bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 **Microbiota in milk from healthy and mastitis cows varies greatly in diversity, species**
2 **richness and composition, as revealed by PacBio sequencing**

3

4 **Teng Ma, Lingling Shen, Qiannan Wen, Ruirui lv, Qiangchuan Hou, Lai Yu Kwok, Zhihong Sun,**
5 **and Heping Zhang***

6

7 Key Laboratory of Dairy Biotechnology and Engineering, Ministry of Education, Key Laboratory of
8 Dairy Products Processing, Ministry of Agriculture, Inner Mongolia Agricultural University, Huhhot, P.
9 R. China

10

11

12 **Correspondence**
13 Heping Zhang, College of Food Science and Engineering, Inner Mongolia Agricultural University,
14 Huhhot, P. R. China.
15 E-mail: hepingdd@vip.sina.com.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

16 **Abstract**

17 Mastitis is the most economically important disease of dairy cows. This study used PacBio
18 single-molecule real-time sequencing technology to sequence the full-length of the l6S rRNA from the
19 microbiota in 27 milk samples (18 from mastitis and 9 from healthy cows; the cows were at different
20 stages of lactation). We observed that healthy or late stage milk microbiota had significantly higher
21 microbial diversity and richness. The community composition of the microbiota from different groups
22 also varied greatly. In milk from healthy cows the microbiota was predominantly comprised of
23 *Lactococcus lactis*, *Acinetobacter johnsonii* and *Bacteroides dorei*, while from mastitis cows it was
24 predominantly comprised of *Bacillus cereus*, *Clostridium cadaveris* and *Streptococcus suis*. The
25 prevalence of *La. lactis* and *B. cereus* in milk from healthy and mastitis cows was confirmed by digital
26 droplets PCR. Differences in milk microbiota composition could suggest an important role for these
27 microbes in protecting the host from mastitis. Based on the milk microbiota profiles, the Udder Health
28 Index was constructed to predict the risk of bovine mastitis. Application of this predictive model could
29 aid early identification and prevention of mastitis in dairy cows, though the model requires further
30 optimisation using a larger dataset.

31

32 **KEYWORDS**

33 bovine mastitis, PacBio single-molecule real-time sequencing, l6S rRNA, digital droplets PCR, milk
34 microbiota

35

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

36 **1 INTRODUCTION**

37 Bovine mastitis is an inflammatory reaction that occurs in the cow's mammary glands; it is caused either
38 by invasion of pathogenic microbes or by physical/chemical stimulants (Rasmussen, Fogsgaard,
39 Rontved, Klaas, & Herskin, 2011). In general, depending on the degree of inflammation, mastitis can be
40 divided into subclinical and clinical mastitis. Clinical mastitis was diagnosed when the milk from the
41 udder had visible abnormalities (e.g. the presence of floc and/or granules in milk or watery/reduced
42 milk/no milk) or cows showing clinical symptoms, such as elevated body temperature, loss of
43 appetite/refusal to eat, swelling/oedema of the affected quarter (Carolina Espeche et al., 2012; Metzger et
44 al., 2018). In subclinical cases, animals are outwardly healthy, but when the somatic cell count (SCC) of
45 milk samples is greater than 200,000 cells/mL and there is no evidence of clinical infection, they are
46 classified as having subclinical mastitis (Pantoja, Hulland, & Ruegg, 2009). Bovine mastitis is one of the
47 most common and serious diseases in the dairy farming industry; not only does it lead to a decline in
48 milk production and milk quality, but it also increases the elimination and mortality rate of cows.
49 Approximately 33% of the 231 million cows in the world have mastitis, and annual losses due to mastitis
50 are as high as $3.5 billion (Halasa, Huijps, Osteras, & Hogeveen, 2007). Owing to the huge economic
51 significance of bovine mastitis, it is of major global concern.

52 The etiology of bovine mastitis is complicated and varies amongst countries and farms; it is
53 influenced by environmental/geographical factors, feeding management methods, sanitary conditions,
54 microbial infections, and attributes of individual cows (e.g. age, lactation stage, parity, body type,
55 heredity). In general, the main cause of mastitis is bacterial infection (Haltia, Honkanen-Buzalski,
56 Spiridonova, Olkonen, & Myllys, 2006), as supported by numerous reports of the isolation and

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

57 identification of mastitis-causing microbes using traditional microbiological methods. However,
58 traditional bacterial culture is relatively slow and laborious and no pathogenic bacteria are detected, using
59 conventional methods, in approximately 25% of milk samples collected from mastitis cows (Taponen,
60 Salmikivi, Simojoki, Koskinen, & Pyorala, 2009). Second-generation sequencing technology only
61 produces short sequence reads with low taxonomic resolution and has failed to identify mastitis-causing
62 microbes (Dohoo et al., 2011; Liao, Lin, & Lin, 2015). Third-generation sequencing technology, such as
63 the PacBio single-molecule real-time (SMRT) sequencing platform, is increasingly being used to
64 characterize the microbiota of environmental samples (e.g. dairy products (Hui et al., 2017), milk (Hou et
65 al., 2015) and food (Nakano et al., 2016) etc. In contrast to the older sequencing technologies, PacBio
66 SMRT sequencing is high throughput and produces long reads capable of microbial identification to
67 species level when used in conjunction with polymerase chain reaction (PCR) of full-length 16S rRNA
68 genes (Mosher et al., 2014).

69 This study aimed to identify differences in the microbiota of milk from healthy and mastitis cows at
70 different stages of lactation. Milk samples were collected from healthy and mastitis cows from the same
71 dairy. Microbiota profiles in these samples were characterized at the species level by PacBio SMRT
72 sequencing. Abundances of specific microbes were determined using ddPCR. With these data, an Udder
73 Health Index was constructed using the Random Forest Algorithm. Our results serve as a useful reference
74 for designing strategies to prevent and treat mastitis.

75

76 **2 MATERIALS AND METHODS**

77 **2.1 Experimental design and selection of cows**

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

78 This present study was conducted at Aoya Modern Ranch in Tai'an City (Shandong Province, China)
79 using 2- to 6-year-old Holstein cows. The average age of calving is 23-76 months, and the milk
80 production is about 3800 kg. From 26 dairy cows with severe clinical mastitis identified by professional
81 veterinarians, 18 were randomly selected as mastitis group. For comparison, 9 milk samples obtained
82 from healthy cows was used, and these cows had no history of mastitis and found to have a SCC lower
83 than 15,000 cells/mL. Milk from the mastitis cows contained obvious floc or granules which prevented
84 measurement of SCC. According to Vijayakumar et al (Vijayakumar et al., 2017). lactation in cows can
85 be divided into early ($1 \leq d \leq 100$), middle ($101 \leq d \leq 200$), and late lactation ($201 \leq d \leq$ dry milk
86 period). Dairy cows are managed in accordance with the practices of the herd in the pasture, providing
87 them with green fodder and calculating the amount of concentrated mixture. Furthermore, none of the
88 cows had received antibiotics or any other medication known to influence the microbiota of their milk, in
89 the three months before the study.
90

91 **2.2 Collection of milk samples**

92 Milk samples were collected from 27 cows in total (18 mastitis and 9 healthy cows; 8 early, 5 middle and
93 14 late lactation stage cows); detailed information was shown in Table S1. Quarter milk samples were
94 collected during the morning milking from all cows in the following way: Udders were thoroughly wiped
95 using a clean dry cloth to remove bedding and visible contaminants; teats were sanitized; two or three
96 streams of foremilk per teat were discarded; a 0.5% iodine solution was applied to the udder; after 60s
97 the iodine was wiped off using another clean dry cloth towel; the collector then put on clean gloves; a
98 further two to three streams of milk per teat were discarded; all teats were scrubbed with 70%

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

99 isopropanol; a further two streams of milk per teat were discarded after the isopropanol had dried;
100 finally, approximately 10ml were collected into a sterile vial. All raw milk was stored and transported
101 back to the laboratory on ice.
102

103 **2.3 Extraction of metagenomic DNA**

104 DNA was extracted from 3 ml of each raw milk sample using the PowerFood™ Microbial DNA
105 Isolation Kit (MoBio Laboratories, Qiagen, USA) following the manufacturer's instructions. The quality
106 of the extracted genomic DNA was examined by agarose gel electrophoresis and spectrophotometer
107 analysis (ratio of optical density at 260 nm/280 nm). High quality DNA samples were temporarily stored
108 in the refrigerator at -20°C prior use and for no longer than 6 hours.
109

