.net](http://www.phyloviz.net/))<sup>112</sup>. Node sizes reflect the number of isolates with specific MLST profile. Numbers within the nodes indicate the ST. Node colours refer to types of Global Clonal Complexes (GCC): GCC Strain Type (ST) 86 (green; left side of the plot), GCC ST143 (orange, right-bottom side of the plot), GCC ST5 (black, right-top side of the plot), isolates from Queensland are in red colour and Victorian isolates are in blue. Dashed rectangle show the location of the seven isolates to their possible ST founder ST251 (asterisk).

Pangenome breakdown	Number of isolates <sup>a</sup>	Number of genes	Percentage of pangenome (%)
Core genes	27	1509	40.09
Soft core genes	26	33	0.88
Shell genes	5–25	534	14.19
Cloud genes	1–4	1688	44.85
Total genes	1–27	3764	100

**Table 2.** Pangenome analysis results. The total number of genes and percent of total pangenome is shown. <sup>a</sup>Number of isolates out of the total ( $n = 27$ ).

(Data S3). Secondly, experimentally-confirmed virulence factors identified from a number of publications<sup>26–41</sup>. Finally, inclusion of genes with virulence-associated GO descriptors (Data S5). Despite HasA<sup>42</sup>, PauA<sup>43</sup>, and PauB<sup>39</sup> being non-essential for bovine mastitis, these three were included in the search for the sake of completeness. MtsB and regulator ScaR were also included as potential virulence factors because they were associated with the MtuA operon. To identify homologous virulence factors in *S. uberis*, the *Streptococcal* VFDB amino acid sequences were screened against the translated pangenome. Any sequences with > 70% query cover and > 60% similarity were considered putative homologues. Twenty-eight previously undescribed putative virulence factors were identified in *S. uberis* based on homology to known virulence factors in other *Streptococcal* species (Data S3). This custom database consisted of 53 total sequences (Data S2), which were screened against each isolate, as well as all 40 *S. uberis* isolates present within GenBank (Data S6).

The GO virulence-associated genes encoded an enolase/phosphopyruvate hydratase (WP\_012658173.1), LytR family regulatory protein (SQG46876.1), YSIRK-type signal peptide-containing protein (WP\_154591130.1), PauA a plasminogen activator A (WP\_046388868.1), and a HU family DNA-binding protein, HlpA (WP\_012658741.1). 90 soft core proteins were predicted to be secreted including PauA, LytR, and the YSIRK-type signal peptide-containing protein were all predicted to be secreted, along with pneumococcal-type histidine triad protein (Pht).

Isolate/resistance gene	Lincosamides	Streptogramins	Macrolides, oxazolidinones, phenicols, pleuromutilins, streptogramins, tetracyclines
<i>S. uberis</i> 45	<i>lnuD</i>		
<i>S. uberis</i> 47	<i>lnuC</i>		
<i>S. uberis</i> 48	<i>lnuD, mel, mrsE,</i>	<i>vatD</i>	<i>mel, mrsE</i>
<i>S. uberis</i> 51	<i>lnuD, mel, mrsE,</i>	<i>vatD</i>	<i>mel, mrsE</i>

**Table 3.** Table showing antibiotic resistance profiles of *S. uberis* isolates in this study. The antibiotic class is shown along the x-axis, and the isolate is shown along the y-axis. The gene responsible for resistance is shown.

The average number of virulence factors per isolate was 27.7 factors ( $\pm$  Standard deviation [ $\sigma$ ] = 2.6) from the current study. When the publicly available *S. uberis* genomes were screened, an average of  $27.86 \pm 2.14 \sigma$  putative virulence factors were identified from isolates associated with bovine mastitis (Data S6). The 3 human mastitis isolates also contained a similar amount,  $27.67 \pm 4.04 \sigma$ . The aquatic *S. uberis* isolate CAIM 1894 (SAMN03093229) contained a total of 3 putative virulence factors.

In the current study, the number of virulence factors per isolate ranged from 35 in *S. uberis* 36A to 24 in *S. uberis* 124A. Twenty-two putative virulence factors were found in every isolate. These factors included *biofilm putative glycosyltransferase*, *biofilm putative glycosyltransferase 2*, *cpsB*, *cpsC*, *cpsD/cps4D*, *cylA*, *fabG 1/cylG*, *fbpS*, *gtaB/hasC*, *hasC* homologue *gpsA*, *lmb*, *mga*, *mtuA*, *oppF*, *pauA*, *putative surface-anchored protein*, *rqcH/fbp54*, *scaR*, *scpA*, *sua*, *tagU 3/cps4A* and *srtA* (Fig. 4, Data S4). Moreover, there was a tendency for virulence factors to cluster by state ( $p = 0.09$ ). No clustering was observed between isolation source and virulence factors. The accessory virulome was made up of putative virulence factors not present in every isolate. They consisted of 24 sequences (Fig. 4), with individual isolates containing the majority of them. *S. uberis* 56A was the only isolate to contain the putative virulence factors *cpsL* and *neuABCD* (*epsM/neuD*, *legI/neuB*, *neuA\_1*, *neuC*).

**Putative sortase substrates, antimicrobial resistance and mobile genetic elements.** *Streptococcal* sortase *srtA* has been shown to be essential for bovine mastitis<sup>26</sup> and a previous study identified 9 proteins attached to the cell wall of *S. uberis* which were likely virulence-associated genes<sup>44</sup>. This enzyme cleaves and processes proteins containing particular motifs to allow external cell wall-anchoring. To confirm the importance of cell wall-anchored proteins, isolates were screened for the *srtA* gene, which had two variants with 85.9% similarity between them. 22/27 strains contained one *srtA* variant (WP\_046390884.1) while 5/27 strains contained another (WP\_037592200.1). This was the result of the 90% identity cut-off criteria used in the pangenome analysis. Within this constraint, sortase *srtA* was still considered a core gene.

To analyse SrtA substrates, core protein-coding sequences were screened for several motifs including LPXTG<sup>44, 45</sup>, LPXXD<sup>44</sup>, LPXTA<sup>46, 47</sup>, QVPTGV and LPSTGE<sup>48</sup>. 12 protein-coding sequences were identified when excluding the more variable LPXXD motif, and 102 including it. As only core genes were screened, there was an increased likelihood of finding essential virulence factors. Some putative sortase substrates of note included a transmembrane protein containing a xanthine permease domain (WP\_046392124.1), 7 coding sequences (CDS) associated with the bacterial global response to DNA damage (SOS), peroxiredoxin *ahpC* (WP\_015912043.1), and transmembrane branched-chain amino acid transport protein (WP\_046393325.1) (Data S5).

Overall, antibiotic resistance genes were not commonly found within the genomes. Only *S. uberis* 45, 47, 48 and 51 contained resistance genes and most of these were predicted to encode resistance to lincosamides, due to presence of *lnuC* or *lnuD* (Table 3). Antibiotic resistance genes were found in both Victorian and Queensland isolates and in a number of STs. Screening NCBI for *LnuC* revealed it was not identified in any *S. uberis* isolates in the database. A multidrug antibiotic resistance cluster was identified in both *S. uberis* 48 and 51. The cluster consisted of *mel/mef(A)*, *mrsE*, *vatD* and *lnuD*. Despite *S. uberis* 47 containing *lnuD*, it did not contain the cluster. The two isolates containing the cluster were found in different isolation sources within Queensland. They both belonged to GCC ST86. The gene cluster is associated with bacterial resistance to lincosamides, macrolides, oxazolidinones, phenicols, pleuromutilins, streptogramins and tetracyclines.

One-third (9/27) of the isolates contained putatively intact bacteriophages (Table 4). *S. uberis* 38B contained two bacteriophages. The *Streptococcus* prophage 315.2 was found in isolates *S. uberis* 38B and *S. uberis* 47, both Queensland isolates. There were no virulence genes associated with any bacteriophages. Transposons and insertion sequences were identified in 21/27 isolates (77.78%) (Table 5). None were associated with virulence factors. No plasmids were found in the isolates.

## Discussion

Twenty-nine herds located in three climatically distinct geographic regions of Australia were screened and from them 27 independent *S. uberis* isolates were obtained and their genetic relatedness was assessed. The diverse range of MLSTs observed in this study is consistent with previous surveys of bovine mastitis-associated *S. uberis* in that there was a large variety of STs identified<sup>49</sup>. In one study, 33 MLST STs were identified in their group of 46 isolates, the most common STs being MLST ST60 and ST155<sup>49</sup>. In the current study, none of the MLST STs in the isolates were identical, although some were similar such as MLST ST ~ 1223. However, GCC STs were more consistently represented (Table 1) with three GCC STs (ST86, ST143, GCC ST5) representing 20 of the 27 isolates. This finding is mostly consistent with the published literature that has identified GCC ST143 and GCC ST5 as the most common GCCs found in *S. uberis* isolates obtained from Australian cases of clinical mastitis<sup>11, 16, 17, 23</sup>.

Isolate	Region length (kb)	Closest bacteriophage hit	Accession of closest hit
<i>S. uberis</i> 38B	36.8	<i>Streptococcus</i> phage SMP	NC_008721
<i>S. uberis</i> 38B	42.2	<i>Streptococcus</i> prophage 315.2	NC_004585
<i>S. uberis</i> 47	43.9	<i>Streptococcus</i> prophage 315.2	NC_004585
<i>S. uberis</i> 49	40.8	<i>Enterococcus</i> phage phiFLAA	NC_013644
<i>S. uberis</i> 50	55.5	<i>Streptococcus</i> phage 5093	NC_012753
<i>S. uberis</i> 60A	59.5	<i>Streptococcus</i> phage PH10	NC_012756
<i>S. uberis</i> 70A	36.4	<i>Streptococcus</i> phage T12	NC_028700
<i>S. uberis</i> 93A	36.7	<i>Streptococcus</i> phage 315.3	NC_004586
<i>S. uberis</i> 118A	42.8	<i>Streptococcus</i> phage P9	NC_009819
<i>S. uberis</i> 122A	41.9	<i>Streptococcus</i> phage T12	NC_028700

**Table 4.** Table of putatively intact bacteriophage presence within the *S. uberis* isolates identified in this study. The closest-related phage are also shown, along with the accession numbers.

