Targeted syndromic next-generation sequencing panel for simultaneous detection of pathogens associated with bovine reproductive failure

Dhinesh Periyasamy[1,2], Yanyun Huang[2], Janet E. Hill[1,*]

[1]Department of Veterinary Microbiology, University of Saskatchewan, Saskatoon, Canada

[2]Prairie Diagnostic Services, Inc., Saskatoon, Canada

*Correspondence: Janet.Hill@usask.ca

Abstract

Bovine reproductive failure, which includes infertility, abortion, and stillbirth in cattle, leads to significant economic losses for beef and milk producers. Diagnosing the infectious causes of bovine reproductive failure is challenging as there are multiple pathogens associated with it. The traditional stepwise approach to diagnostic testing is time-consuming and can cause significant delays. In this study, we have developed a syndromic next-generation sequencing panel (BovReproSeq), for the simultaneous detection of 17 pathogens associated with bovine reproductive failure. This targeted approach involves amplifying multiple pathogen-specific targets using ultra-multiplex PCR, followed by sequencing with the Oxford Nanopore platform and subsequent analysis of the data using a custom bioinformatic pipeline to determine the presence or absence of pathogens. We tested 116 clinical samples and found that BovReproSeq results matched with current diagnostic methods for 93% of the samples, and most of the disagreements occurring in samples with very low pathogen loads (Ct > 35). At the optimal read-count threshold of 10 reads (minimum number of reads to classify the sample as positive), the sensitivity of the assay was approximately 82%, while specificity was 100%. The overall accuracy of the assay was 98.8%. Matthew's Correlation Coefficient was approximately 0.90 and F1 score (harmonic mean of Precision and Recall) was 0.90, indicating excellent overall performance. Our study presents a significant advancement in detecting the infectious agents associated with bovine reproductive failure and the BovReproSeq panel's ability to detect 17 pathogens makes it a promising tool for veterinary diagnostics.

Introduction

Bovine reproductive failure is a collective term used to refer to infertility, abortion, and stillbirth in cattle. It causes substantial economic losses to beef and milk producers, estimated to be approximately $1 billion annually in the USA (1). Accurate diagnosis is essential to control and manage this condition. Although non-infectious causes, such as poor nutrition, breeding practices and management, account for most of bovine reproductive failure cases in Western Canada, infectious diseases also contribute to bovine reproductive failure (2). In extreme cases, infectious diseases can lead to abortion storms if not diagnosed correctly. The pathogens associated with

37 this condition also pose a concern to humans as some of them belong to risk group-3 or are
38 zoonotic.

39 Multiple pathogens are associated with bovine reproductive failure including viruses, bacteria,
40 fungi, and protozoa. Infections with some of these pathogens are asymptomatic or cause similar
41 clinical signs and differentiation between these pathogens requires a variety of laboratory tests.
42 There is no recent data available on the diagnostic success rate in Western Canada, however,
43 based on 338 aborted fetuses submitted to Prairie Diagnostic Services in 2010, 52% had no
44 diagnosis, 41% were attributed to an infectious disease, and 7% to a developmental anomaly (2).
45 The success rate for diagnosing bovine reproductive failure is typically low as there is a
46 significant delay between infection and the occurrence of abortion (3). The diagnostic success
47 rate also depends on the availability of appropriate samples (e.g., fresh fetus, placenta, amniotic
48 fluid etc.), and this is not always possible as aborted fetuses are often autolyzed (4).
49 Veterinarians and producers usually focus on testing the most likely cause and if the initial test
50 fails to confirm the suspected cause, subsequent tests are requested. This stepwise approach to
51 diagnostic testing is time-consuming and increases the risk of missing the correct diagnosis as
52 veterinarians and producers are less likely to follow through a diagnostic plan if the initial tests
53 are negative. A syndromic panel approach in which multiple pathogens are tested for
54 simultaneously, would increase the chance of successful diagnosis and significantly reduce the
55 turnaround time for diagnosis (5). Currently, there is no syndromic panel test available in Canada
56 for bovine reproductive pathogens.

