***Materials and methods***

STROBE-VET (Strengthening the Reporting of Observational Studies in Epidemiology–Veterinary Extension) statement guidelines were followed in the reporting of this study (O'Connor et al., 2016). Animal use for this project was approved by the University of Vermont Institutional Animal Care and Use Committee (IACUC; protocol #19-001).

*Sample origination*

Isolates included in the current study originate from milk samples collected during a longitudinal, cross-sectional observational study of 10 certified organic dairy farms in Vermont (US) carried out in Winter 2019-2020. Enrolled farms were a non-probability subsample of certified organic dairies in Vermont which had participated in previous studies, and inclusion criteria included: 1) milking between 35-120 cows and 2) using either a tiestall barn bedded with shavings/sawdust or a bedded pack system to house lactating dairy cows. The inclusive term “bedded pack” is used here to encompass both aerobically composting bedded packs and deep bedded packs, and was defined as an enclosed loose housing facility deeply bedded with organic material (Jeffrey et al., 2024). Around the time of the first farm visit, herd records were captured from the record processing center working with 9 of the participating herds (Lancaster DHIA, Manheim, PA; Dairy One Co-Op. Inc., Ithaca, NY) to obtain freshening date and parity for the current lactation. Freshening date and parity for 1 herd was obtained from personal communication with the producer who kept written records. The goal was to enroll 35 cows of varying parity in early- to mid-lactation from each herd for the duration of the study. In 1 herd with approximately 35 lactating cows, all cows were sampled. In 8 herds with ≥ 35 cows and with available DHIA data, a stratified random approach was used with cows stratified by SCC, lactation number, and DIM and then randomly selected across these variables. In 1 herd with ≥ 35 cows and no DHIA data, the producer generated a list of 35 cows in early lactation so that they would continue to be milking for the duration of the study. Cows that were unable to be sampled at a follow-up visit (dried off, left the herd) were replaced with another lactating cow dictated by convenience. At each farm visit, duplicate quarter-milk samples were aseptically collected from each lactating quarter immediately before milking for all enrolled cows according to NMC guidelines (NMC, 2017). Briefly, after routine pre-milking teat disinfection was completed, researchers (wearing clean disposable gloves) scrubbed teat ends and the distal third of teats with 70% isopropyl alcohol-moistened gauze swabs until teat ends were visibly clean, stripped the quarters (discarding 3-5 squirts of foremilk), and sequentially collected approximately 5-6 mL of milk into each of two sterile 11-mL flip-top vials. If any sign of clinical mastitis was present, it was noted and that sample was excluded from the inclusion in this study. Samples were kept on ice in a cooler during transport until stored temporarily overnight at 4°C in the laboratory, where an aliquot was frozen for SCC measurement and the remaining milk sample was processed for bacteriological culture.

*SCC measurement*

Aliquots of frozen quarter-milk samples were sent to the Vermont State Agricultural and Environmental Laboratory, where samples were gradually thawed under refrigeration at time of processing and quarter-level somatic cell count was determined using flow cytometry (Somacount FC, Bentley Instruments).

*Aerobic culture of milk samples and determination of bacteriological status*

Standard aerobic bacteriological culture of quarter-milk was performed in duplicate within 24 hours of collection to identify bacterial species present in the sample. After being homogenized by gentle inversion, tryptic soy agar plates with 5% sheep blood (Northeast Laboratory, Waterville, ME) were inoculated with 10 μL of milk using disposable calibrated plastic inoculating loops. Plates were then incubated in aerobic conditions at 37°C before being read at approximately 24 and 48 hrs.

Aerobic culture results of both samples were then used together to determine the overall bacteriological status of each quarter-milk sample into the following categories: 1) “no significant growth,” when there was no growth on both plates, or ≤ 200 CFU/mL on one plate and no growth on the other plate, or ≤ 200 CFU/mL on both plates and morphology of isolates on each plate was different; 2) “pure culture,” when there was ≥ 100 CFU/mL of a particular isolate identified with the same morphology on both plates; 3) “mixed culture,” when there was ≥ 100 CFU/mL of two phenotypically-distinct isolates identified, each growing on both plates; 4) “contaminated,” when 1 or both of the 2 samples had more than 2 morphologically distinct isolates growing on a plate; 5) and “indeterminate,” when the set of quarter-milk samples did not meet the criteria for any of the previous categories (e.g., missing duplicate).