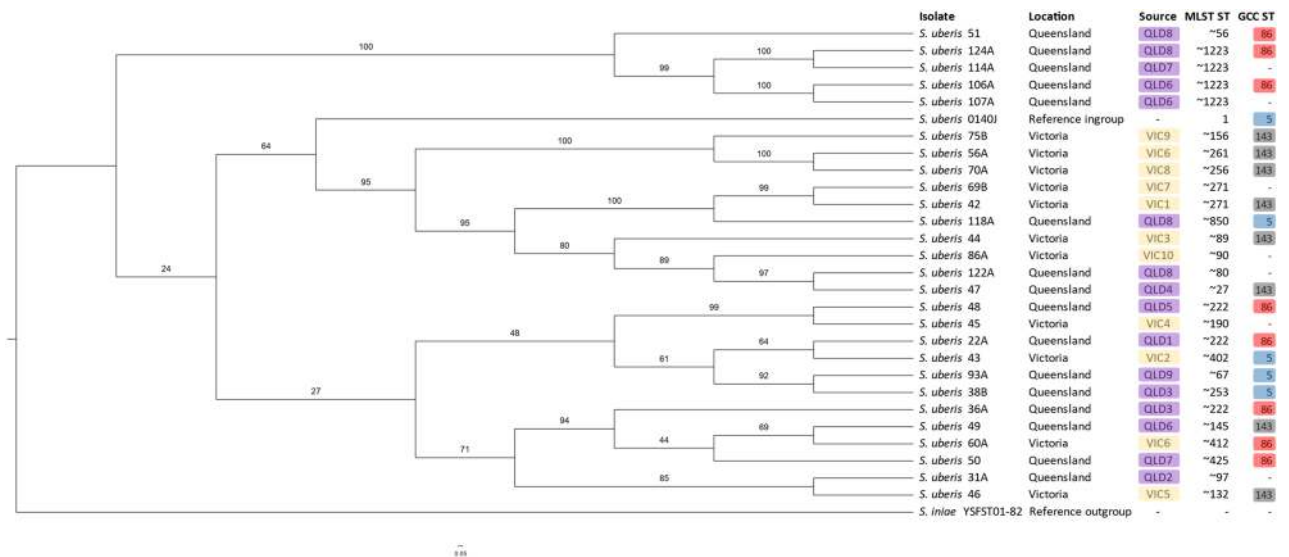
IS family	Isolates	Total
IS256	<i>S. uberis</i> 46	1
IS3	<i>S. uberis</i> 106A; <i>S. uberis</i> 107A; <i>S. uberis</i> 114A; <i>S. uberis</i> 122A; <i>S. uberis</i> 122A; <i>S. uberis</i> 124A; <i>S. uberis</i> 22A; <i>S. uberis</i> 38B; <i>S. uberis</i> 43; <i>S. uberis</i> 44; <i>S. uberis</i> 47; <i>S. uberis</i> 49; <i>S. uberis</i> 51; <i>S. uberis</i> 56A; <i>S. uberis</i> 70A; <i>S. uberis</i> 75B; <i>S. uberis</i> 86A	19
IS30	<i>S. uberis</i> 118A; <i>S. uberis</i> 42; <i>S. uberis</i> 44; <i>S. uberis</i> 56A; <i>S. uberis</i> 69B; <i>S. uberis</i> 70A; <i>S. uberis</i> 75B	7
IS6	<i>S. uberis</i> 31A	1
ISL3	<i>S. uberis</i> 86A	1
Novel	<i>S. uberis</i> 22A	1

**Table 5.** Transposon and insertion sequences within the *S. uberis* isolates identified in this study. Isolates which contain IS families are separated by semicolons. *S. uberis* 122A contains two different IS3 regions and is included twice in this table.

The observation that GCCs clustered by region was notable, given that GCC ST86 isolates were more prevalent in QLD compared to Victoria, while GCC ST143 isolates were more prevalent in Victoria. This study provides observational evidence that *S. uberis* GCC ST86 isolates, which had been previously thought to lack pathogenicity required to cause clinical mastitis<sup>49</sup>, can contribute to disease. Further investigations such as experimental animal trials will be required to confirm this observation. If GCC ST86 isolates were indeed responsible for clinical mastitis, it would also be consistent with the lack of strong phylogenetic clustering of GCC ST86 (Fig. 3); the putative virulence factor profiles which showed consistent differences between GCCs (Fig. 4). Isolation of *S. uberis* GCC ST86 strains from cows afflicted with bovine mastitis may be indicative of a shift in pathogenic lineages within the Australian *S. uberis* population. Alternatively, this provides evidence that sequence typing alone is not a sufficiently granular approach to associate with pathogenicity of *S. uberis* in the bovine mammary gland, neither the core genome-based phylogenetic clustering (Fig. 3) nor the putative virulence factor profiles (Fig. 4) align closely with GCC typing in parallel. Our results indicate that all isolates likely possess virulence factors promoting invasion of host tissue, survival in the host environment, evasion of the host immune response, and internalization in the mammary gland cells<sup>49</sup>.

These observations indicate that the relative importance of an environmental source and possible cow-to-cow transmission in the development and risk of mastitis also depend on geographical factors, and the transmissibility and disease potential of the particular strains circulating in the cow's environment<sup>10, 20, 50</sup>. It could also impact on the relative success rate of treatment protocols<sup>51</sup> and management practices and the apparent 'success' and 'failure' rates in different regions or on different properties. At present, attention to possible environmental sources of infection, good standards of milking hygiene, and proper treatment of *S. uberis* mastitis remain important measures in attempts to control transmission of *S. uberis*<sup>20, 23</sup>.

A maximum likelihood phylogeny of the core genome demonstrated the clustering of MLST STs and indicated that the scheme was, at times, suitable (Fig. 3) for *S. uberis*. For example, MLST ST ~ 1223 clustered well, however MLST ST ~ 222 did not. Given the amount of novel STs identified (100% of isolates), the usefulness of the ST was of limited value to describing the evolutionary lineage of this highly recombinant species<sup>52</sup>. While the GCCs clustered with other bovine mastitis-associated *S. uberis* (Fig. 2), the inconsistent phylogenetic clustering between GCC STs (Fig. 3) meant that these GCCs did not appear to accurately capture the assumed evolutionary lineages of the bovine mastitis-associated *S. uberis* isolates identified in this study. Given the reducing cost of whole genome sequencing and greater power for genomic and phylogenetic analysis compared with the multiple sanger sequencing reactions required for MLST and GCC ST analysis, it would be appropriate to rely more on higher resolution WGS in future diagnostics. There appeared to be two clades containing state-only isolates: a Queensland clade including *S. uberis* 51, 124A, 114A, 106A and 107A, which almost all shared the same MLST



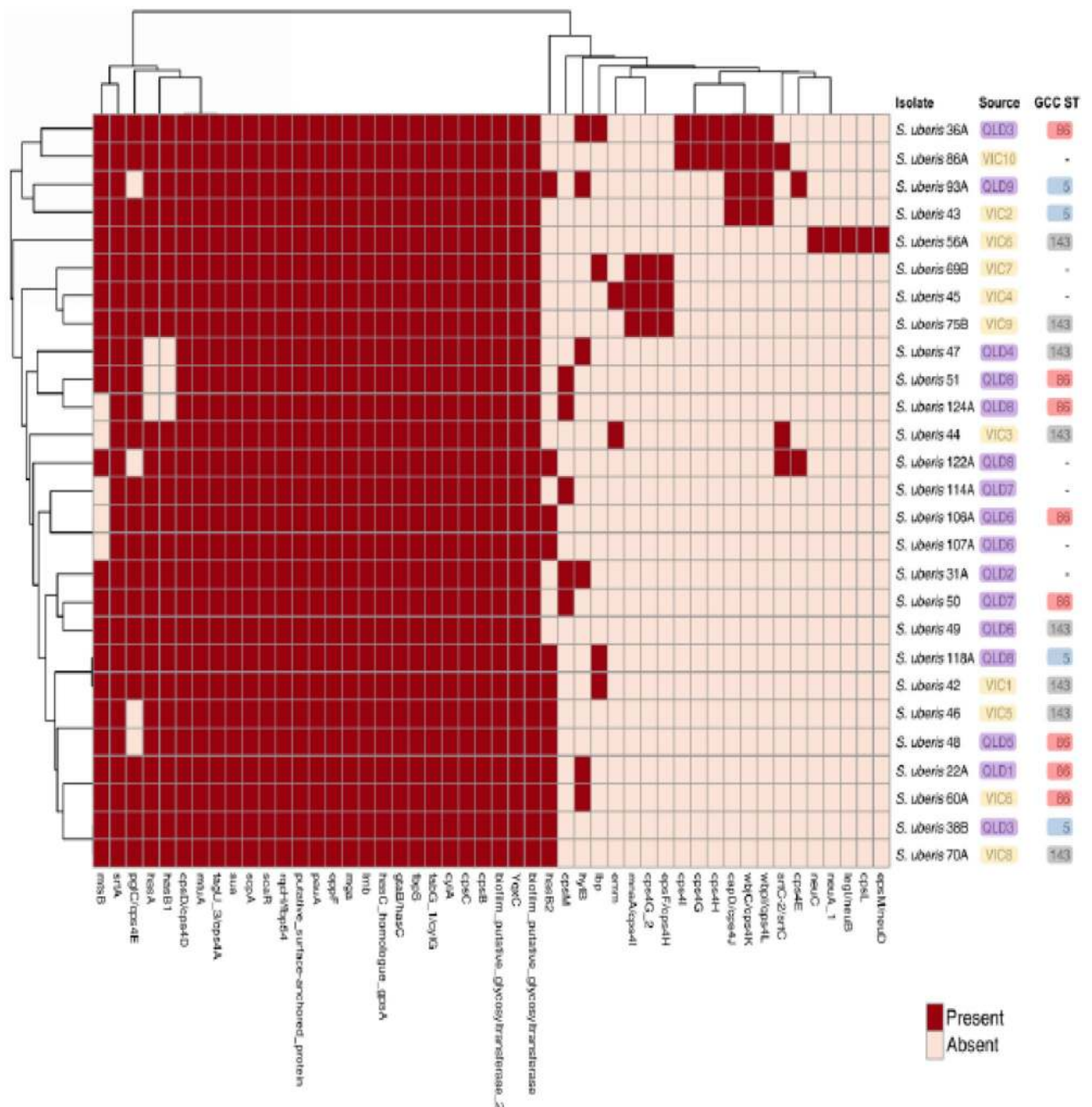
**Figure 3.** Maximum likelihood phylogenetic tree based on core genome alignment between the *S. uberis* isolates identified in this study. *S. iniae* YSFST01-82 (accession number: GCF\_000831485.1) was used as an outgroup and *S. uberis* 0140J (accession number: AM946015) (MLST ST1, GCC ST5) as an ingroup. Bootstrap values are shown in numbers along branches. Purple labels show Queensland isolates. Yellow labels show Victorian isolates. MLST STs are shown, as well as GCC STs. GCCs are coloured, GCC ST86 coloured red, GCC ST 5 coloured blue, GCC ST143 coloured grey. Core genome alignment was performed in roary (version 3.13.0)<sup>115</sup> and maximum likelihood tree was built in RAxML version 8.2.10 (raxmlHPC-PTHREADS-SSE3)<sup>120</sup> with 1000 replicates. Tree visualised in FigTree version 1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>). For figure clarity, the phylogenetic tree was visualised using a cladogram transformation. Data S1 shows the untransformed tree, along with an isolate-only tree which shows very high bootstrap support at almost every node (~100%) due to the increased number of genes in the core genome alignment.