110 **2.4 Droplet digital PCR**

111 To verify the sequencing results, ddPCR was used to quantify *La. lactis* and *B. cereus*, as they were the
112 most abundant species in milk from healthy and mastitis cows, respectively. Nine samples from each
113 group were included in this analysis. The ddPCR was done using the QX200 system (Bio-Rad, Hercules,
114 CA, USA). Primer 5.0 software was used to design the primers targeting *La. lactis* (lactisF,
115 5'-AGCAGTAGGGAATCTTCGGCA-3'; lactisR, 5'-GGGTAGTTACCGTCACTTGATGAG-3') and *B.*
116 *cereus* (PCERF, 5'-GGATTCATGGAGCGGCAGTA-3'; PCERR3,
117 5'-GCTTACCTGTCATGGTGTAACTTCA-3') based on the species-specific genomic region (Francisco
118 Martinez-Blanch, Sanchez, Garay, & Aznar, 2011; Ma et al., 2018). The ddPCR reaction solution was
119 prepared by mixing 2 μl DNA, 0.2 μl each of the upstream and downstream primers, 10 μl 2-fold

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

120 EvaGreen ddPCRSuperMix and 7.6 μl of sterilized ultra-pure water. The prepared ddPCR reaction
121 solution and the droplet generation oil were added to the droplet generation card and partitioned into
122 20,000 droplets per sample by the QX200 droplet generator (Cremonesi et al., 2016). The droplets
123 produced by each sample were transferred to a 96-well plate and PCR amplification carried out using the
124 EvaGreen program: 95°C for 10 min, 40-cycles of 94°C for 30 sec, 60°C for *La. lactis* and 62°C for *B.*
125 *cereus* for 1 min, and 4°C for 5 min, followed by 90°C for 5 min and a hold at 4°C. After thermal
126 cycling, the 96-well plate was loaded into the QX200 droplet reader. Data were collected using
127 QuantaSoft software, and the number of target DNA molecules calculated based on Poisson statistics.
128
129 **2.5 Amplification of full-length 16S rRNA genes and SMRT sequencing**
130 The 16S rRNA genes were amplified by PCR using the purified extracted genomic DNA as templates,
131 with the universal primer pair 27F (5′-AGAGTTTGATCCTGGCTCAG-3′) and 1495R
132 (5′-CTACGGCTACCTTGTTACGA-3′) (Mosher, Bernberg, Shevchenko, Kan, & Kaplan, 2013). At the
133 same time, to identify different samples in the same library, a 16-base identification barcode was added
134 to both ends of all primers. The specific amplification conditions were as follows: 95°C for 3 min, 98°C
135 for 20 s, 60°C for 15 s, 72°C for 30 s, 30 cycles and then 72°C for 2 min.
136 PCR products from each sample were purified and equal mass-mixed to construct the DNA library
137 for sequencing by the Agilent DNA 1000 Kit, according to the manufacturer's instructions. The
138 constructed library, sequencing primers and DNA polymerase, were added to the SMRT cells for
139 sequencing using the PacBio RS II instrument.
140

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

141 **2.6 Bioinformatics processing of high-quality sequences**

142 Quality control of raw data was achieved using the RS_ReadsOfinsert.1 protocol available in the SMRT
143 Portal version 2.3. Specific quality control conditions were based on the following criteria: minimum
144 cycle sequencing number, minimum prediction accuracy, minimum insertion sequence length and
145 maximum insertion sequence length, which were set to 5, 90, 1,400 and 1,800 respectively (Hou et al.,
146 2015). Then all sequences were divided into different samples according to the barcodes on the
147 amplicons. After removal of barcodes and primer sequences, bioinformatics analysis of high-quality
148 sequences was done using the QIIME package (version 1.7), briefly, PyNAST (Caporaso et al., 2010)
149 was used to align the sequences, and UCLUST (Edgar, 2010) was done under 100% clustering of
150 sequence identity to obtain representative sequences. Afterwards, sequences were classified into
151 operational taxonomic units (OTU) according to 97% similarity. Chimeric OTU sequences were detected
152 and removed using ChimeraSlayer (Haas et al., 2011). The Ribosomal Database Project (RDP, Release
153 11.5) (Cole et al., 2007), Greengenes (version 13.8) (DeSantis et al., 2006) and Silva (version
154 128) (Quast et al., 2013) databases were used to assign the taxonomy of each OTU representative
155 sequence with an 80% confidence threshold (Hou et al., 2015). A *de novo* taxonomic tree was
156 constructed for representative OTU sets using FastTree for downstream analysis (Price, Dehal, & Arkin,
157 2009). The alpha diversity of each sample was assessed based on the sample with the lowest sequencing
158 depth. All sequencing data generated have been uploaded to the MG-RAST database under the project
159 number mgp88495 (http://www.mg-rast.org).

160

161 **2.7 Random Forest algorithm for predicting cow udder health**

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

162 Here a mastitis prediction model was built to predict the udder health status of dairy cows based on the
163 default settings of the Random Forest machine learning algorithm available in R package (3.1.1) (Cutler,
164 Edwards, Beard, Cutler, & Hess, 2007). Briefly, the prediction was based on regression of the milk
165 microbiota profiles of all 27 cows against their mastitis health status. The Random Forest algorithm
166 ranked all species into 'character importance' and used the 'rfcv' function to generate more than 100
167 predictions to determine the number of top discriminant species required for calculating the Udder Health
168 Index. Then the Udder Health Index was calculated based on the relative abundance of the predicted top
169 discriminatory milk microbes.

170

171 **2.8 Statistical analyses**

172 The R software package was used for statistical analysis (http://www.rproject.org/). Wilcoxon and
173 Kruskal-Wallis tests were used to evaluate differences in microbial between groups with a cut-off
174 confidence level of 95%. Principal coordinate analysis (PCoA) was used to observe the distribution of
175 samples. Permutational multivariate analysis of variance (PERMANOVA) (Anderson & Walsh, 2013)
176 was done on the different groups of samples in the R language 'vegan' (Nilsson et al., 2010) package to
177 reveal the effect of different groups on the target microbiota of healthy and mastitis cows. The 'ggplot2'
178 package (Ginestet, 2011) was used to analyze and visualize the raw results obtained by QIIME.
179 Cytoscape 3.5.1 was used for network building.

180

181 **3 RESULTS**

182 **3.1 Sequence coverage and alpha diversity of microbiota from milk**

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

183 A total of 169,308 high-quality 16S rRNA sequences were produced from the 27 samples included in this
184 study (average = 6,270; range = 1,010-15,174; SD = 3,905). After PyNAST alignment and UCLUST
185 classification at 100% similarity, 82,592 representative sequences remained. Upon removal of chimeric
186 sequences, they were classed into 6,850 OTUs for downstream analysis.

187 All Shannon diversity curves leveled off, suggesting that the sequencing depth was sufficient to
188 capture representative microbial populations present in the samples (Figure S1), although new
189 phylotypes may still be found with further sequencing. The microbial abundance and species richness in
190 each sample were assessed by the number of observed OTUs and the Shannon diversity index,
191 respectively. Both the number of observed OTUs and Shannon diversity index were significantly higher
192 in healthy compared with mastitis cows (Figure 1a, b; $P < 0.001$). We also found that both indices were
193 significantly higher during late compared with early lactation, while no significant difference was
194 observed amongst other lactation stages (Figure 1c, d; $P < 0.05$).