*Identification of bacterial isolates to species*

Isolates from both pure and mixed culture quarter-milk samples were selected and grown in pure culture on blood agar. Standard benchtop tests were done to presumptively identify bacteria following NMC procedure guidelines, including differential growth on selective media, colony morphology, hemolytic pattern, catalase reaction, Gram stain, and coagulase testing (NMC, 2017). Isolates were preserved in tryptic soy broth with a final concentration of 15% glycerol in cryovials and stored at -80°C. Frozen isolates were sent overnight on ice to the University of Missouri for speciation using MALDI-TOF mass spectrometry (Microflex, Bruker Daltonics) with Flex Control software (Bruker Daltonics). The protocol for identifying bacterial isolates with MALDI-TOF mass spectrometry has been described previously (Haw et al., 2024). Briefly, generated spectra were assigned a score based on similarity with spectra in the manufacturer’s database (MBT 8468 MSP Library), as well as the University of Missouri laboratory custom database (Adkins et al., 2018). The confidence levels used for NASM species identification were applied as previously described (Cameron et al., 2017), in which ≥ 1.7 was used for species-level identification and < 1.7 was classified as inconclusive. Suspect staphylococci and mammaliicocci isolates unable to be identified to the species level and those identified as *Staphylococcus agnetis* or *Staphylococcus hyicus* by MALDI-TOF were speciated using *tuf*gene sequences with a cut-off of 98% identity as previously described (Hwang et al., 2011).

*Determination of IMI status and selection of isolates*

Using the bacteriological status and speciation information, a quarter-day IMI status was assigned to each quarter observation: 1) “healthy,” when there was no significant growth; 2) “infected with *S. chromogenes* only,” when ≥ 100 CFU/mL of *S. chromogenes* was identified in pure culture on both plates (interpretation in series; Dohoo et al., 2011); 3) “mixed infection with *S. chromogenes*,” when ≥ 100 CFU/mL of *S. chromogenes* and an additional species were identified in mixed culture on both plates; 4) “infected with pathogen other than *S. chromogenes*,” when ≥ 100 CFU/mL of a species besides *S. chromogenes* was identified in pure or mixed culture on both plates; and 5) “unknown” if the sample status had been identified as contaminated or indeterminate as previously described. Quarter-day observations were eligible for inclusion in further analysis if they had an associated quarter-level SCC measurement and the IMI status was classified as infected with *S. chromogenes* only.

A given quarter was considered to have a potentially persistent *S. chromogenes* IMI if: 1) it had ≥ 2 quarter-day observations (from sequential sampling events approximately 30 days apart); 2) IMI status could be determined for all sampling events associated with that quarter; and 3) it was infected with *S. chromogenes* only for all associated quarter-day observations throughout the study. Two groups were then selected from all potentially persistent *S. chromogenes* IMI: 1) IMI associated with high SCC, where all quarter-day observations had an associated SCC of ≥ 200,000 cells/mL; and 2) IMI associated with low SCC, where all quarter-day observations had an associated SCC of < 200,000 cells/mL. Any potentially persistent *S. chromogenes* IMI that did not fit into one of these two categories was excluded from further analysis (e.g., had an SCC of < 200,000 cells/mL for one quarter-day observation and an SCC of ≥ 200,000 cells/mL for the next).