ST (~1223) and GCC ST (86); and a Victorian clade containing *S. uberis* 75B, 56A, and 70A, in which almost all sharing the same GCC ST (143). Within these two specific clades, it is likely that the isolates diverged from a common lineage within their respective states. Every other identified clade appeared to contain a mixture of Victorian and Queensland isolates. Phylogenetic clustering within a source point was not observed in many cases, such as with QLD7, QLD8, QLD9 and VIC6 isolates. This observed divergence was much greater than originally expected. The most likely explanation of this finding is environmental acquisition of novel *S. uberis* bovine mastitis-associated isolates<sup>23,41,50</sup>, rather than an acquisition of monophyletic descendants such as would be expected with a contagious pathogen. Given this observation, further analysis of environmental *S. uberis* reservoirs is recommended.

The pan and core genome genes were consistent with previous pangenome analyses, showing a core genome of 1550 genes using 13 clinical and sub-clinical bovine mastitis isolates<sup>27</sup> and ~1530 genes using 21 bovine mastitis isolates<sup>52</sup>. Core genes associated with nutrients found in abundance within the mammary tissue of cows, including two fibrinogen binding proteins, were identified (WP\_046389274.1, WP\_012658173.1). This is a notable finding given the fibrinogen concentration in the milk of cows affected by mastitis is significantly higher than that of healthy cows<sup>53</sup>. Two genes associated with lactose catabolism were identified, *lacA* (WP\_000215993.1) and *lacB* (WP\_000686149.1), which is also notable given the abundance of lactose in bovine milk. This may explain the presence or adaptation of *S. uberis* to mammary tissue.

As strain specific pathogenicity has been observed with *S. uberis*<sup>54</sup>, this indicates bacterial pathogenic factors are involved, not host-specific factors. Notably, the isolates from this study shared similar numbers of putative virulence factors as the publicly available bovine mastitis isolates, as well as human mastitis isolates (Fig. 4, Data S4, Data S6). This indicates a common set of virulence determinants responsible for cross-host mastitis pathogenesis.

The core virulence factors identified in every isolate (Fig. 4, Data S4) may partially explain the pathogenic behaviour of *S. uberis*. There were several core genes of note including *cpsBCD* and *cps4a* which encode products involved in capsular polysaccharide formation<sup>40</sup>, as-well-as, two putative glycosyltransferases involved with biofilm formation<sup>38</sup>. *cylAG* were identified, which contribute to hemolytic activity<sup>55</sup>, corresponding to an ABC transporter and an enoyl reductase involved in fatty acid biosynthesis<sup>56</sup>. This analysis confirmed *hasA* as a non-essential virulence factor, as it was only found in 24/27 isolates, and also confirmed *mtuA* as an essential virulence factor<sup>33</sup>, as it was found in all isolates. *rqcH/fbp54*, a fibronectin-binding protein and putative fibronectin-binding protein *fbpS* were also found in all isolates. This has been shown to be important for adhesion and a key virulence factor in *Streptococcus pyogenes*<sup>57</sup>, as-well-as, in *Streptococcus suis*, but not essential for virulence<sup>58</sup>. Also found in every isolate was *scpA*, a C5a peptidase, which has been shown to delay accumulation of leukocytes attracted to *S. pyogenes* by cleaving the serum chemotaxin C5a, indicating it is an essential virulence factor<sup>59</sup>. *hasC* was also found in all isolates; it is essential for hyaluronic acid capsule production<sup>29</sup>. Upon loss of this gene, *S. uberis*



**Figure 4.** Heatmap showing the presence/absence of virulence and putative virulence genes (x-axis) within the *S. uberis* isolates identified in this study (y-axis). Presence of virulence genes shown in red, absence in light pink. Core virulence genes are found in every isolate (27/27), while accessory virulence genes are found in < 27 isolates. Victorian isolates are coloured in yellow, Queensland in purple. GCCs are coloured, GCC ST86 coloured red, GCC ST5 coloured blue and GCC ST143 coloured grey. Data generated using ABRicate version 1.0.1 (<https://github.com/tseemann/abricate>)<sup>121</sup> with the VFDB<sup>122</sup> and the custom *Su*PVDB (Data S2). Heatmap produced using ClustVis (<https://bit.cs.ut.ee/clustvis/>)<sup>124</sup> and Inkscape version 0.92 (<https://www.inkscape.org>).

becomes susceptible to phagocytosis by bovine neutrophils, indicating it is an important virulence factor<sup>29</sup>. *lmb*, encoding a laminin binding protein important for bacterial colonisation<sup>37</sup> was also found in every isolate. *Mga*, a virulence regulator has been shown to be important for a number of phenotypes including biofilm formation, growth in whole human blood and soft tissue and phagocytosis resistance<sup>60,61</sup>. This was also identified in all isolates of *S. uberis* in this study. *sua*, an adhesin, has been experimentally verified in *S. uberis* isolates from dairy cows and identified as a conserved gene<sup>62</sup>. Our study confirms this as it is conserved in all isolates. *oppF*, an oligopeptide permease, shown to be important for auxotrophic amino acid acquisition from bovine milk<sup>34</sup>, was also found in every isolate. *scaR*, another virulence regulator which upregulates during low levels of Mn<sup>2+</sup><sup>63</sup> was also found in every isolate. Given that manganese is present in cow's milk at 0.02–0.05 µg/mL<sup>64</sup>, it is likely *scaR* upregulates virulence-associated genes in response to mastitis infection. Finally, a putative surface anchored



protein was also found in all isolates. This was experimentally identified as a sortase substrate, anchored to the surface in *S. uberis*<sup>44</sup>. It contains a G5 domain, which is involved with N-acetylglucosamine binding and is found in a number of different enzymes, ranging from biofilm proteins to IgA-cleaving peptidases<sup>65</sup>. The *neuABCD* locus is involved in sialic acid synthesis and their presence in *S. uberis* 56A indicates it may produce a capsule to assist with immune system evasion<sup>66,67</sup>.

Other core virulence genes included enolase/phosphopyruvate hydratase (WP\_012658173.1), involved in plasminogen and adhesion in *Streptococcal* species<sup>68</sup> and LytR family regulatory protein (SQG46876.1) essential for attaching capsules and teichoic acids to cell wall<sup>69</sup>. Also a YSIRK-type signal peptide-containing protein (WP\_154591130.1), *PauA* a plasminogen activator A (WP\_046388868.1) which has been already described as a non-essential virulence factor in *S. uberis*-related bovine mastitis<sup>43</sup>. *Pht*, a putatively secreted gene has been implicated in adhesion to epithelium in *Streptococcus pneumoniae*<sup>70</sup>. Finally, a HU family DNA-binding protein, *HlpA* (WP\_012658741.1) has been shown to bind to double-stranded DNA involved in repair and recombination<sup>71</sup> in times of cellular stress<sup>72</sup> and is involved in tissue inflammation of *Streptococcal* species<sup>73</sup>.

Scanning the core genome for putative sortase substrates revealed several genes which may also explain behaviour of *S. uberis*, including a transmembrane branched-chain amino acid transport protein (WP\_046393325.1), which has been shown to transport leucine, isoleucine and valine via proton motive force<sup>74</sup>, all three of which are found in excess within bovine milk<sup>75</sup>. Furthermore, bacterial loads of *Staphylococcus aureus*, another bovine mastitis-causing organism have been correlated with these amino acids in milk<sup>76</sup>.