195

196 **3.2 Identity of microbiota in milk**

197 *Firmicutes* (73.0%) and *Bacteroidetes* (10.9%) were the two main bacterial phyla. At the genus level,
198 there were 12 genera with an average relative abundance of more than 1% (Figure 2a), including *Bacillus*
199 (28.5%), *Clostridium* (10.6%), *Lactobacillus* (10.4%), *Lactococcus* (7.9%) and *Bacteroides* (6.1%). At
200 the species level, 15 of which had an average relative abundance of over 1% (Figure 2c), including *B.*
201 *cereus* (27.6%), *La. lactis* (7.8%), *Lactobacillus helveticus* (6.3%), *Clostridium limosum* (4.4%),
202 *Helcococcus ovis* (3.4%), and *Clostridium cadaveris* (3.2%) et al.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

203 The mean microbiota composition varied largely between groups. Whereas the horizontal lines in the
204 box-plots indicate the median, we found that more *Lactobacillus*, *Lactococcus*, *Bacteroides* and
205 *Acinetobacter* were detected in healthy cows, while more *Bacillus*, *Clostridium* and *Streptococcus* were
206 observed in mastitis cows (Figure 2b). Significantly more *La. lactis* was found in the milk of healthy
207 cows, while *B. cereus* was found in the mastitis cows ($P < 0.05$, Figure 2d). The relative abundance of
208 *Bacillus* and *Clostridium* were not significantly different in milk collected during the early and middle
209 compared with late lactation. In addition, significantly more *La. lactis*, *S. suis*, *A. johnsonii*, and *B. dorei*
210 were detected in milk from late compared with early and middle lactation cows ($P < 0.05$). The relative
211 abundance of *H. ovis* and *P. heparinolytica* gradually decreased as lactation day increased (Figure S2).
212

213 **3.3 Comparative analysis of community structure in milk from cows**

214 A PCoA was done based on the Bray Curtis distance to visualize differences in the milk microbiota
215 between the healthy and mastitis groups. Two distinct clusters formed on the PCoA score plot,
216 representing milk samples collected from healthy and mastitis cows, respectively (Figure 3a). This
217 suggests large difference in the microbiota community structure and compositon between the two groups.
218 Results of the PERMANOVA test revealed that mastitis was a significant factor contributing to
219 differences in microbiota between the two groups ($P = 0.001$).

220 When samples were grouped based on the three lactation stages, the distinction between groups were
221 less clear. Clusters associated with early and middle lactation milk overlapped each other and were
222 separate from the cluster representing late lactation milk. This indicates that the former two groups were
223 more similar to each other than they were to the latter group (Figure 3b). The PERMANOVA test

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

224 revealed significant difference in the milk microbiota composition of the three lactation stages (P =
225 0.016). However, the effect size of lactation stage on milk microbiota was smaller than that of mastitis
226 status. It is worth noting that most late lactation dairy cows were healthy.

227 Wilcoxon and Kruskal-Wallis tests were done to identify differentially abundant taxa in the different
228 groups. Species that were significantly differentially abundant (average relative abundance > 0.5%) are
229 listed in Table S2 and S3. Significantly more *La. lactis*, *A. johnsonii* and *B. dorei* were found in the milk
230 of healthy ($P < 0.01$) compared with mastitis cows, while significantly more *B. cereus* and *S. suis* were
231 found in the milk of mastitis compared with healthy cows ($P < 0.05$). Meanwhile, significantly more *La.*
232 *lactis*, *A. johnsonii*, *B. dorei*, *S. stercoricanis* and *S. suis* were present in the milk of late lactation cows
233 compared with other lactation stages ($P < 0.05$); the two latter species were found only during late
234 lactation.

235

236 **3.4 Correlation analysis of microbiota community structure in milk from cows**

237 Spearman correlation analyses were done to identify co-occurrence relationships amongst the major
238 bacterial genera (an average relative abundance > 0.5%) in milk from healthy and mastitis cows. The
239 results are expressed as correlation network diagrams. The milk microbiota of the healthy group seemed
240 to be closely interrelated, contrasting with the overall weak networking apparent amongst bacterial
241 genera in milk from mastitis cows. In milk from healthy cows, *Lactococcus* was significantly and
242 negatively correlated with *Acinetobacter* ($r = -0.78$; $p = 0.01$) and *Massilia* ($r = -0.70$; $p = 0.04$, Figure
243 3c), respectively. It is interesting to note that the *Faecalibacterium* formed the highest number of

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

244 significant correlations with other genera from milk. In contrast, *Faecalibacterium* didn't correlate with
245 any genera in the milk from mastitis cows (Figure 3d).
246

247 **3.5 Quantification of microbial in cows milk by ddPCR and construction of Udder Health Index**

248 To verify the sequencing results, ddPCR was done to quantify the number of *La. lactis* and *B. cereus*, as
249 they were the most abundant species in the healthy and mastitis groups, respectively. For the healthy
250 group, there were significantly more *La. lactis* (2.16×10^3 copies/ml) compared with *B. cereus* ($P < 0.01$),
251 and *B. cereus* was only detected in the sample C17 (healthy group). An opposite trend was observed for
252 the mastitis group, with a significantly higher abundance of *B. cereus* (3.25×10^4 copies/ml) compared
253 with *La. lactis* (9.00×10^2 copies/mL; $P < 0.001$) (Figure 4a). These results were consistent with those
254 found by DNA sequencing.
255 The Random Forest regression model was applied to predict the udder health of cows. The relative
256 abundances of bacterial species detected in milk were regressed against each cow's health status. The top
257 nine mastitis-discriminatory marker species were selected based on the minimum 'CV error', and an
258 Udder Health Index was constructed (Figure 4b,d). Udder Health Indices for the healthy and mastitis
259 groups were calculated; a higher value represented a healthier status. As expected, the Udder Health
260 Index was significantly higher for the healthy compared with the mastitis group ($P < 0.001$, Figure 4c).
261 To verify the accuracy of the constructed model, the SCC of the healthy milk samples was measured
262 (Table 1). But this was not possible for the mastitis group due to the milk flocs were obvious and the
263 granules were large, accompanied by blood and different degree of coagulation. However, the SCC
264 associated well with the predicted Udder Health Index of the healthy cows, validating the current model.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

265

266 **4 DISCUSSION**

267 Mastitis is a common disease among dairy cows; symptoms include shortening of lactation period and
268 milk production, increase in leukocytes in the milk, and lesions in mammary tissues (Paster, Dewhirst,
269 Olsen, & Fraser, 1994). Since bovine mastitis results in huge economical losses, its prevention and
270 treatment have attracted wide attention. Bovine mastitis can also increase the microbial load in milk
271 resulting in rapid changes in the quality and shelf life of raw milk and related products (Murphy, Martin,
272 Barbano, & Wiedmann, 2016). Pathogenic bacteria in raw milk from mastitis cows are a serious food
273 safety issue as they may lead to human disease. In addition, clinical mastitis is a serious animal welfare
274 concern because mastitis is debilitating and painful (Fromm & Boor, 2004; Halasa et al., 2007). Milk is
275 produced by the mammary tissues of lactating cows; the appearance, texture, and quantity of secreted
276 cow milk are indicative of the health of the mammary tissue and thus serve as indicators for clinical
277 mastitis. This study used PacBio SMRT sequencing technology to reveal the community composition of
278 microbiota in milk from dairy cows, and identified differences in the milk microbiota of healthy and
279 mastitis cows, and cows at different stages of lactation. We observed a significantly higher microbial
280 diversity and richness in milk from healthy compared with mastitis cows, as reported in other
281 studies (Braem et al., 2012; Kuehn et al., 2013). Our data also showed that the microbial diversity and
282 richness of milk also increased significantly during late lactation. On the other hand, the increase in
283 microbial diversity could indicate a healthy state of the cows, as the mastitis milk might be dominated by
284 certain pathogens, which would be reflected by the diminished microbiota diversity. Moreover, changes
285 in milk microbiota at different lactation stages could be associated with changes in the nutritional content

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

286 of the milk, e.g. the total concentration of milk oligosaccharides was found to decrease in the early and
287 mid lactation stages. The anionic oligosaccharides including N-glycolylneuraminic acid decreased more
288 rapidly than the neutral oligosaccharides in lactation (Nakamura et al., 2003; Tao, DePeters, German,
289 Grimm, & Lebrilla, 2009). Therefore, when cows suffer from mastitis, the mammary tissues may be
290 overwhelmed with harmful bacteria; this could induce localized immune responses that suppress the
291 healthy resident microbiota and reduce microbial diversity and richness in the mammary gland (Kuang et
292 al., 2009).