*Strain-typing and selection of isolates*

All isolates associated with each potentially persistent high and low SCC *S. chromogenes* IMI were strain typed using random amplification of polymorphic DNA (RAPD)-PCR. DNA was extracted using a commercial kit from overnight broth culture following the manufacturer’s instructions (DNeasy Blood and Tissue Kit, Qiagen) and then stored at -20 °C until further analysis. RAPD-PCR was performed as described by Wuytack et al. (2020) using the primer set D11344 (as described by Fitzgerald et al., 1997) with the following PCR conditions: 4 cycles of 94 °C at 5 min, 36 °C at 5 min, and 72 °C at 5 min and 30 cycles of 94 °C at 1 min, 36 °C at 1 min, and 72 °C at 2 min. Amplified DNA fragments were then separated on 1.5% (wt/vol) agarose gels stained with SYBR Safe (0.1 µL/mL; ThermoFisher Scientific), at 120 V for 75 min, and then photographed by UV transillumination (Image Lab, Bio-Rad). All isolates from a given persistent IMI were analyzed in the same PCR run and were run side-by-side on the gel. The images were inspected visually, and isolates with the same banding pattern, number, and size of bands were considered to be the same RAPD type. If theisolates from all quarter observations of an *S. chromogenes* IMI belonged to the same RAPD type, the quarter was considered persistently infected with the same strain.

From among the confirmed persistent *S. chromogenes* IMI, 15 quarters with a persistently low SCC IMI (LOW) were selected to match the 15 quarters with a persistently high SCC IMI (HIGH). LOW quarters were matched to HIGH quarters belonging to the same cow (different quarter) when possible. If this was not possible, LOW and HIGH quarters were matched on farm, or facility type (bedded pack vs. tiestall) when same farm was not possible. When LOW and HIGH quarters were paired between different cows, quarters were matched as closely as possible to ensure a similar DIM and parity. From each of the 15 HIGH and 15 LOW IMI, a representative isolate was chosen to undergo whole-genome sequencing (WGS). For each persistent IMI which had 3 associated quarter-day observations, the middle isolate in the series was submitted for WGS. For persistent IMI which had 2 associated quarter-day observations, 1 of the 2 isolates in the series was haphazardly selected for WGS.

*DNA extraction, whole-genome sequencing, assembly, and annotation*

Each of the 30 isolates selected for WGS were grown from frozen stock on blood agar in aerobic conditions at 37°C, and read at approximately 24 and 48 hrs. All plates were then inspected to ensure purity, and a single colony was selected and passed to a new blood agar plate. After again being incubated at 37°C, read at approximately 24 and 48 hrs, and checked for contamination, 48-hr growth plates were wrapped in Parafilm (Amcor). Wrapped plates were sent overnight to a commercial sequencing facility (SeqCoast Genomics; Portsmouth, NH, USA) for DNA extraction, library preparation, long read sequencing using GridION Oxford Nanopore, paired-end sequencing using Illumina, assembly, and annotation. DNA extraction was performed on colony material collected from the agar plates with a commercial kit using bead beating lysis (MagMAX Microbiome Ultra Nucleic Acid Isolation Kit, Applied Biosystems). Library preparation was completed using Illumina DNA Prep tagmentation kit (Illumina), and paired-end sequencing (2x150bp) was run on the Illumina NextSeq2000 platform (Illumina). 1-2% PhiX control was spiked into the run to support optimal base calling. Read demultiplexing, read trimming, quality control (FastQC) and run analytics were performed on the instrument using DRAGEN v3.10.12. Library preparation for long-read sequencing was completed using the Oxford Nanopore Technologies SQK-LSK114 native barcoding kit, and sequencing was performed on the GridION platform (FLOW-MIN114 Spot-ON Flow Cell, vR10). Quality-trimming of raw reads was completed using Trimmomatic v0.39 (Bolger et al., 2014), and read error correction and assembly of short and long reads were completed using SPAdes v3.14.0 (Bankevich et al., 2012) wrapped in Unicycler v0.4.4 (Wick et al., 2017). Read mapping was completed using Bowtie2 (Langmead and Salzberg, 2012) and SAMtools (Li et al., 2009), and the assembly was polished using Pilon (Walker et al., 2014), with all three tools wrapped in Unicycler. Annotation was completed using BAKTA v1.5.1 (Schwengers et al., 2021).