Another putative *srtA* substrate of note is a transmembrane protein containing a xanthine permease domain (WP\_046392124.1), which allows uptake of xanthine. Xanthine and hypoxanthine are commonly found in bovine milk<sup>77</sup>, along with bovine enzyme xanthine oxidoreductase. Bovine xanthine oxidoreductase catalyses oxidation of hypoxanthine to hydrogen peroxide<sup>78</sup> and xanthine, and of xanthine to uric acid, among many other substrates<sup>79</sup>. Of note, is the demonstrated antimicrobial activity against Gram-positive and negative bacteria in milk via production of reactive oxygen/nitrogen species, namely hydrogen peroxide and nitric oxide in both humans and cows<sup>78,80</sup>. All *S. uberis* isolates identified in this study containing an externally-located protein which competes with this precursor antimicrobial compound may indicate a novel virulence factor which allows *S. uberis* to colonise and infect the udder, causing bovine mastitis. The core genome contained 7 genes associated with the bacterial SOS response (Data S5), which may be involved in combating damage caused to DNA by reactive oxygen/nitrogen species<sup>81</sup>. Also found in the core genome was peroxiredoxin *ahpC* (WP\_015912043.1), which breaks down hydrogen peroxide and has been shown in *Streptococcus agalactiae* to bind heme<sup>82</sup>.

We propose the xanthine permease domain protein uptakes free hypoxanthine and xanthine found in bovine milk. This likely competes with bovine enzyme xanthine oxidoreductase which usually converts xanthine to the antimicrobial, reactive oxygen species, hydrogen peroxide. At least 7 genes associated with the bacterial SOS response are likely involved in combating damage caused to DNA by reactive oxygen/nitrogen species. The presence of core genes such as these demonstrate adaptation of *S. uberis* to the bovine host, and as one facet of *S. uberis*-associated bovine mastitis. It is possible the genes highlighted here demonstrate the Red Queen hypothesis<sup>83</sup> between host and pathogen<sup>84,85</sup>, which can be summarised as follows: organisms living within a dynamic environment (bovine host) require matching rates of evolution to maintain colonisation. Co-evolution between host and parasite is a widely recognised consequence of natural selection<sup>86</sup> and continually develops as genetic arms-race.

There were a limited number of antibiotic resistance genes found in this study. This finding is consistent with some previous dairy herd surveys<sup>9</sup>, but inconsistent with others that detected a broad and diverse group of antibiotic resistance genes<sup>13,87,88</sup>. Four *S. uberis* isolates found across Victoria and Queensland herds contained a combination of *InuC* and a multidrug resistance cluster including *mel/mef(A)*, *mrsE*, *vatD* and *InuD*. The presence of an antimicrobial resistance gene cluster not associated with any known insertion sequences or mobilizable elements indicates this may have been acquired through uptake and integration of free DNA. The two isolates that contain this cluster, *S. uberis* 48 and 51 are not directly phylogenetically related in the context of other isolates, which may indicate this chromosomal integration occurred independently. Further evidence of this can be seen in the variable genomic context of this cluster, which is not consistently placed within the genome (data not shown). The *mel/mef(A)* gene has been reported in *S. pneumoniae*<sup>89</sup> and *S. pyogenes*<sup>90</sup> but not in *S. uberis*. *S. uberis* was then screened via NCBI for *mel/mef(A)*. Only one protein sequence was identified (accession: WP\_046391446.1) from *S. uberis* 6780 (accession: JATD00000000), also isolated from milk from a cow udder afflicted with bovine mastitis<sup>27</sup>. This provided further evidence that this observation is a rare occurrence not associated with known mobilizable elements. The *mel/mef(A)* gene encodes resistance to lincosamides, macrolides, oxazolidinones, phenicols, pleuromutilins, streptogramins and tetracyclines<sup>91</sup>. The *msr* family ABC-F type ribosomal protection gene, which is homologous to the coding sequence of MrsE (99% query coverage, 79% similarity), encodes resistance to streptogramins, phenicols, pleuromutilins, lincosamides, oxazolidinones, tetracyclines, macrolides and other antibiotics which target the peptidyl-transferase region of the ribosomal subunit<sup>92</sup>. The next gene in the cluster encodes a hypothetical protein with no known function or domains. This is followed by *vatD*, a streptogramin A O-acetyltransferase resistance gene<sup>93</sup>. The final gene in the cluster is *InuD*, encoding a nucleotidyltransferase, which has been shown to provide lincosamides resistance. The *InuD* and non-cluster *InuC* genes encode O-nucleotidyltransferases and inactivate lincosamides by adenylation<sup>94</sup>. The *InuD* has been previously reported in *S. uberis* isolated from clinical bovine mastitis<sup>95</sup>, but the *InuC* gene has only been identified in human *S. agalactiae* isolates<sup>96,97</sup>, suggesting recent acquisition by bovine mastitis-associated *S. uberis*. As *InuC* is a transposon-mediated resistance gene<sup>97</sup>, this may represent the first observed interspecies jump into bovine mastitis-associated *S. uberis*. Lincosamide (lincomycin, clindamycin, and pirlimycin) usage and resistance should be carefully monitored now this mobilizable gene is within bovine mastitis-associated *S. uberis* populations. It would be of value to determine if a similar pattern exists in other streptococcal species present in the dairy farm environment. Further work is needed to verify these findings and determine antimicrobial gene

reservoirs and the mechanism of transfer. A limitation of this study was the small number of herds and limited number of isolates cultured from clinical mastitis cases. However, herd (farm) type, breed of dairy cow, and herd size were representative of Australian herds. Nevertheless, the presence of these genes in some of the isolates indicates that antibiotic usage should be carefully tracked and monitored to produce and maintain an up-to-date understanding of the level of antimicrobial resistance within dairy populations in the country.

## Materials and methods

The study and all experimental procedures and protocols were approved by the Animal Ethics Committee at the University of Queensland (animal ethics approval numbers: SVS/ANRFA/540/18 and SVS/043/18/TERRA-GEN), and all methods were performed in accordance with the relevant institutional guidelines and regulations. The methods described in the current study and reported results were compliant with ARRIVE guidelines for reporting animal research<sup>98</sup>.

**Sample collection.** This study was part of a larger project conducted between March and June 2019 based on 430 milk samples collected from 29 dairy herds located in Queensland [QLD] (North Queensland,  $n=9$  herds; Southeast QLD,  $n=9$  herds), and Victoria ([VIC],  $n=11$  herds) (Fig. 1; map figure was generated using `ozmap` package version 0.3.6 [<https://CRAN.R-project.org/package=ozmaps>]<sup>99</sup> implemented within R version 4.0.2 [<https://www.R-project.org/>]<sup>100</sup>). A 2-week rolling average of herd bulk somatic cell counts (BMTSCCs) was used to arbitrarily divide the herds into three categories ( $\leq 150 \times 1000$  cells/mL,  $> 150$  to  $300 \times 1000$  cells/mL, and  $> 300 \times 1000$  cells/mL). BMTSCCs categories were used as an a priori assumption for the risk of mastitis in the herd. Herd selection was based on ease of access to the farm location and the cooperation of the dairy farm owners and their associated veterinary practices.

Detailed methodology is described elsewhere<sup>101</sup>. Briefly, milk samples were collected from eligible dairy cows with a new case of clinical mastitis. Chronic mastitis cases (apparently healthy cow with lumps palpable in the udder, and mild changes to milk) and subclinical mastitis cases were not eligible for study enrolment. An enrolment eligible clinical mastitis case was defined as a previously apparently healthy lactating dairy cow of any age, breed, or stage of lactation that was experiencing a new case of clinical mastitis, defined as either the first occurrence of a mastitis event in the current lactation or a mastitis event occurring at least 21 days following a previous mastitis event that has clinically resolved or achieved a clinical cure<sup>102</sup>. Eligible cases must have not received systemic or intramammary antimicrobials, anti-inflammatory medications, or topical treatments in the 2 weeks prior to becoming a new case, nor immediately prior to sample collection. Milk samples were collected aseptically from individual quarters into separate sterile tubes, immediately capped and placed in a  $-20$  °C freezer. Collected samples were transported and delivered frozen to the Veterinary Laboratory Services of the University of Queensland for bacterial culture.

***Streptococcus uberis* isolation.** At the laboratory, milk samples were mixed thoroughly, 100  $\mu$ L streaked onto Sheep Blood Agar (SBA, P2133 Sheep Blood Columbia Agar Plates, Thermofisher), and the SBA plate incubated aerobically at 37 °C for 18–24 h. *S. uberis* isolates were initially identified using conventional microbiology laboratory tests (Gram stain appearance and catalase production). Matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS; Bruker Daltonik, Bremen, Germany) was used to confirm the identification. Individual colonies were sub-cultured on SBA plates and incubated aerobically at 37 °C for 18–24 h. Pure isolates were then incubated in 2 mL of Brain Heart Infusion (BHI) broth, subsequently mixed with 20% glycerol, and stored at  $-80$  °C.

**DNA extraction.** Stored *S. uberis* isolates were batch thawed and cultured in BHI broth (37 °C, orbital shaker at 300 rpm, 18–24 h). Genomic DNA was extracted using DNeasy PowerFood Microbial Kit (QIAGEN Chadstone, Victoria, Australia) with minor modifications. Eight mL of culture liquid was centrifuged (15 min, 4 °C,  $20,000 \times g$ ) to pellet the bacteria. The pellet was then resuspended in 450  $\mu$ L of lysis buffer and incubated for 10 min at 65 °C. Proteinase K (Proteinase K, QIAGEN Chadstone, Victoria, Australia) (25  $\mu$ L) was added and incubation continued for an additional 20 min at 65 °C. This Proteinase K modification to the DNeasy protocol increased the final quantity of extracted DNA (data not shown). Thereafter, the whole component was transferred to a Powerbead tube, secured horizontally to a vortex adapter, and vortexed at maximum speed for 10 min. After washing to remove protein and other inhibitors, purified DNA was eluted and the concentration and purity of the isolated genomic DNA determined using a NanoDrop ND-1000 spectrophotometer. A sample of DNA was considered acceptable if the A260/280 ratio was  $\sim 1.8$ .