57 PCR is the most preferred method for detection of reproductive failure pathogens due to its high
58 analytical sensitivity and specificity. Even though multiple targets can be detected using PCR,
59 there is a limit to the number of targets that can be included in multiplex assays (6).
60 Metagenomic next-generation sequencing (mNGS) can overcome this limitation by either
61 sequence-independent amplification of the genetic material in the sample or by directly
62 sequencing the genetic material, thereby providing comprehensive microbial analysis. The
63 sensitivity of this method, however, is lower compared to species specific real-time PCR due to
64 the abundance of host genetic material in the sample (7). Additionally, there is a high cost
65 associated with mNGS and the turnaround times are longer compared to real-time PCR. These

limitations can be overcome by combining PCR with high-throughput next-generation sequencing. With this targeted approach, multiple pathogen-specific targets are first amplified in an ultra-multiplex PCR and amplified fragments are then sequenced and analysed (8, 9). While there are several high-throughput sequencing platforms available, the Oxford Nanopore platform offers real-time sequence data and analysis, enabling rapid and timely diagnosis. Multiple samples can be processed in a single sequencing run and diagnostic costs can be significantly reduced. In addition to detecting pathogens of interest, targeted metagenomic sequencing offers the ability to determine strain types (*in silico* serotyping, virulence typing, multi-locus sequence typing (MLST), etc.) and identify variants, which could be useful in differentiating closely related pathogens such as *Campylobacter fetus* subsp. *fetus* and *Campylobacter fetus* subsp. *venerealis* (10). There is also potential to detect emerging or novel pathogens with this approach using random primers along with the targeted primers (11).

In this study, we have developed a syndromic next-generation sequencing panel (BovReproSeq) for simultaneous detection of 17 pathogens associated bovine reproductive failure. The BovReproSeq method was tested on clinical samples, and it successfully detected all 17 pathogens associated with bovine reproductive failure. Despite some limitations in detecting pathogens at very low pathogen loads, the overall performance was excellent with a Matthew's Correlation Coefficient of 0.90.

Materials and methods

Target pathogen selection

Seventeen pathogens including bacteria, protozoa and viruses were selected for inclusion in the panel (Table 1). Pathogens were selected based on the results of reproductive failure cases that were submitted to Prairie Diagnostic Services since 2014 and additional advice from veterinarians and diagnostic pathologists. Originally 12 pathogens were included and after successful results from initial experiments, five more pathogens were added to the panel.

92 Table 1: Pathogens in the BovReproSeq panels and current diagnostic test

Pathogen	Current diagnostic test
Bacteria	
Coxiella burnetti	Real-time PCR
Chlamydia abortus	Real-time PCR
Ureaplasma diversum	Real-time PCR
Campylobacter fetus. subsp. *fetus*	Real-time and conventional PCR
Campylobacter jejuni	Conventional PCR
Campylobacter fetus. subsp. *venerealis*	Real-time and conventional PCR
Leptospira spp.	Immunohistochemistry
Listeria monocytogenes	Bacterial culture
Trueperella pyogenes	Bacterial culture
Mycoplasma bovis	Real-time PCR
Bacillus licheniformis	Bacterial culture
Protozoa	
Neospora caninum	Immunohistochemistry, serology
Tritrichomonas fetus	Real-time PCR
Toxoplasma gondii	Real-time PCR
Sarcocystis spp.	Immunohistochemistry
Viruses	
Bovine alphaherpesvirus 1 (BoHV-1)	Real-time PCR, virus isolation
Bovine viral diarrhea virus (BVDv)	Real-time PCR, virus isolation

93

94 **PCR Primer design and *in silico* PCR verification**

95 Multiple PCR assays for each pathogen were identified from published papers. The complete
96 RefSeq genomes for each pathogen were downloaded from NCBI RefSeq database. If complete
97 genomes are not available in RefSeq database, sequences of the target genes or genome contigs
98 of the pathogen were obtained from the NCBI GenBank database. *In silico* PCR was performed
99 in UGENE (version 48) using primer sets from published papers and the downloaded sequences
100 as templates. *In silico* PCR results were evaluated and the primer set that consistently produced
101 the expected amplicon across the highest proportion of representative genomes was chosen for
102 the BovReproSeq panel. A total of 37 primer sets were included based on the *in silico* PCR
103 results, typically two primer pairs targeting each pathogen (Table Sl). For Coccidian parasites
104 (*Neospora caninum*, *Toxoplasma gondii, Sarcocystis* spp.), one specific primer set for each
105 pathogen, and a pan-Coccidia specific primer set were added. An additional seven MLST primer