*Bioinformatic analyses, in silico multilocus sequence typing, and detection of ARG genes and pVF*

Whole genome multilocus sequence types were predicted *in silico* from the annotated genomes for the 7-locus scheme described for *S. chromogenes* (Huebner et al., 2021) using the MLST 2.0 tool (Center for Genomic Epidemiology, Technical University of Denmark, Kongens Lyngby, Denmark; software v2.0.9, database v2023-06-19; MLST allele sequence and profile data obtained from PubMLST.org). Any novel alleles identified by were confirmed using PCR and Sanger sequencing. The 7-locus concatenated nucleotide sequence data were then combined with all 386 available concatenated MLST sequences for *S. chromogenes* in PubMLST. The resulting FASTA file was used for the construction of phylogenetic trees by maximum-likelihood algorithm with the optimal model and 100 bootstrap replications in MEGA-X (Kumar et al., 2018). Isolates which grouped together with a bootstrap value of ≥ 65% were classified as clusters.

Prevalence of ARGs was evaluated using data from 5 databases using assembled genomes: 1) ResFinder from the Center for Genomic Epidemiology v4.5.0 (Camacho et al., 2009; Bortolaia et al., 2020), 2) Comprehensive Antibiotic Resistance Database vXXX (CARD) (Alcock et al., 2020), 3) MegaRES v3.0 (Bonin et al., 2023), 4) ARG-ANNOT vXXXX (Antibiotic Resistance Gene-ANNOTation) (XXX), and 5) AMRFinderPlus from NCBI v3.12.8 (Feldgarden et al., 2021).

Deduplication?

Cut-offs?

Specific to CNS?

pVF – construction of pipeline, cite Naushad

Deduplication, cut-offs for percent identity? (look at Naushad)

Classification and distribution of virulence factors

*Statistical analysis*

A spreadsheet with isolate identification, associated metadata, and outcome variables was made (Microsoft Excel, Redmond, WA) and imported into the R Statistical Programming Environment (R Development Core Team, 2023) for analysis. Descriptive statistics were generated to compare parity and DIM of the cow, quarter location, and average SCC associated with each persistent IMI between the two SCC groupings (HIGH vs. LOW). Normality of the data was checked using a Shapiro test. For outcomes which were not normally distributed (parity, DIM, average SCC), a Mann Whitney U test was used to compare metrics between the HIGH and LOW groups. For outcomes which were normally distributed (quarter location), Fisher’s Exact test was used to compare the two groups. Statistical significance for these tests were declared at P ≤ 0.05.

*Data availability*

The raw reads from ONT and Illumina for all 30 genomes are available under NCBI Bioproject accession number PRJNA1130504 (Biosamples SAMN42232476 to SAMN42232505) in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/).

Adkins, P. R. F., S. Dufour, J. N. Spain, M. J. Calcutt, T. J. Reilly, G. C. Stewart, and J. R. Middleton. 2018. Molecular characterization of non-aureus Staphylococcus spp. from heifer intramammary infections and body sites. J. Dairy Sci. 101(6):5388-5403.

Alcock, B. P., A. R. Raphenya, T. T. Y. Lau, K. K. Tsang, M. Bouchard, A. Edalatmand, W. Huynh, A. V. Nguyen, A. A. Cheng, S. Liu, S. Y. Min, A. Miroshnichenko, H. K. Tran, R. E. Werfalli, J. A. Nasir, M. Oloni, D. J. Speicher, A. Florescu, B. Singh, M. Faltyn, A. Hernandez-Koutoucheva, A. N. Sharma, E. Bordeleau, A. C. Pawlowski, H. L. Zubyk, D. Dooley, E. Griffiths, F. Maguire, G. L. Winsor, R. G. Beiko, F. S. L. Brinkman, W. W. L. Hsiao, G. V. Domselaar, and A. G. McArthur. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 48(D1):D517-d525.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, and P. A. Pevzner. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5):455-477.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114-2120.

Bonin, N., E. Doster, H. Worley, L. J. Pinnell, J. E. Bravo, P. Ferm, S. Marini, M. Prosperi, N. Noyes, P. S. Morley, and C. Boucher. 2023. MEGARes and AMR++, v3.0: an updated comprehensive database of antimicrobial resistance determinants and an improved software pipeline for classification using high-throughput sequencing. Nucleic Acids Res 51(D1):D744-d752.