**Whole genome sequencing.** The New England Biolabs NEBNext Ultra II FS DNA Library Prep Kit for Illumina for multiplexing was used according to the manufacturer's instructions (New England Biolabs, Notting Hill, Victoria, Australia) to construct whole genome sequencing libraries. Libraries were sequenced (Illumina NextSeq 500 instrument, 300 cycle mid-output kit;  $2 \times 150$  bp paired end) at the Centre of Health Genomics and Informatics, University of Calgary, Canada, to obtain an average coverage depth of  $> 100$  and a read retention of  $> 98\%$ . The draft genome sequences of the 27 *S. uberis* isolates have been deposited with GenBank<sup>103</sup> under BioProject PRJNA669385.

**Whole genome assembly and annotation.** Primer sequences were removed and reads were quality trimmed using cutAdapt<sup>104</sup>. The Nullarbor pipeline<sup>105</sup> was used to process the samples. SKESA assembler (version 2.3.0)<sup>106</sup> was used for de-novo assembly. Assemblies were annotated using prokka (version 1.14.5)<sup>107</sup>.

Details about the parameters used in different packages can be found at [https://github.com/bananabencana/genomics\\_methods\\_S.uberis/blob/main/commands\\_run](https://github.com/bananabencana/genomics_methods_S.uberis/blob/main/commands_run).

**Multi-locus sequence typing.** Assemblies were taxonomically assigned, analysed for MLST and were possible assigned a GCC, sequence types (STs)<sup>108, 109</sup>, and scored for genome distance using Type (Strain) Genome Server (TYGS) (<https://tygs.dsmz.de>)<sup>110</sup> (date accessed: 19 May 2020). MLSTs were confirmed using `mlst_check` version 2.1.1706216<sup>111</sup> and MLST databases accessed on 5 June 2020. PhyloViz version 2.0 ([www.phyloViz.net](http://www.phyloViz.net))<sup>112</sup> and the `goeBURT` algorithm<sup>113</sup> were used to visualise the data. Pairwise comparison of genome sequences among the set of genomes were conducted by calculating precise distances using the Genome BLAST Distance Phylogeny approach (GBDP) under the algorithm ‘coverage’ and distance formula  $d5$ <sup>114</sup>. These distances were used to determine the genome similarities for each of the genome pairs. 100 distance replicates were calculated each. Digital DDH values and confidence intervals were calculated using the recommended settings of the GGDC 2.1<sup>114</sup>.

**Pangenome analysis.** Pangenome analysis was performed using `roary` (version 3.13.0)<sup>115</sup>, splitting paralogues and setting the identity threshold at 90%. This list was then annotated and characterised via screening sequences against NCBI and Gene Ontology<sup>116</sup> using the PANNZER2 server<sup>117</sup>. Signal sequences were obtained from the core genome using SignalP 5.0 (Linux x86\_64)<sup>118, 119</sup>.

**Phylogenetic analysis.** A core genome alignment generated by `roary` was performed, including an out-group *S. iniae* YSFST01-82 (accession number: GCF\_000831485.1) and in-group *S. uberis* 0140J (accession number: AM946015) ST1. RAxML (`raxmlHPC-PTHREADS-SSE3` version 8.2.10)<sup>120</sup> was used for maximum likelihood and rapid bootstrap analysis with 1000 replicates. The bipartitions output file was visualised in FigTree version 1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) and bootstraps visualised. Inkscape version 0.92 (<http://www.inkscape.org>) was used to prepare the figure.

**Identification of putative virulence factors.** ABRicate version 1.0.1 (<https://github.com/tseemann/abricate>)<sup>121</sup> was used with the VFDB<sup>122</sup> and a custom gene list, the *Su*PVDB (Data S2). The custom gene list was based on sequences identified in the literature described to be *S. uberis* virulence factors with experimental or strong computational evidence<sup>26–41</sup>, as well identifying homologues of *Streptococcal* virulence factors. In addition to the isolates genomes reported in this current study, 40 additional publically available genomes (Data S6)<sup>27, 123</sup> were retrieved from GenBank and screened for putative virulence factors. To identify homologous virulence factors in *S. uberis*, the *Streptococcal* VFDB amino acid sequences were screened against the translated pangenome. Any sequences with >70% query cover and >60% similarity were considered putative homologues (Data S3). The output was then visualised in ClustVis (<https://biit.cs.ut.ee/clustvis/>)<sup>123</sup>. No scaling was performed on the data and clustering was performed on rows and columns using the ‘Euclidean’ measure. To identify core, putative sortase A substrates, `grep` version 3.1 (<https://git.savannah.gnu.org/cgi/grep.git/tree/AUTHORS>) was used.

**Antibiotic resistance determination.** ABRicate version 1.0.1 (<https://github.com/tseemann/abricate>)<sup>121</sup> was used with the CARD<sup>125</sup>, database accessed 4/5/2020.

**Bacteriocin identification.** Bacteriocins were identified by downloading the BAGEL4 database<sup>126</sup> and converting into a multifasta file. The pangenome was then aligned against this using `blastp` version 2.9.0+<sup>127, 128</sup> and results were filtered for  $\geq 80\%$  coverage and  $\geq 50\%$  similarity/positives (to account for different signal sequences).

**Plasmid identification and analysis.** Mob-suite version 3.0.0<sup>129</sup> was used to identify plasmids using the `mob recon` function, then visualised in ClustVis (<https://biit.cs.ut.ee/clustvis/>)<sup>124</sup>. No scaling was performed on the data and clustering was performed on rows and columns using the ‘correlation’ distance measure. Putative mobilization of plasmids was determined using the `mob_typer` function.

**Bacteriophage and prophages identification.** The PHASTER (<https://phaster.ca/>)<sup>130</sup> web server was used to identify bacteriophage and prophage regions using the API tool. ‘Questionable’ phages (score of 70–90) were grouped with ‘incomplete’.

**Transposon/Insertion sequence identification.** ISEScan version 1.7.2 (<https://github.com/xiezhq/ISEScan>)<sup>131</sup> was used to identify transposons and insertion sequences.

**Statistical analysis.** Fisher’s exact test was performed to identify significance in the proportion of strain types, virulence genes, and plasmids from isolates (unit of analysis) recovered from NQLD, QLD and VICT. A two sided *p*-value obtained by Monte-Carlo simulation ( $n = 2000$ ) of at least 0.05 was considered to be significant. Statistical analysis was conducted using `stats` package implement in R.