293 We found that the distribution of bacterial taxa present in milk also varied greatly between healthy
294 and mastitis cows, as found in other studies (Falentin et al., 2016; Oikonomou, Machado, Santisteban,
295 Schukken, & Bicalho, 2012). The predominant genera detected in this work included *Lactobacillus*,
296 *Streptococcus*, *Acinetobacter* and *Bacillus*, which are known to be common in milk (N. Li et al., 2018).
297 We found significantly more *Lactobacillus*, *Lactococcus* and *Acinetobacter* in milk from healthy
298 compared with mastitis cows. The relative abundance of *Lactobacillus* in milk is known to be negatively
299 associated with milk SCC, an indicator of the seriousness of mastitis; thus, these bacteria are crucial in
300 maintaining the health of the mammary tissues and suppressing local infection and inflammation (Yu,
301 Ren, Xi, Huang, & Zhang, 2017). In contrast, *Lactococcus* is known to be able to cause bovine
302 mastitis (Rodrigues, Lima, Higgins, Canniatti-Brazaca, & Bicalho, 2016). *Acinetobacter* is widely
303 distributed in nature and is often detected in milk, soil and water. It is also considered as an opportunistic
304 pathogen associated with wounds and skin infections (Dortet, Legrand, Soussy, & Cattoir, 2006; L. Li et
305 al., 2016). Results of the Spearman correlation analysis in our study indicated a negative correlation
306 between *Lactococcus* and *Acinetobacter*, suggesting that there may be a competition inhibition

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

307 relationship between them. *Bacteroides* are typical milk bacteria that may also have a role in maintaining
308 healthy mammary tissues (Quigley et al., 2013). Differences in the correlation patterns between the milk
309 microbiota from healthy and mastitis cows may suggest an important role of milk microbiota in
310 protecting cows from mastitis. Our data revealed an increase in the relative abundances of *Bacillus*,
311 *Clostridium* and *Streptococcus* in milk from mastitis compared with healthy cows. *Staphylococcus* has
312 been reported as the most common cause of mastitis in cows worldwide, followed by
313 *Streptococcus* (Moroni et al., 2006). *Clostridium* species have been reported to cause abscesses in
314 mammary tissues of sows and humans as well as gangrene in cows (Durojaiye, Gaur, & Alsaffar, 2011;
315 Osman, El-Enbaawy, Ezzeldeen, & Hussein, 2009). The *Bacillus* genus includes some important
316 causative agents of mastitis, e.g. *B. cereus* (Parkinson, Merrall, & Fenwick, 1999); in this study *B. cereus*
317 increased in relative abundance (up to 27.55%) in milk from mastitis compared with healthy cows. By
318 ddPCR, we confirmed the elevated abundance of *Bacillus cereus* in the mastitis group (3.25×10^4
319 copies/mL); it was practically absent in the healthy group. These results suggest a possible link between
320 *B. cereus* and bovine mastitis.

321 Another spectrum of mastitis-associated bacterial sequences detected in this study was the anaerobes,
322 including *F. necrophorum* and *B. dorei*. Although these bacteria are unlikely to be causative agents of
323 clinical bovine mastitis, they are known to be associated with summer mastitis and may interact with
324 other pathogens such as *Trueperella pyogenes* (Oikonomou et al., 2012; Pyorala, Jousimies-Somer, &
325 Mero, 1992). Significantly more *P. heparinolytica* and *H. ovis* sequences were detected in milk from
326 cows at the early lactation stage in this study, both *P. heparinolytica* and *H. ovis* are potential pathogens
327 that have previously been isolated from mammary gland wounds causing localized infection (Paster et

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

328 al., 1994). It is suspected that the presence of *H. ovis* in cows milk indicates involvement in the
329 pathogenesis of mastitis (Schwaiger et al., 2012).The relative abundances of *La. lactis*, *A. johnsonii* and
330 *B. dorei* were higher during late lactation compared with the early and middle lactation stages. ***B.** dorei*
331 can suppress the production of lipopolysaccharides by intestinal gut microbes and thus reduces
332 pro-inflammatory immune responses (Yoshida et al., 2018). Furthermore, changes in susceptibility to
333 mastitis might be coupled with changes in community composition of the milk microbiota, such as
334 increases in *B. dorei* and *La. lactis* (Xu et al., 2017).

335 Cows suffering from subclinical mastitis can easily develop clinical mastitis if they are not spotted
336 early enough and managed appropriately. Therefore, it is important to predict the likely risk of mastitis
337 developing. Current practice defines subclinical mastitis using a cut-off SCC threshold level of 200 000
338 cells/mL. However, the determination of milk SCC is widely used to monitor udder health, but there are
339 still some limitations: (1) it may take a longer time for the SCC to return to that of a healthy state after
340 pathogen clearance, so there is a window period when the diagnosis based on SCC level would not be
341 accurate; (2) the SCC value fluctuates largely with individual milk yield and other environmental factors;
342 (3) diagnosis based purely on SCC is mainly applicable to mastitis cows caused by contagious pathogens,
343 while mastitis caused by environmental factors would be hard to detect due to the relatively small
344 changes in SCC (Pyorala et al., 1992; Sharma, Singh, & Bhadwal, 2011). In this study we used the
345 Random Forest model identified the 9 top marker species that were indicative of bovine mastitis, and
346 constructed an Udder Health Index based on the relative abundances of these species. We validated the
347 model by associating the index with the SCC of healthy cows. Our results showed that the accuracy of the
348 model was 88.89%. Thus, this index would be a good complementary indicator that helps detect early

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

349 changes in cow health. One limitation of the current model is the small sample size, therefore, larger
350 scale future works will be necessary to optimize and verify this model.
351
352 **5 CONCLUSION**
353 The results of this study showed great variation in the microbiota in milk from healthy and mastitis cows.
354 Overall there was a high relative abundance of commensals in healthy milk and a high relative abundance
355 of potential pathogens in milk from mastitis cows. Finally, the Udder Health Index constructed is helpful
356 for early identification of mastitis risk and has the potential to be used to prevent mastitis from developing,
357 and prompt measures should be taken to prevent further development of clinical mastitis.
358
359 **ACKNOWLEDGEMENTS**
360 This research was supported by the China Agriculture Research System (Grant CARS-36) of the Inner
361 Mongolia Science & Technology Projects. The authors wish to sincerely thank Mr. Yang Ku for allowing
362 the collection of milk samples from the cows on his farm. We also thank Dr. Lai Yu-Kwok for their
363 valuable revision of the English text.
364

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

365 **REFERENCES**

366 Anderson, M. J., & Walsh, D. C. I. (2013). PERMANOVA, ANOSIM, and the Mantel test in the face of
367 heterogeneous dispersions: what null hypothesis are you testing? *Ecological Monographs, 83*,
368 557-574. https://doi.org/10.1890/12-2010.1
369 Braem, G., De Vliegher, S., Verbist, B., Heyndrickx, M., Leroy, F., & De Vuyst, L. (2012).
370 Culture-independent exploration of the teat apex microbiota of dairy cows reveals a wide
371 bacterial species diversity. *Veterinary Microbiology, 157*, 383-390.
372 https://doi.org/10.1016/j.vetmic.2011.12.031
373 Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., . . . Knight, R.
374 (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods,*
375 *7*, 335-336. https://doi.org/10.1038/nmeth.f.303
376 Carolina Espeche, M., Pellegrino, M., Frola, I., Larriestra, A., Bogni, C., & Fatima Nader-Macias, M. E.
377 (2012). Lactic acid bacteria from raw milk as potentially beneficial strains to prevent bovine
378 mastitis. *Anaerobe, 18*, 103-109. https://doi.org/10.1016/j.anaerobe.2012.01.002
379 Cole, J. R., Chai, B., Farris, R. J., Wang, Q., Kulam-Syed-Mohideen, A. S., McGarrell, D. M., . . . Tiedje,
380 J. M. (2007). The ribosomal database project (RDP-II): introducing myRDP space and quality
381 controlled public data. *Nucleic Acids Research, 35*, D169-D172.
382 https://doi.org/10.1093/nar/gkl889
383 Cremonesi, P., Cortimiglia, C., Picozzi, C., Minozzi, G., Malvisi, M., Luini, M., & Castiglioni, B. (2016).
384 Development of a droplet digital polymerase chain reaction for rapid and simultaneous
385 identification of common food borne pathogens in soft cheese. *Frontiers in Microbiology, 7,*
386 1725. https://doi.org/10.3389/fmicb.2016.01725
387 Cutler, D. R., Edwards, T. C., Jr., Beard, K. H., Cutler, A., & Hess, K. T. (2007). Random forests for
388 classification in ecology. *Ecology, 88*, 2783-2792. https://doi.org/10.1890/07-0539.1
389 DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., . . . Andersen, G. L.
390 (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible
391 with ARB. *Applied and Environmental Microbiology, 72*, 5069-5072.
392 https://doi.org/10.1128/aem.03006-05
393 Dohoo, I. R., Smith, J., Andersen, S., Kelton, D. F., Godden, S., & Mastitis Res Workers, C. (2011).
394 Diagnosing intramammary infections: evaluation of definitions based on a single milk sample.
395 *Journal of Dairy Science, 94*, 250-261. https://doi.org/10.3168/jds.2010-3559
396 Dortet, L., Legrand, P., Soussy, C.-J., & Cattoir, V. (2006). Bacterial identification, clinical significance,
397 and antimicrobial susceptibilities of *Acinetobacter ursingii* and *Acinetobacter schindleri*, two
398 frequently misidentified opportunistic pathogens. *Journal of Clinical Microbiology, 44*,
399 4471-4478. https://doi.org/10.1128/jcm.01535-06
400 Durojaiye, O., Gaur, S., & Alsaffar, L. (2011). Bacteraemia and breast abscess: unusual extra-intestinal
401 manifestations of *Clostridium difficile* infection. *Journal of Medical Microbiology, 60*, 378-380.
402 https://doi.org/10.1099/jmm.0.027409-0
403 Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics, 26*,
404 2460-2461. https://doi.org/10.1093/bioinformatics/btq461
405 Falentin, H., Rault, L., Nicolas, A., Bouchard, D. S., Lassalas, J., Lamberton, P., . . . Even, S. (2016).
406 Bovine teat microbiome analysis revealed reduced alpha diversity and significant changes in