Bortolaia, V., R. S. Kaas, E. Ruppe, M. C. Roberts, S. Schwarz, V. Cattoir, A. Philippon, R. L. Allesoe, A. R. Rebelo, A. F. Florensa, L. Fagelhauer, T. Chakraborty, B. Neumann, G. Werner, J. K. Bender, K. Stingl, M. Nguyen, J. Coppens, B. B. Xavier, S. Malhotra-Kumar, H. Westh, M. Pinholt, M. F. Anjum, N. A. Duggett, I. Kempf, S. Nykäsenoja, S. Olkkola, K. Wieczorek, A. Amaro, L. Clemente, J. Mossong, S. Losch, C. Ragimbeau, O. Lund, and F. M. Aarestrup. 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother 75(12):3491-3500.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10(1):421.

Cameron, M., H. W. Barkema, J. De Buck, S. De Vliegher, M. Chaffer, J. Lewis, and G. P. Keefe. 2017. Identification of bovine-associated coagulase-negative staphylococci by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry using a direct transfer protocol. J. Dairy Sci. 100(3):2137-2147.

Dohoo, I., S. Andersen, R. Dingwell, K. Hand, D. Kelton, K. Leslie, Y. Schukken, and S. Godden. 2011. Diagnosing intramammary infections: Comparison of multiple versus single quarter milk samples for the identification of intramammary infections in lactating dairy cows. J. Dairy Sci. 94(11):5515-5522.

Feldgarden, M., V. Brover, N. Gonzalez-Escalona, J. G. Frye, J. Haendiges, D. H. Haft, M. Hoffmann, J. B. Pettengill, A. B. Prasad, G. E. Tillman, G. H. Tyson, and W. Klimke. 2021. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci Rep 11(1):12728.

Fitzgerald, J. R., W. J. Meaney, P. J. Hartigan, C. J. Smyth, and V. Kapur. 1997. Fine-structure molecular epidemiological analysis of <i>Staphylococcus aureus</i> recovered from cows. Epidemiology and Infection 119(2):261-269.

Haw, S. R., P. R. F. Adkins, V. Bernier Gosselin, S. E. Poock, and J. R. Middleton. 2024. Intramammary infections in lactating Jersey cows: Prevalence of microbial organisms and association with milk somatic cell count and persistence of infection. J. Dairy Sci. 107(5):3157-3167.

Huebner, R., R. Mugabi, G. Hetesy, L. Fox, S. De Vliegher, A. De Visscher, J. W. Barlow, and G. Sensabaugh. 2021. Characterization of genetic diversity and population structure within Staphylococcus chromogenes by multilocus sequence typing. PLoS One 16(3):e0243688.

Hwang, S. M., M. S. Kim, K. U. Park, J. Song, and E. C. Kim. 2011. Tuf gene sequence analysis has greater discriminatory power than 16S rRNA sequence analysis in identification of clinical isolates of coagulase-negative staphylococci. J Clin Microbiol 49(12):4142-4149.

Jeffrey, C. E., T. Andrews, S. M. Godden, D. A. Neher, and J. W. Barlow. 2024. Relationship Between Facility Type and Bulk Tank Milk Bacteriology, Udder Health, Udder Hygiene, and Milk Production on Vermont Organic Dairy Farms. J. Dairy Sci.

Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35(6):1547-1549.

Langmead, B. and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9(4):357-359.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078-2079.

National Mastitis Council. 2017. Laboratory Handbook on Bovine Mastitis. Third ed. National Mastitis Council, Inc., New Prague, MI.

R Development Core Team. 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Schwengers, O., L. Jelonek, M. A. Dieckmann, S. Beyvers, J. Blom, and A. Goesmann. 2021. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microb Genom 7(11).

Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, and A. M. Earl. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9(11):e112963.

Wick, R. R., L. M. Judd, C. L. Gorrie, and K. E. Holt. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLOS Computational Biology 13(6):e1005595.

Wuytack, A., A. De Visscher, S. Piepers, F. Haesebrouck, and S. De Vliegher. 2020. Fecal non-aureus Staphylococci are a potential cause of bovine intramammary infection. Vet Res 51(1):32.