Received: 21 October 2020; Accepted: 20 January 2021

Published online: 04 February 2021

## References

- Rollin, E., Dhuyvetter, K. C. & Overton, M. W. The cost of clinical mastitis in the first 30 days of lactation: An economic modeling tool. *Prev. Vet. Med.* **122**, 257–264. <https://doi.org/10.1016/j.prevetmed.2015.11.006> (2015).
- Huijps, K., Lam, T. J. G. M. & Hogeveen, H. Costs of mastitis: Facts and perception. *J. Dairy Res.* **75**, 113–120. <https://doi.org/10.1017/S0022029907002932> (2008).
- Hogeveen, H., Huijps, K. & Lam, T. J. G. M. Economic aspects of mastitis: New developments. *N. Z. Vet. J.* **59**, 16–23 (2011).
- Shum, L. W., McConnel, C. S., Gunn, A. A. & House, J. K. Environmental mastitis in intensive high-producing dairy herds in New South Wales. *Aust. Vet. J.* **87**, 469–475. <https://doi.org/10.1111/j.1751-0813.2009.00523.x> (2009).
- Ruegg, P. L. A 100-Year Review: Mastitis detection, management, and prevention. *J. Dairy Sci.* **100**, 10381–10397. <https://doi.org/10.3168/jds.2017-13023> (2017).
- Verbeke, J., Piepers, S., Supre, K. & De Vliegher, S. Pathogen-specific incidence rate of clinical mastitis in Flemish dairy herds, severity, and association with herd hygiene. *J. Dairy Sci.* **97**, 6926–6934. <https://doi.org/10.3168/jds.2014-8173> (2014).
- Phuektes, P. *et al.* Molecular epidemiology of *Streptococcus uberis* isolates from dairy cows with mastitis. *J. Clin. Microbiol.* **39**, 1460–1466 (2001).
- Bradley, A., Leach, K., Breen, J., Green, L. & Green, M. Survey of the incidence and aetiology of mastitis on dairy farms in England and Wales. *Vet. Rec.* **160**, 253–258 (2007).
- Charman, N., Dyson, R., Hodge, A., Robertson, N. & Chaplin, S. *A Survey of Mastitis Pathogens in the South Eastern Australian Dairy Industry* (Dairy Focus, Moncton, 2012).
- Zadoks, R. N. *et al.* Clinical, epidemiological and molecular characteristics of *Streptococcus uberis* infections in dairy herds. *Epidemiol. Infect.* **130**, 335–349. <https://doi.org/10.1017/S0950268802008221> (2003).
- Pullinger, G. D., Coffey, T. J., Maiden, M. C. & Leigh, J. A. Multilocus-sequence typing analysis reveals similar populations of *Streptococcus uberis* are responsible for bovine intramammary infections of short and long duration. *Vet. Microbiol.* **119**, 194–204. <https://doi.org/10.1016/j.vetmic.2006.08.015> (2007).
- Leelahapongsathon, K., Schukken, Y. H., Srihanasawan, A. & Suriyathaporn, W. Molecular epidemiology of *Streptococcus uberis* intramammary infections: Persistent and transient patterns of infection in a dairy herd. *J. Dairy Sci.* **103**, 3565–3576. <https://doi.org/10.3168/jds.2019-17281> (2020).
- Tomazi, T. *et al.* Genotyping and antimicrobial resistance of *Streptococcus uberis* isolated from bovine clinical mastitis. *PLoS One* **14**, e0223719 (2019).
- Hassan, A., Khan, I., Abdulmawjood, A. & Lämmler, C. Evaluation of PCR methods for rapid identification and Differentiation of *Streptococcus uberis* and *Streptococcus parauberis*. *J. Clin. Microbiol.* **39**, 1618–1621 (2001).
- Lundberg, A., Nyman, A., Unnerstad, H. E. & Waller, K. P. Prevalence of bacterial genotypes and outcome of bovine clinical mastitis due to *Streptococcus dysgalactiae* and *Streptococcus uberis*. *Acta Vet. Scand.* <https://doi.org/10.1186/s13028-014-0080-0> (2014).
- Coffey, T. J. *et al.* First insights into the evolution of *Streptococcus uberis*: A multilocus sequence typing scheme that enables investigation of its population biology. *Appl. Environ. Microbiol.* **72**, 1420–1428. <https://doi.org/10.1128/AEM.72.2.1420-1428.2006> (2006).
- Douglas, V. L., Fenwick, S. G., Pfeiffer, D. U., Williamson, N. B. & Holmes, C. W. Genomic typing of *Streptococcus uberis* isolates from cases of mastitis, in New Zealand dairy cows, using pulsed-field gel electrophoresis. *Vet. Microbiol.* **75**, 27–41. [https://doi.org/10.1016/S0378-1135\(00\)00184-X](https://doi.org/10.1016/S0378-1135(00)00184-X) (2000).
- Ward, P. N. *et al.* Evidence for niche adaptation in the genome of the bovine pathogen *Streptococcus uberis*. *BMC Genom.* **10**, 54 (2009).
- Pullinger, G. D. *et al.* Application of *Streptococcus uberis* multilocus sequence typing: Analysis of the population structure detected among environmental and bovine isolates from New Zealand and the United Kingdom. *Appl. Environ. Microbiol.* **72**, 1429–1436. <https://doi.org/10.1128/AEM.72.2.1429-1436.2006> (2006).
- Lundberg, Å. *et al.* Udder infections with *Staphylococcus aureus*, *Streptococcus dysgalactiae*, and *Streptococcus uberis* at calving in dairy herds with suboptimal udder health. *J. Dairy Sci.* **99**, 2102–2117 (2016).
- Reinoso, E. B., Lasagno, M. C. & Odierno, L. M. Genetic patterns of *Streptococcus uberis* isolated from bovine mastitis. *Rev. Argent. Microbiol.* **47**, 108–111 (2015).
- McDougall, S., Parkinson, T. J., Leyland, M., Anniss, F. M. & Fenwick, S. G. Duration of infection and strain variation in *Streptococcus uberis* isolated from cows' milk. *J. Dairy Sci.* **87**, 2062–2072. [https://doi.org/10.3168/jds.S0022-0302\(04\)70024-7](https://doi.org/10.3168/jds.S0022-0302(04)70024-7) (2004).
- Abureema, S., Smooker, P., Malmo, J. & Deighton, M. Molecular epidemiology of recurrent clinical mastitis due to *Streptococcus uberis*: Evidence of both an environmental source and recurring infection with the same strain. *J. Dairy Sci.* **97**, 285–290. <https://doi.org/10.3168/jds.2013-7074> (2014).
- Werner, C., Sauerwald, C., Sundrum, A., El-Sayed, A. & Zschöck, M. Genotyping of *Streptococcus uberis* isolates in healing process of bovine clinical mastitis. *Int. J. Vet. Sci. Med.* **6**, 274–278 (2018).
- Jolley, K., Bray, J. & Maiden, M. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications [version 1; peer review: 2 approved]. *Wellcome Open Res.* <https://doi.org/10.12688/wellcomeopenres.14826.1> (2018).
- Leigh, J. A., Egan, S. A., Ward, P. N., Field, T. R. & Coffey, T. J. Sortase anchored proteins of *Streptococcus uberis* play major roles in the pathogenesis of bovine mastitis in dairy cattle. *Vet. Res.* **41**, 63 (2010).
- Hossain, M. *et al.* Virulence related sequences; insights provided by comparative genomics of *Streptococcus uberis* of differing virulence. *BMC Genom.* **16**, 334. <https://doi.org/10.1186/s12864-015-1512-6> (2015).
- Egan, S. A., Ward, P. N., Watson, M., Field, T. R. & Leigh, J. A. Vru (Sub0144) controls expression of proven and putative virulence determinants and alters the ability of *Streptococcus uberis* to cause disease in dairy cattle. *Microbiology* **158**, 1581–1592. <https://doi.org/10.1099/mic.0.055863-0> (2012).
- Ward, P. N., Field, T. R., Ditcham, W. G. F., Maguin, E. & Leigh, J. A. Identification and disruption of two discrete loci encoding Hyaluronic acid capsule biosynthesis genes *hasA*, *hasB*, and *hasC* in *Streptococcus uberis*. *Infect. Immun.* **69**, 392–399. <https://doi.org/10.1128/iai.69.1.392-399.2001> (2001).
- Oliver, S. P., Almeida, R. A. & Calvino, L. F. Virulence factors of *Streptococcus uberis* isolated from cows with mastitis. *J. Vet. Med. B* **45**, 461–471. <https://doi.org/10.1111/j.1439-0450.1998.tb00817.x> (1998).
- Skalka, B. & Smola, J. Lethal effect of CAMP-factor and UBERIS-factor—a new finding about diffusible exo-substances of *Streptococcus agalactiae* and *Streptococcus uberis*. *Zentralbl. Bakteriolog. Mikrobiol. Hyg. A* **249**, 190–194. [https://doi.org/10.1016/S0174-3031\(81\)80073-X](https://doi.org/10.1016/S0174-3031(81)80073-X) (1981).
- Jiang, M., Babiuik, L. A. & Potter, A. A. Cloning, sequencing and expression of the CAMP factor gene of *Streptococcus uberis*. *Microb. Pathog.* **20**, 297–307. <https://doi.org/10.1006/mpat.1996.0028> (1996).