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

407 taxonomic profiles in quarters with a history of mastitis. *Frontiers in Microbiology, 7*, 480.
408 https://doi.org/10.3389/fmicb.2016.00480
409 Francisco Martinez-Blanch, J., Sanchez, G., Garay, E., & Aznar, R. (2011). Detection and quantification of
410 viable *Bacillus cereus* in food by RT-qPCR. *European Food Research and Technology, 232*,
411 951-955. https://doi.org/10.1007/s00217-011-1465-1
412 Fromm, H. I., & Boor, K. J. (2004). Characterization of pasteurized fluid milk shelf-life attributes. *Journal*
413 *of Food Science, 69*, M207-M214. https://doi.org/10.1111/j.1365-2621.2004.tb09889.x
414 Ginestet, C. (2011). ggplot2: elegant graphics for data analysis. *Journal of the Royal Statistical Society*
415 *Series a-Statistics in Society, 174*, 245-245. https://doi.org/10.1111/j.1467-985X.2010.00676_9.x
416 Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., . . . Human
417 Microbiome, C. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and
418 454-pyrosequenced PCR amplicons. *Genome Research, 21*, 494-504.
419 https://doi.org/10.1101/gr.112730.110
420 Halasa, T., Huijps, K., Osteras, O., & Hogeveen, H. (2007). Economic effects of bovine mastitis and
421 mastitis management: a review. *Veterinary Quarterly, 29*, 18-31.
422 https://doi.org/10.1080/01652176.2007.9695224
423 Haltia, L., Honkanen-Buzalski, T., Spiridonova, I., Olkonen, A., & Myllys, V. (2006). A study of bovine
424 mastitis, milking procedures and management practices on 25 Estonian dairy herds. *Acta*
425 *Veterinaria Scandinavica, 48*, 22. https://doi.org/10.1186/1751-0147-48-22
426 Hou, Q., Xu, H., Zheng, Y., Xi, X., Kwok, L.-Y., Sun, Z., . . . Zhang, W. (2015). Evaluation of bacterial
427 contamination in raw milk, ultra-high temperature milk and infant formula using single molecule,
428 real-time sequencing technology. *Journal of Dairy Science, 98*, 8464-8472.
429 https://doi.org/10.3168/jds.2015-9886
430 Hui, W., Hou, Q., Cao, C., Xu, H., Zhen, Y., Kwok, L.-Y., . . . Zhang, W. (2017). Identification of
431 microbial profile of Koji using single molecule, real-time sequencing technology. *Journal of*
432 *Food Science, 82*, 1193-1199. https://doi.org/10.1111/1750-3841.13699
433 Kuang, Y., Tani, K., Synnott, A. J., Ohshima, K., Higuchi, H., Nagahata, H., & Tanji, Y. (2009).
434 Characterization of bacterial population of raw milk from bovine mastitis by culture-independent
435 PCR-DGGE method. *Biochemical Engineering Journal, 45*, 76-81.
436 https://doi.org/10.1016/j.bej.2009.02.010
437 Kuehn, J. S., Gorden, P. J., Munro, D., Rong, R., Dong, Q., Plummer, P. J., . . . Phillips, G. J. (2013).
438 Bacterial community profiling of milk samples as a means to understand culture-negative bovine
439 clinical mastitis. *PloS One, 8*, e61959. https://doi.org/10.1371/journal.pone.0061959
440 Li, L., Renye, J. A., Jr., Feng, L., Zeng, Q., Tang, Y., Huang, L., . . . Yang, P. (2016). Characterization of
441 the indigenous microflora in raw and pasteurized buffalo milk during storage at refrigeration
442 temperature by high-throughput sequencing. *Journal of Dairy Science, 99*, 7016-7024.
443 https://doi.org/10.3168/jds.2016-11041
444 Li, N., Wang, Y., You, C., Ren, J., Chen, W., Zheng, H., & Liu, Z. (2018). Variation in raw milk
445 microbiota throughout 12 months and the impact of weather conditions. *Scientific Reports, 8*,
446 1-10. https://doi.org/10.1038/s41598-018-20862-8
447 Liao, Y.-C., Lin, S.-H., & Lin, H.-H. (2015). Completing bacterial genome assemblies: strategy and
448 performance comparisons. *Scientific Reports, 5*, 8747. https://doi.org/10.1038/srep08747

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

449 Ma, C., Sun, Z., Zeng, B., Huang, S., Zhao, J., Zhang, Y., . . . Zhang, H. (2018). Cow-to-mouse fecal
450 transplantations suggest intestinal microbiome as one cause of mastitis. *Microbiome, 6*, 200.
451 https://doi.org/10.1186/s40168-018-0578-1
452 Metzger, S. A., Hernandez, L. L., Skarlupka, J. H., Walker, T. M., Suen, G., & Ruegg, P. L. (2018). A
453 cohort study of the milk microbiota of healthy and inflamed bovine mammary glands from dryoff
454 through 150 days in milk. *Frontiers in Veterinary Science, 5*, 247.
455 https://doi.org/10.3389/fvets.2018.00247
456 Moroni, P., Rossi, C. S., Pisoni, G., Bronzo, V., Castiglioni, B., & Boettcher, P. J. (2006). Relationships
457 between somatic cell count and intramammary infection in buffaloes. *Journal of Dairy Science,*
458 *89*, 998-1003. https://doi.org/10.3168/jds.S0022-0302(06)72165-8
459 Mosher, J. J., Bernberg, E. L., Shevchenko, O., Kan, J., & Kaplan, L. A. (2013). Efficacy of a 3rd
460 generation high-throughput sequencing platform for analyses of 16S rRNA genes from
461 environmental samples. *Journal of Microbiological Methods, 95*, 175-181.
462 https://doi.org/10.1016/j.mimet.2013.08.009
463 Mosher, J. J., Bowman, B., Bernberg, E. L., Shevchenko, O., Kan, J., Korlach, J., & Kaplan, L. A. (2014).
464 Improved performance of the PacSio SMRT technology for 16S rDNA sequencing. *Journal of*
465 *Microbiological Methods, 104*, 59-60. https://doi.org/10.1016/j.mimet.2014.06.012
466 Murphy, S. C., Martin, N. H., Barbano, D. M., & Wiedmann, M. (2016). Influence of raw milk quality on
467 processed dairy products: how do raw milk quality test results relate to product quality and yield?
468 *Journal of Dairy Science, 99*, 10128-10149. https://doi.org/10.3168/jds.2016-11172
469 Nakamura, T., Kawase, H., Kimura, K., Watanabe, Y., Ohtani, M., Arai, I., & Urashima, T. (2003).
470 Concentrations of sialyloligosaccharides in bovine colostrum and milk during the prepartum and
471 early lactation. *Journal of Dairy Science, 86*, 1315-1320.
472 https://doi.org/10.3168/jds.S0022-0302(03)73715-1
473 Nakano, K., Shiroma, A., Tamotsu, H., Ohki, S., Shimoji, M., Ashimine, N., . . . Hirano, T. (2016). First
474 complete genome sequence of the skin-improving *Lactobacillus curvatus* Strain FBA2, isolated
475 from fermented vegetables, determined by PacBio single-molecule real-time technology. *Genome*
476 *announcements, 4*, e00884-e008816. https://doi.org/10.1128/genomeA.00884-16
477 Nilsson, R. H., Veldre, V., Hartmann, M., Unterseher, M., Amend, A., Bergsten, J., . . . Abarenkov, K.
478 (2010). An open source software package for automated extraction of ITS1 and ITS2 from fungal
479 ITS sequences for use in high-throughput community assays and molecular ecology. *Fungal*
480 *Ecology, 3*, 284-287. https://doi.org/10.1016/j.funeco.2010.05.002
481 Oikonomou, G., Machado, V. S., Santisteban, C., Schukken, Y. H., & Bicalho, R. C. (2012). Microbial
482 diversity of bovine mastitis milk as described by pyrosequencing of metagenomic 16s rDNA.
483 *PloS One, 7*, e47671. https://doi.org/10.1371/journal.pone.0047671
484 Osman, K. M., El-Enbaawy, M. I., Ezzeldeen, N. A., & Hussein, H. M. G. (2009). Mastitis in dairy buffalo
485 and cattle in Egypt due to *Clostridium perfringens*: prevalence, incidence, risk factors and costs.
486 *Revue Scientifique Et Technique-Office International Des Epizooties, 28*, 975-986.
487 https://doi.org/10.20506/rst.28.3.1936
488 Pantoja, J. C. F., Hulland, C., & Ruegg, P. L. (2009). Somatic cell count status across the dry period as a
489 risk factor for the development of clinical mastitis in the subsequent lactation. *Journal of Dairy*
490 *Science, 92*, 139-148. https://doi.org/10.3168/jds.2008-1477

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

491 Parkinson, T. J., Merrall, M., & Fenwick, S. G. (1999). A case of bovine mastitis caused by *Bacillus*
492 *cereus*. *New Zealand Veterinary Journal, 47*, 151-152.
493 https://doi.org/10.1080/00480169.1999.36134
494 Paster, B. J., Dewhirst, F. E., Olsen, I., & Fraser, G. J. (1994). Phylogeny of *Bacteroides*, *Prevotella*, and
495 *Porphyromonas* spp. and related bacteria. *Journal of Bacteriology, 176*, 725-732.
496 https://doi.org/10.1128/jb.176.3.725-732.1994
497 Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with
498 profiles instead of a distance matrix. *Molecular Biology and Evolution, 26*, 1641-1650.
499 https://doi.org/10.1093/molbev/msp077
500 Pyorala, S., Jousimies-Somer, H., & Mero, M. (1992). Clinical, bacteriological and therapeutic aspects of
501 bovine mastitis caused by aerobic and anaerobic pathogens. *The British veterinary journal, 148*,
502 54-62.
503 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., . . . Gloeckner, F. O. (2013). The
504 SILVA ribosomal RNA gene database project: improved data processing and web-based tools.
505 *Nucleic Acids Research, 41*, D590-D596. https://doi.org/10.1093/nar/gks1219
506 Quigley, L., O'Sullivan, O., Stanton, C., Beresford, T. P., Ross, R. P., Fitzgerald, G. F., & Cotter, P. D.
507 (2013). The complex microbiota of raw milk. *FEMS Microbiology Reviews, 37*, 664-698.
508 https://doi.org/10.1111/1574-6976.12030
509 Rasmussen, D. B., Fogsgaard, K., Rontved, C. M., Klaas, I. C., & Herskin, M. S. (2011). Changes in
510 thermal nociceptive responses in dairy cows following experimentally induced *Escherichia coli*
511 mastitis. *Acta Veterinaria Scandinavica, 53*, 32. https://doi.org/10.1186/1751-0147-53-32
512 Rodrigues, M. X., Lima, S. F., Higgins, C. H., Canniatti-Brazaca, S. G., & Bicalho, R. C. (2016). The
513 *Lactococcus* genus as a potential emerging mastitis pathogen group: a report on an outbreak
514 investigation. *Journal of Dairy Science, 99*, 9864-9874. https://doi.org/10.3168/jds.2016-11143
515 Schwaiger, K., Wimmer, M., Huber-Schlenstedt, R., Fehlings, K., Hoelzel, C. S., & Bauer, J. (2012). Hot
516 topic: bovine milk samples yielding negative or nonspecific results in bacterial culturing-The
517 possible role of PCR-single strand conformation polymorphism in mastitis diagnosis. *Journal of*
518 *Dairy Science, 95*, 98-101. https://doi.org/10.3168/jds.2011-4700
519 Sharma, N., Singh, N. K., & Bhadwal, M. S. (2011). Relationship of somatic cell count and mastitis: an
520 overview. *Asian-Australasian Journal of Animal Sciences, 24*, 429-438.
521 https://doi.org/10.5713/ajas.2011.10233
522 Tao, N., DePeters, E. J., German, J. B., Grimm, R., & Lebrilla, C. B. (2009). Variations in bovine milk
523 oligosaccharides during early and middle lactation stages analyzed by high-performance liquid
524 chromatography-chip/mass spectrometry. *Journal of Dairy Science, 92*, 2991-3001.
525 https://doi.org/10.3168/jds.2008-1642
526 Taponen, S., Salmikivi, L., Simojoki, H., Koskinen, M. T., & Pyorala, S. (2009). Real-time polymerase
527 chain reaction-based identification of bacteria in milk samples from bovine clinical mastitis with
528 no growth in conventional culturing. *Journal of Dairy Science, 92*, 2610-2617.
529 https://doi.org/10.3168/jds.2008-1729
530 Vijayakumar, M., Park, J. H., Ki, K. S., Lim, D. H., Kim, S. B., Park, S. M., . . . Kim, T. I. (2017). The
531 effect of lactation number, stage, length, and milking frequency on milk yield in Korean Holstein
532 dairy cows using automatic milking system. *Asian-Australasian Journal of Animal Sciences, 30*,
533 1093-1098. https://doi.org/10.5713/ajas.16.0882

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

534 Xu, H., Huang, W., Hou, Q., Kwok, L.-y., Sun, Z., Ma, H., . . . Zhang, H. (2017). The effects of probiotics
535 administration on the milk production, milk components and fecal bacteria microbiota of dairy
536 cows. *Science Bulletin, 62*, 767-774. https://doi.org/10.1016/j.scib.2017.04.019
537 Yoshida, N., Emoto, T., Yamashita, T., Watanabe, H., Hayashi, T., Tabata, T., . . . Hirata, K.-i. (2018).
538 *Bacteroides vulgatus* and *Bacteroides dorei* reduce gut microbial lipopolysaccharide production
539 and inhibit atherosclerosis. *Circulation, 138*, 2486-2498.
540 https://doi.org/10.1161/circulationaha.118.033714
541 Yu, J., Ren, Y., Xi, X., Huang, W., & Zhang, H. (2017). A novel Lactobacilli-based teat disinfectant for
542 improving bacterial communities in the milks of cow teats with subclinical mastitis. *Frontiers in*
543 *Microbiology, 8*, 1782. https://doi.org/10.3389/fmicb.2017.01782

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

544 **TABLES**

545 **Table 1** Udder Health Index, somatic cell count (SCC), and total bacterial count of cow milk samples

Cow milk sample	Udder Health Index	SCC (x10^4 cells/mL)	Total bacteria count (x10^4 CFU/mL)
C2	0.142	12.2	0.863
C4	0.253	14.8	0.948
C14	0.846	13.8	0.882
C13	0.849	13.3	1.050
C18	1.334	12.9	0.819
C17	2.180	12.0	0.796
C15	4.062	11.1	0.859
C20	5.545	10.7	0.671
C12	9.753	9.9	0.698

546

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

547 **FIGURE LEGENDS**

548 **Figure 1** Comparison of the number of observed OTUs and Shannon diversity index for the microbiota
549 in milk. The healthy and mastitis cows (a, b); early, middle and late lactation (c, d). Healthy and mastitis
550 cows are represented by 'HC' and 'MC', respectively. Early, middle, and late lactation stages are
551 represented by 'EL', 'ML' and 'LL', respectively. Single and triple asterisks represent $P < 0.05$ and $P <$
552 0.001, respectively.

553

554 **Figure 2** Community composition of the bacterial microbiota in milk from healthy and mastitis cows.
555 Milk bacterial community at the genus (a and b) and species (c and d) levels. Healthy and mastitis cows
556 are represented by 'HC' and 'MC', respectively.