33. Smith, A. J. *et al.* MtuA, a lipoprotein receptor antigen from *Streptococcus uberis*, is responsible for acquisition of manganese during growth in milk and is essential for infection of the lactating bovine mammary gland. *Infect. Immun.* **71**, 4842–4849. <https://doi.org/10.1128/iai.71.9.4842-4849.2003> (2003).
34. Smith, A. J., Kitt, A. J., Ward, P. N. & Leigh, J. A. Isolation and characterization of a mutant strain of *Streptococcus uberis*, which fails to utilize a plasmin derived  $\beta$ -casein peptide for the acquisition of methionine. *J. Appl. Microbiol.* **93**, 631–639. <https://doi.org/10.1046/j.1365-2672.2002.01723.x> (2002).
35. Tian, X. Y. *et al.* Antimicrobial resistance and virulence genes of *Streptococcus* isolated from dairy cows with mastitis in China. *Microb. Pathog.* **131**, 33–39. <https://doi.org/10.1016/j.micpath.2019.03.035> (2019).
36. Lakshmi, R. Understanding the regulation of Group B *Streptococcal* virulence factors. *Future Microbiol.* **4**, 201–221. <https://doi.org/10.2217/17460913.4.2.201> (2009).
37. Spellerberg, B. *et al.* Lmb, a protein with similarities to the Lral adhesin family, mediates attachment of *Streptococcus agalactiae* to human Laminin. *Infect. Immun.* **67**, 871–878. <https://doi.org/10.1128/iai.67.2.871-878.1999> (1999).
38. Czabanska, A. *Immunochemical Investigations of the Cell Envelope Components Isolated from Streptococcus uberis* (University of Lübeck, Lübeck, 2013).
39. Ward, P. N. & Leigh, J. A. Characterization of PauB, a novel broad-spectrum plasminogen activator from *Streptococcus uberis*. *J. Bacteriol.* **184**, 119–125. <https://doi.org/10.1128/jb.184.1.119-125.2002> (2002).
40. Morona, J. K., Miller, D. C., Morona, R. & Paton, J. C. The effect that mutations in the conserved capsular polysaccharide biosynthesis genes *cpsA*, *cpsB*, and *cpsD* have on virulence of *Streptococcus pneumoniae*. *J. Infect. Dis.* **189**, 1905–1913. <https://doi.org/10.1086/383352> (2004).
41. Zadoks, R. N., Schukken, Y. H. & Wiedmann, M. Multilocus sequence typing of *Streptococcus uberis* provides sensitive and epidemiologically relevant subtype information and reveals positive selection in the virulence gene *pauA*. *J. Clin. Microbiol.* **43**, 2407–2417 (2005).
42. Field, T. R., Ward, P. N., Pedersen, L. H. & Leigh, J. A. The Hyaluronic acid capsule of *Streptococcus uberis* is not required for the development of infection and clinical mastitis. *Infect. Immun.* **71**, 132–139. <https://doi.org/10.1128/iai.71.1.132-139.2003> (2003).
43. Ward, P. N., Field, T. R., Rapier, C. D. & Leigh, J. A. The activation of bovine plasminogen by PauA is not required for virulence of *Streptococcus uberis*. *Infect. Immun.* **71**, 7193–7196. <https://doi.org/10.1128/iai.71.12.7193-7196.2003> (2003).
44. Egan, S. A., Kurian, D., Ward, P. N., Hunt, L. & Leigh, J. A. Identification of sortase A (SrtA) substrates in *Streptococcus uberis*: Evidence for an additional hexapeptide (LPXXXD) sorting motif. *J. Proteome Res.* **9**, 1088–1095. <https://doi.org/10.1021/pr901025w> (2010).
45. Lalioui, L. *et al.* The SrtA sortase of *Streptococcus agalactiae* is required for cell wall anchoring of proteins containing the LPXTG motif, for adhesion to epithelial cells, and for colonization of the mouse intestine. *Infect. Immun.* **73**, 3342. <https://doi.org/10.1128/IAI.73.6.3342-3350.2005> (2005).
46. Antos, J. M. *et al.* Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J. Am. Chem. Soc.* **131**, 10800–10801. <https://doi.org/10.1021/ja902681k> (2009).
47. Wuehrich, I. *et al.* Site-specific chemoenzymatic labeling of aerolysin enables the identification of new aerolysin receptors. *PLoS One* **9**, e109883. <https://doi.org/10.1371/journal.pone.0109883> (2014).
48. Barnett, T. C., Patel, A. R. & Scott, J. R. A novel Sortase, SrtC2, from *Streptococcus pyogenes* anchors a surface protein containing a QVPTGV Motif to the cell wall. *J. Bacteriol.* **186**, 5865–5875. <https://doi.org/10.1128/jb.186.17.5865-5875.2004> (2004).
49. Tomita, T. *et al.* Identification of *Streptococcus uberis* multilocus sequence types highly associated with mastitis. *Appl. Environ. Microbiol.* **74**, 114–124. <https://doi.org/10.1128/aem.01373-07> (2008).
50. Zadoks, R. N., Tikofsky, L. L. & Boor, K. J. Ribotyping of *Streptococcus uberis* from a dairy's environment, bovine feces and milk. *Vet. Microbiol.* **109**, 257–265. <https://doi.org/10.1016/j.vetmic.2005.05.008> (2005).
51. AVPG. *Australian Veterinary Prescribing Guidelines*. <https://vetantibiotics.fvas.unimelb.edu.au/bovine-medicine-guidelines/mastitis/>. (2020).
52. Lang, P. *et al.* Gene content differences across strains of *Streptococcus uberis* identified using oligonucleotide microarray comparative genomic hybridization. *Infect. Genet. Evol.* **9**, 179–188. <https://doi.org/10.1016/j.meegid.2008.10.015> (2009).
53. Tabrizi, A. D., Batavani, R. A., Rezaei, S. A. & Ahmadi, M. Fibrinogen and ceruloplasmin in plasma and milk from dairy cows with subclinical and clinical mastitis. *Pak. J. Biol. Sci.* **11**, 571–576 (2008).
54. Tassi, R. *et al.* Strain-specific pathogenicity of putative host-adapted and nonadapted strains of *Streptococcus uberis* in dairy cattle. *J. Dairy Sci.* **96**, 5129–5145. <https://doi.org/10.3168/jds.2013-6741> (2013).
55. Spellerberg, B., Martin, S., Brandt, C. & Lütticken, R. The *cyl* genes of *Streptococcus agalactiae* are involved in the production of pigment. *FEMS Microbiol. Lett.* **188**, 125–128. <https://doi.org/10.1111/j.1574-6968.2000.tb09182.x> (2000).
56. Spellerberg, B. *et al.* Identification of genetic determinants for the hemolytic activity of *Streptococcus agalactiae* by ISS1 transposition. *J. Bacteriol.* **181**, 3212–3219. <https://doi.org/10.1128/jb.181.10.3212-3219.1999> (1999).
57. Terao, Y. *et al.* Fba, a novel fibronectin-binding protein from *Streptococcus pyogenes*, promotes bacterial entry into epithelial cells, and the *fba* gene is positively transcribed under the Mga regulator. *Mol. Microbiol.* **42**, 75–86. <https://doi.org/10.1046/j.1365-2958.2001.02579.x> (2001).
58. Li, Q. *et al.* Fibronectin/fibrinogen-binding protein (FBPS) is not a critical virulence factor for the *Streptococcus suis* serotype 2 strain ZY05719. *Vet. Microbiol.* **208**, 38–46. <https://doi.org/10.1016/j.vetmic.2017.07.010> (2017).
59. O'Connor, S. P. & Cleary, P. P. *In Vivo Streptococcus pyogenes* C5a peptidase activity: Analysis using transposon- and nitroguanidine-Induced mutants. *J. Infect. Dis.* **156**, 495–504. <https://doi.org/10.1093/infdis/156.3.495> (1987).
60. Le Breton, Y. *et al.* Genome-wide discovery of novel M1T1 group A streptococcal determinants important for fitness and virulence during soft-tissue infection. *PLoS Pathog.* **13**, e1006584. <https://doi.org/10.1371/journal.ppat.1006584> (2017).
61. Cho, K. H. & Caparon, M. G. Patterns of virulence gene expression differ between biofilm and tissue communities of *Streptococcus pyogenes*. *Mol. Microbiol.* **57**, 1545–1556. <https://doi.org/10.1111/j.1365-2958.2005.04786.x> (2005).
62. Luther, D. A., Almeida, R. A. & Oliver, S. P. Elucidation of the DNA sequence of *Streptococcus uberis* adhesion molecule gene (*sua*) and detection of *sua* in strains of *Streptococcus uberis* isolated from geographically diverse locations. *Vet. Microbiol.* **128**, 304–312. <https://doi.org/10.1016/j.vetmic.2007.10.015> (2008).
63. Jakubovics, N. S., Smith, A. W. & Jenkinson, H. F. Expression of the virulence-related Sca (Mn<sup>2+</sup>) permease in *Streptococcus gordonii* is regulated by a diphtheria toxin metalloregulator-like protein ScaR. *Mol. Microbiol.* **38**, 140–153. <https://doi.org/10.1046/j.1365-2958.2000.02122.x> (2000).
64. Lönnerdal, B., Keen, C. L. & Hurley, L. S. Iron, copper, zinc, and manganese in milk. *Annu. Rev. Nutr.* **1**, 149–174. <https://doi.org/10.1146/annurev.nu.01.070181.001053> (1981).
65. Bateman, A., Holden, M. T. G. & Yeats, C. The G5 domain: A potential N-acetylglucosamine recognition domain involved in biofilm formation. *Bioinformatics* **21**, 1301–1303. <https://doi.org/10.1093/bioinformatics/bti206> (2004).
66. Feng, Y. *et al.* Attenuation of *Streptococcus suis* virulence by the alteration of bacterial surface architecture. *Sci. Rep.* **2**, 710. <https://doi.org/10.1038/srep00710> (2012).
67. Auger, J.-P., Dolbec, D., Roy, D., Segura, M. & Gottschalk, M. Role of the *Streptococcus suis* serotype 2 capsular polysaccharide in the interactions with dendritic cells is strain-dependent but remains critical for virulence. *PLoS One* **13**, e0200453. <https://doi.org/10.1371/journal.pone.0200453> (2018).

68. Fulde, M. & Bergmann, S. *Moonlighting Proteins: Novel Virulence Factors in Bacterial Infections* 245–268 (Wiley, New York, 2017).
69. Ye, W. *et al.* Pneumococcal LytR protein is required for the surface attachment of both capsular polysaccharide and teichoic acids: Essential for pneumococcal virulence. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2018.01199> (2018).
70. Kallio, A. *et al.* Role of Pht proteins in attachment of *Streptococcus pneumoniae* to respiratory epithelial cells. *Infect. Immun.* **82**, 1683–1691. <https://doi.org/10.1128/iai.00699-13> (2014).
71. Kamshayev, D. & Rouviere-Yaniv, J. The histone-like protein HU binds specifically to DNA recombination and repair intermediates. *EMBO J.* **19**, 6527–6535. <https://doi.org/10.1093/emboj/19.23.6527> (2000).
72. Oberto, J., Nabti, S., Jooste, V., Mignot, H. & Rouviere-Yaniv, J. The HU Regulon is composed of genes responding to anaerobiosis, acid stress, high osmolarity and SOS induction. *PLoS One* **4**, e4367. <https://doi.org/10.1371/journal.pone.0004367> (2009).
73. Stinson, M. W., McLaughlin, R., Choi, S. H., Juarez, Z. E. & Barnard, J. Streptococcal histone-like protein: Primary structure of hlpA and protein binding to lipoteichoic acid and epithelial cells. *Infect. Immun.* **66**, 259–265 (1998).
74. Stucky, K. *et al.* Cloning and characterization of brnQ, a gene encoding a low-affinity, branched-chain amino acid carrier in *Lactobacillus delbrückii* subsp. *lacticus* DSM7290. *Mol. Gen. Genet.* **249**, 682–690. <https://doi.org/10.1007/BF00418038> (1995).
75. Wohlt, J. E., Clark, J. H., Derrig, R. G. & Davis, C. L. Valine, leucine, and isoleucine metabolism by lactating bovine mammary tissue. *J. Dairy Sci.* **60**, 1875–1882. [https://doi.org/10.3168/jds.S0022-0302\(77\)84118-0](https://doi.org/10.3168/jds.S0022-0302(77)84118-0) (1977).
76. Grisoldi, L. *et al.* The relationship between *S. aureus* and branched-chain amino acids content in composite cow milk. *Animals* **9**, 20 (2019).
77. Tiemeyer, W., Stohrer, M. & Giesecke, D. Metabolites of nucleic acids in bovine milk. *J. Dairy Sci.* **67**, 723–728. [https://doi.org/10.3168/jds.S0022-0302\(84\)81361-2](https://doi.org/10.3168/jds.S0022-0302(84)81361-2) (1984).
78. Ozturk, G. *et al.* The antimicrobial activity of bovine milk xanthine oxidase. *Int. Dairy J.* **102**, 104581. <https://doi.org/10.1016/j.idairyj.2019.104581> (2020).
79. Harrison, R. Milk xanthine oxidase: Properties and physiological roles. *Int. Dairy J.* **16**, 546–554. <https://doi.org/10.1016/j.idairyj.2005.08.016> (2006).
80. Stevens, C. R. *et al.* Antibacterial properties of xanthine oxidase in human milk. *Lancet* **356**, 829–830. [https://doi.org/10.1016/S0140-6736\(00\)02660-X](https://doi.org/10.1016/S0140-6736(00)02660-X) (2000).
81. Mendoza-Chamizo, B., Löbner-Olesen, A. & Charbon, G. Coping with reactive oxygen species to ensure genome stability in *Escherichia coli*. *Genes* **9**, 20 (2018).
82. Lechardeur, D. *et al.* The 2-cys peroxiredoxin alkyl hydroperoxide reductase C binds heme and participates in its intracellular availability in *Streptococcus agalactiae*. *Int. J. Biol. Chem.* **285**, 16032–16041. <https://doi.org/10.1074/jbc.M109.024505> (2010).
83. van Valen, L. A new evolutionary law. *Evol. Theory* **1**, 1–30 (1973).
84. Stockmann, C. *et al.* Clinical and epidemiological evidence of the Red Queen hypothesis in pneumococcal serotype dynamics. *Clin. Infect. Dis.* **63**, 619–626. <https://doi.org/10.1093/cid/ciw357> (2016).
85. Bonachela, J. A., Wortel, M. T. & Stenseth, N. C. Eco-evolutionary Red Queen dynamics regulate biodiversity in a metabolite-driven microbial system. *Sci. Rep.* **7**, 17655. <https://doi.org/10.1038/s41598-017-17774-4> (2017).
86. Schulte, R., Makus, C., Hasert, B., Michels, N. K. & Schultenburg, H. Multiple reciprocal adaptations and rapid genetic change upon experimental coevolution of an animal host and its microbial parasite. *Proc. Natl. Acad. Sci. USA* **107**, 7359–7364. <https://doi.org/10.1073/pnas.1003113107> (2010).
87. Velez, J. R. *et al.* Whole-genome sequence analysis of antimicrobial resistance genes in *Streptococcus uberis* and *Streptococcus dysgalactiae* isolates from Canadian dairy herds. *Front. Vet. Sci.* **4**, 63. <https://doi.org/10.3389/fvets.2017.00063> (2017).
88. Kappeli, N. *et al.* Sequence types and antimicrobial resistance profiles of *Streptococcus uberis* Isolated from bovine mastitis. *Front. Vet. Sci.* **6**, 234. <https://doi.org/10.3389/fvets.2019.00234> (2019).
89. Xu, X. *et al.* Distribution of serotypes, genotypes, and resistance determinants among macrolide-resistant *Streptococcus pneumoniae* isolates. *Antimicrob. Agents Chemother.* **54**, 1152–1159. <https://doi.org/10.1128/AAC.01268-09> (2010).
90. Clancy, J. *et al.* Molecular cloning and functional analysis of a novel macrolide-resistance determinant, *mefA*, from *Streptococcus pyogenes*. *Mol. Microbiol.* **22**, 867–879. <https://doi.org/10.1046/j.1365-2958.1996.01521.x> (1996).
91. MacAogáin, M. *et al.* Metagenomics reveals a core macrolide resistome related to microbiota in chronic respiratory disease. *Am. J. Respir. Crit. Care Med.* **202**, 433–447. <https://doi.org/10.1164/rccm.201911-2202OC> (2020).
92. Su, W. *et al.* Ribosome protection by antibiotic resistance ATP-binding cassette protein. *Proc. Natl. Acad. Sci.* **115**, 5157. <https://doi.org/10.1073/pnas.1803313115> (2018).
93. Hammerum, A. M., Flannagan, S. E., Clewell, D. B. & Jensen, L. B. Indication of transposition of a mobile DNA element containing the *vat(D)* and *erm(B)* genes in *Enterococcus faecium*. *Antimicrob. Agents Chemother.* **45**, 3223–3225. <https://doi.org/10.1128/AAC.45.11.3223-3225.2001> (2001).
94. Bozdogan, B. *et al.* A new resistance gene, *linB*, conferring resistance to lincosamides by nucleotidylation in *Enterococcus faecium* HM1025. *Antimicrob. Agents Chemother.* **43**, 925–929. <https://doi.org/10.1128/AAC.49.7.2716-2719.2005> (1999).
95. Petinaki, E. *et al.* Lincomycin resistance gene *inu (D)* in *Streptococcus uberis*. *Antimicrob. Agents Chemother.* **52**, 626–630. <https://doi.org/10.1128/AAC.01126-07> (2008).
96. Gravey, F. *et al.* Lincosamide resistance mediated by *inu (C)* (L phenotype) in a *Streptococcus anginosus* clinical isolate. *J. Antimicrob. Chemother.* **68**, 2464–2467. <https://doi.org/10.1093/jac/dkt255> (2013).
97. Achard, A., Villers, C., Pichereau, V. & Leclercq, R. New *inu (C)* gene conferring resistance to lincomycin by nucleotidylation in *Streptococcus agalactiae* UCN36. *Antimicrob. Agents Chemother.* **49**, 2716–2719. <https://doi.org/10.1128/AAC.49.7.2716-2719.2005> (2005).
98. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. *PLoS Biol.* **8**, e1000412. <https://doi.org/10.1371/journal.pbio.1000412> (2010).
99. ozmaps: Australia Maps. R package version 0.3.6. <https://CRAN.R-project.org/package=ozmaps> (2020).
100. R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria, 2020).
101. Alawneh, J. I. *et al.* Survey and sequence characterisation of bovine mastitis-associated *Escherichia coli* in dairy herds. *Front. Vet. Sci.* **7**, 1–16. <https://doi.org/10.3389/fvets.2020.582297> (2020).
102. Santos, J., Cerri, R., Ballou, M., Higginbotham, G. & Kirk, J. Effect of timing of first clinical mastitis occurrence on lactational and reproductive performance of Holstein dairy cows. *Anim. Reprod. Sci.* **80**, 31–45 (2004).
103. Coordinators, N. R. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **44**, D7–19. <https://doi.org/10.1093/nar/gkv1290> (2016).
104. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**(3), 2011. <https://doi.org/10.14806/ej.17.1.200> (2011).
105. Seemann, T. *et al.* Nullarbor. <https://github.com/tseemann/nullarbor> (2020).
106. Souvorov, A., Agarwala, R. & Lipman, D. J. SKESA: Strategic k-mer extension for scrupulous assemblies. *Genome Biol.* **19**, 153. <https://doi.org/10.1186/s13059-018-1540-z> (2018).
107. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153> (2014).
108. Zhou, Z. *et al.* The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny and *Escherichia coli* core genomic diversity. *Genome Res.* **30**, 138–152. <https://doi.org/10.1101/gr.251678.119> (2019).

109. Jolley, K., Bray, J. & Maiden, M. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* <https://doi.org/10.12688/wellcomeopenres.14826.1> (2018).
110. Meier-Kolthoff, J. P. & Göker, M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* **10**, 2182. <https://doi.org/10.1038/s41467-019-10210-3> (2019).
111. Page, A. J., Taylor, B. & Keane, J. A. Multilocus sequence typing by blast from de novo assemblies against PubMLST. *J. Open Source Softw.* **8**, 118–119. <https://doi.org/10.21105/joss.00118> (2016).
112. Nascimento, M. *et al.* PHYLOViZ 2.0: Providing scalable data integration and visualization for multiple phylogenetic inference methods. *Bioinformatics* **33**, 128–129. <https://doi.org/10.1093/bioinformatics/btw582> (2017).
113. Francisco, A. P., Bugalho, M., Ramirez, M. & Carrico, J. A. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinform.* <https://doi.org/10.1186/1471-2105-10-152> (2009).
114. Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P. & Göker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform.* **14**, 60. <https://doi.org/10.1186/1471-2105-14-60> (2013).
115. Page, A. J. *et al.* Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693. <https://doi.org/10.1093/bioinformatics/btv421> (2015).
116. Annon., The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, D258–D261 (2004).
117. Törönen, P., Medlar, A. & Holm, L. PANNZER2: A rapid functional annotation web server. *Nucleic Acids Res.* **46**, W84–W88. <https://doi.org/10.1093/nar/gky350> (2018).
118. Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423. <https://doi.org/10.1038/s41587-019-0036-z> (2019).
119. Käll, L., Krogh, A. & Sonnhammer, E. L. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* **35**, W429–W432 (2007).
120. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033> (2014).
121. Seemann, T. *ABRicate*. <https://github.com/tseemann/abricate>.
122. Chen, L., Zheng, D., Liu, B., Yang, J. & Jin, Q. VFDB 2016: Hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res.* **44**, D694–D697. <https://doi.org/10.1093/nar/gkv1239> (2015).
123. Hill, A. W., Finch, J. M., Field, T. R. & Leigh, J. A. Immune modification of the pathogenesis of *Streptococcus uberis* mastitis in the dairy cow. *FEMS Immunol. Med. Microbiol.* **8**, 109–117. <https://doi.org/10.1111/j.1574-695X.1994.tb00432.x> (1994).
124. Metsalu, T. & Vilo, J. ClustVis: A web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res.* **43**, W566–W570. <https://doi.org/10.1093/nar/gkv468> (2015).
125. Alcock, B. P. *et al.* CARD 2020: Antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **48**, D517–D525. <https://doi.org/10.1093/nar/gkz935> (2019).
126. van Heel, A. J. *et al.* BAGEL4: A user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res.* **46**, W278–W281. <https://doi.org/10.1093/nar/gky383> (2018).
127. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinform.* **10**, 421. <https://doi.org/10.1186/1471-2105-10-421> (2009).
128. Johnson, M. *et al.* NCBI BLAST: A better web interface. *Nucleic Acids Res.* **36**, W5–W9. <https://doi.org/10.1093/nar/gkn201> (2008).
129. Robertson, J. & Nash, J. H. E. MOB-suite: Software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb. Genom.* <https://doi.org/10.1099/mgen.0.000206> (2018).
130. Arndt, D. *et al.* PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21. <https://doi.org/10.1093/nar/gkw387> (2016).
131. Xie, Z. & Tang, H. ISEScan: Automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* **33**, 3340–3347. <https://doi.org/10.1093/bioinformatics/btx433> (2017).

## Acknowledgements

We would like to acknowledge Joanne Mollinger and Ameh James for their contributions to this work.

## Author contributions

Conception and experimental design: J.I.A., T.O., M.S. Acquisition of data: H.A., H.R., J.I.A. Analysis and interpretation of data: B.V., J.I.A., R.M. Drafted the manuscript: B.V., J.I.A., H.A. All authors contributed to and revised the manuscript.

## Competing interests

This work was funded by Science with Impact Scheme, the University of Queensland, Sub-tropical Dairy Board, Dairy Australia, and Terragen Biotech Ltd, Australia. The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-82357-3>.

**Correspondence** and requests for materials should be addressed to J.I.A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://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