557

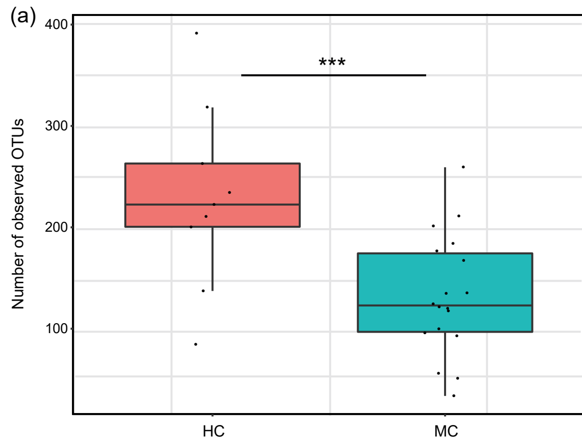
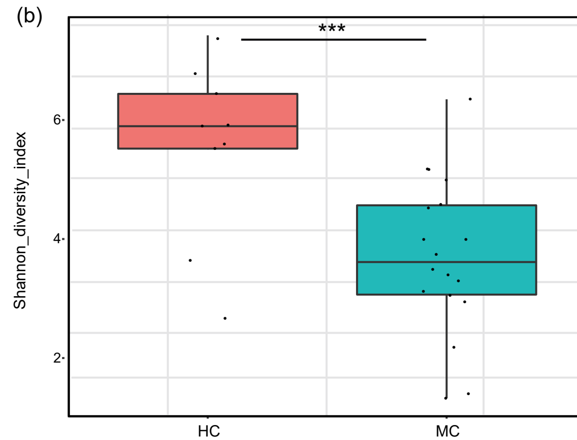
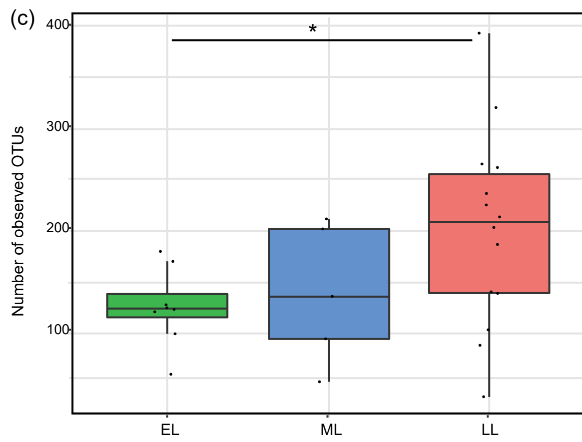
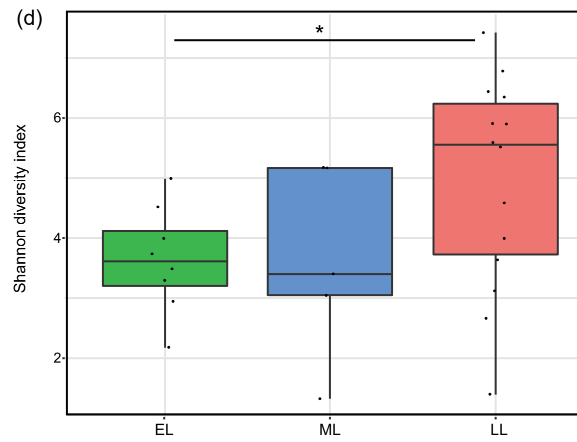
558 **Figure 3** Principal coordinates analysis (PCoA) and correlation networks of the bacterial microbiota in
559 milk from healthy and mastitis cows. They are grouped based on mastitis status (a) and lactation stage
560 (b). The co-occurrence relationship was calculated using Spearman's rank correlation analysis of the
561 microbiota in milk from healthy (c) and mastitis (d) cows. Only major genera of average relative
562 abundance > 0.5% are included in the analysis. The diameter of the circles represents the abundance of
563 the genera. Healthy and mastitis cows are represented by 'HC' and 'MC', respectively. Early, middle, and
564 late lactation stages are represented by 'EL', 'ML' and 'LL', respectively.

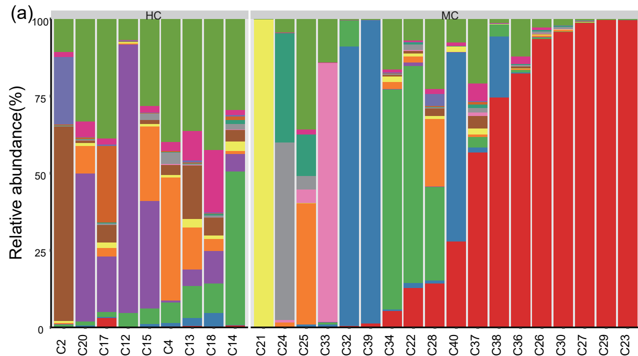
565

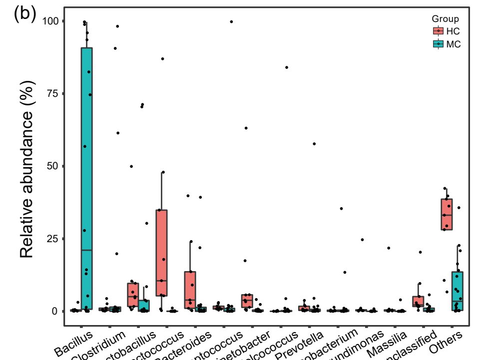
566 **Figure 4** Quantification of *La. lactis* and *B. cereus* in cows' milk by ddPCR and construction of Udder
567 Health Index. (a) The prevalence of *La. lactis* and *B. cereus* in milk from healthy and mastitis cows. (b)

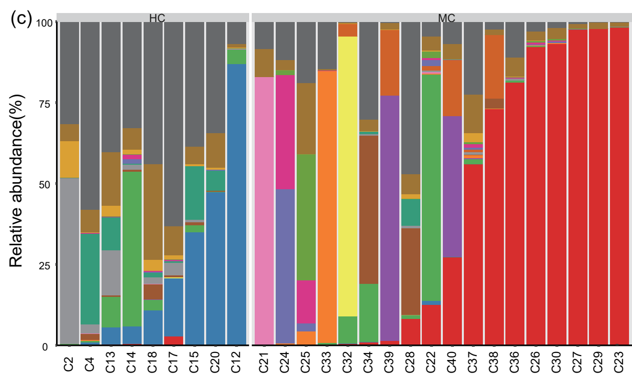
bioRxiv preprint doi: https://doi.org/10.1101/2020.08.13.249524; this version posted August 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

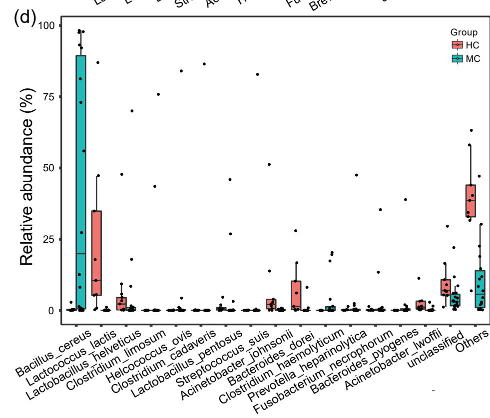
568 Model marker species selected based on minimum CV error. (c) The nine selected marker species of the
569 model. (d) Udder Health Indices in the healthy and mastitis groups. Healthy and mastitis cows are
570 represented by 'HC' and 'MC', respectively. Single and triple asterisks represent $P < 0.05$ and $P < 0.001$,
571 respectively.
572
573

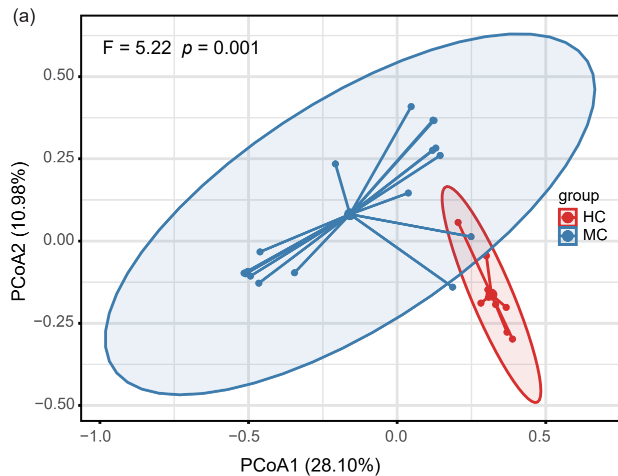

(a)
Number of observed OTUs
400
300
200
100
***
HC
MC

(b)
Shannon_diversity_index
6
4
2
***
HC
MC

(c)
Number of observed OTUs
400
300
200
100
*
EL
ML
LL

(d)
Shannon diversity index
6
4
2
*
EL
ML
LL

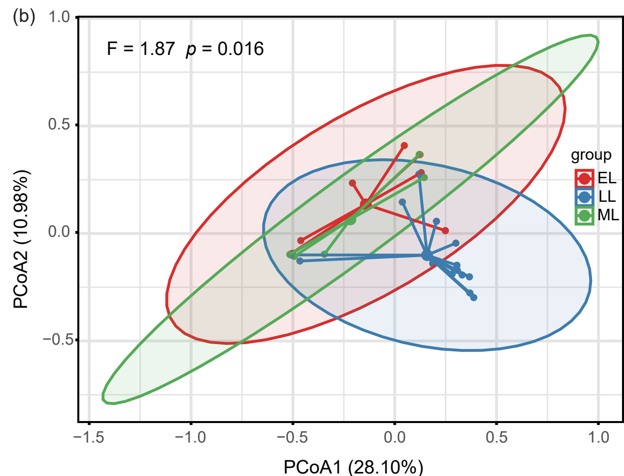

(a)
HC
MC
Relative abundance(%)
Genus
other
unclassified
Massilia
Brevundimonas
Fusobacterium
Prevotella
Helcococcus
Acinetobacter
Streptococcus
Bacteroides
Lactococcus
Lactobacillus
Clostridium
Bacillus


(b)
Group
HC
MC
Relative abundance (%)
Bacillus
Clostridium
Lactobacillus
Lactococcus
Bacteroides
Streptococcus
Acinetobacter
Helcococcus
Prevotella
Fusobacterium
Brevundimonas
Massilia
unclassified
Others

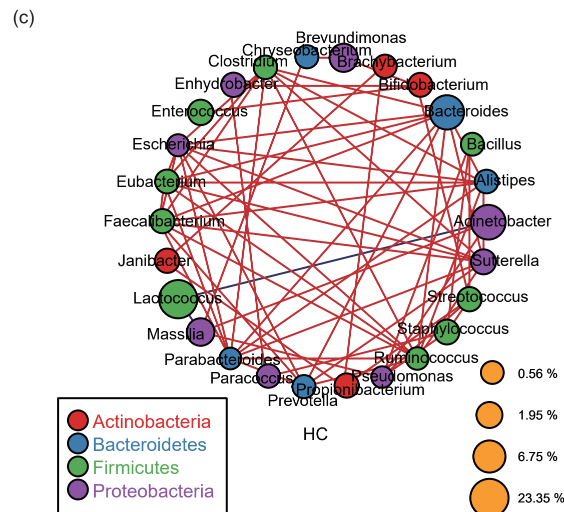

(c)
HC
MC
Relative abundance(%)
Species
other
unclassified
Acinetobacter_lwoffii
Bacteroides_pyogenes
Fusobacterium_necrophorum
Prevotella_heparinolytica
Clostridium_haemolyticum
Bacteroides_dorei
Acinetobacter_johnsonii
Streptococcus_suis
Lactobacillus_pentosus
Clostridium_cadaveris
Helcococcus_ovis
Clostridium_limosum
Lactobacillus_helveticus
Lactococcus_lactis
Bacillus_cereus

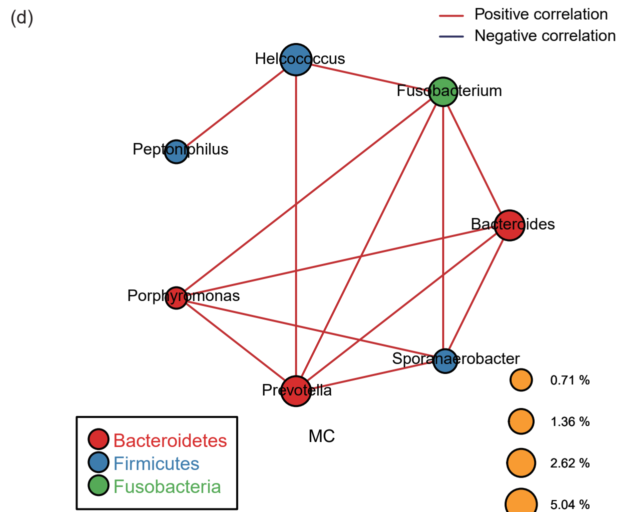

(d)
Group
HC
MC
Relative abundance (%)
Bacillus_cereus
Lactococcus_lactis
Lactobacillus_helveticus
Clostridium_limosum
Helcococcus_ovis
Clostridium_cadaveris
Lactobacillus_pentosus
Streptococcus_suis
Acinetobacter_johnsonii
Bacteroides_dorei
Clostridium_haemolyticum
Prevotella_heparinolytica
Fusobacterium_necrophorum
Bacteroides_pyogenes
Acinetobacter_lwoffii
unclassified
Others

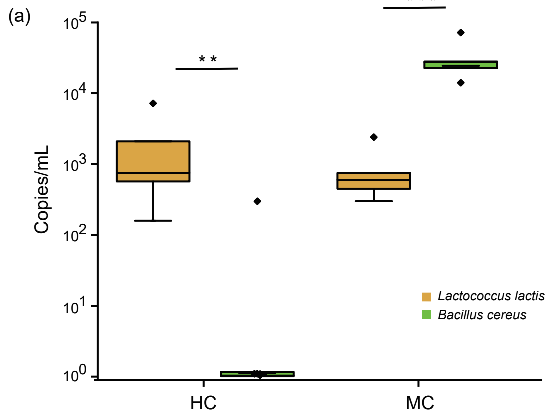


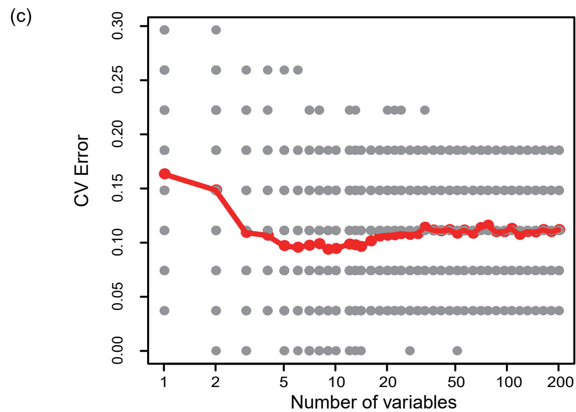


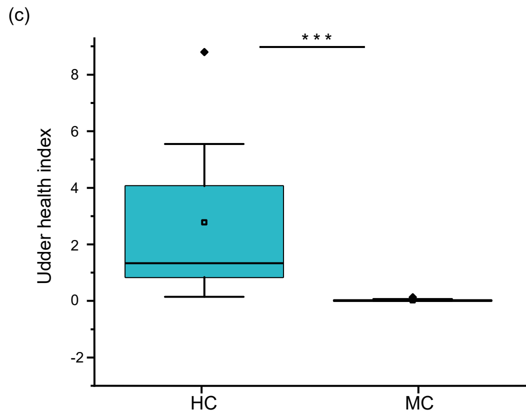
Correlation network

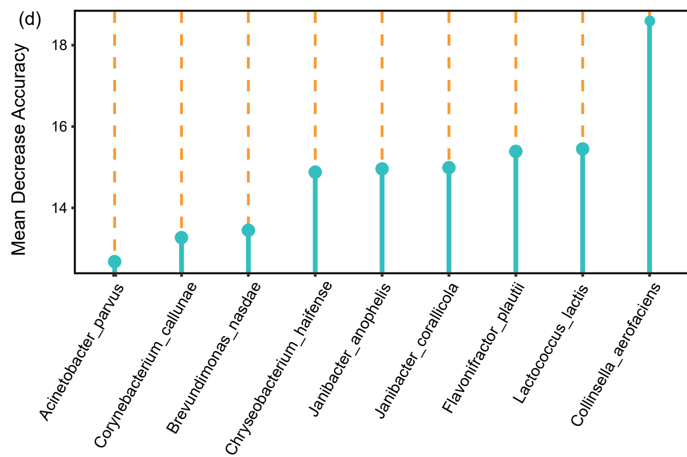





(a)
Copies/mL
10^5
10^4
10^3
10^2
10^1
10^0
* *
* * *
Lactococcus lactis
Bacillus cereus
HC
MC


(c)
CV Error
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1
2
5
10
20
50
100
200
Number of variables


(c)
Udder health index
8
6
4
2
0
-2
* * *
HC
MC


(d)
Mean Decrease Accuracy
18
16
14
Acinetobacter_parvus
Corynebacterium_callunae
Brevundimonas_nasdae
Chryseobacterium_haifense
Janibacter_anophelis
Janibacter_corallicola
Flavonifractor_plautii
Lactococcus_lactis
Collinsella_aerofaciens