106 sets were included for *Campylobacter fetus* subspecies differentiation. A partial sequence
107 encoding Enhanced Green Fluorescent Protein (EGFP) was added as an internal amplification
108 control (IC-EGFP) to rule out false negatives due to presence of PCR inhibitors in the clinical
109 samples. Primers for IC-EGFP were designed using primer3 (12). Primers were synthesized by
110 IDT (Integrated DNA Technologies), and 100 µM stocks were prepared for each primer by
111 suspending the lyophilized primer in nuclease free water. Aliquots (5 µl) of each primer stock
112 (100 µM) were pooled together to create a working stock of pooled primers.

113 **Synthetic controls**

114 Synthetic controls were created and used to optimise and validate the targeted next-generation
115 sequencing approach. These controls were generated by flanking random sequences with target-
116 specific primer binding sites. Random sequences with equal nucleotide frequency were generated
117 using the RSAT random sequence web tool (http://rsat.sb-roscoff.fr/random-seq_form.cgi) and
118 then flanked with target-specific primer binding sites (13). The size of each synthetic control was
119 designed to match the actual size of the target (Supplemental File S3). Two to four of these
120 synthetic controls were integrated into a pUCIDT plasmid containing an ampicillin resistance
121 gene. *E. coli* DH5-alpha was transformed with these plasmids, and control stocks were stored at -
122 80 °C. Subsequently, plasmids were extracted from overnight *E. coli* cultures in lysogeny broth
123 (LB) containing 100 µg/ml ampicillin and plasmid DNA extractions were normalized to 5 ng/µl
124 for further use. Synthetic controls were verified by conventional PCR using both individual and
125 pooled primers. Equimolar amounts of each plasmid were then pooled to make a working stock,
126 ensuring that equal numbers of each synthetic target were in the pooled control. These controls
127 were used as a template for PCR optimisation and as a positive control in clinical sample
128 analysis.

129 **PCR optimisation with synthetic controls**

130 Two high-fidelity DNA polymerase kits (Invitrogen Platinum superfi II – Themofisher Cat no.
131 12361010 – (TF_Plati), and NEB Q5 High-Fidelity 2X Master Mix – NEB Cat no. M0492L
132 (NEB_Q5)) and three annealing temperatures (57 °C, 60 °C, 63 °C) were evaluated. Multiplex
133 PCR was performed with six different combinations of master mixes and annealing temperatures

134 and a primer concentration of 0.015 µM. The pooled plasmids containing synthetic controls (2
135 µl) was used as a template. Resulting PCR products (100 ng each sample) from each
136 combination was barcoded with ONT Native barcoding kit (SQK.NBD114.96) and the resulting
137 pooled library was then sequenced on a GridION sequencer using an R10.4 flowcell. Following
138 these results, a reverse transcriptase step was added to the protocol to enable the detection of
139 *BVDv* (an RNA virus). Three different reverse transcriptase kits (NEB LunaScript Multiplex
140 One-Step RT-PCR Kit (Cat no. E1555S), NEB LunaScript RT SuperMix Kit (Cat no. E3010L)
141 and NEB LunaScript RT Master Mix Kit Primer-free (Cat no. E3025L) with *BVDv* specific
142 primers) were tested using manufacturer's instructions with two *BVDv* positive clinical samples.

143 **Clinical samples**

144 A total of 116 samples were selected from clinical cases submitted to Prairie Diagnostic Services
145 (PDS). Each sample was tested for one or more pathogens as requested by the clients using
146 current diagnostic methods (Table 1). This resulted in 152 outcomes: 64 confirmed positive and
147 88 confirmed negative for the requested tests. Due to the unavailability of positive bovine
148 samples for certain pathogens in the panel, caprine and ovine samples were included, as several
149 abortifacient pathogens are common across cattle, goats and sheep. Various sample types were
150 tested, including placenta, amniotic fluid, vaginal swabs, InPouch media (InPouch TF Bovine,
151 BioMed Diagnostics), semen, and fetal tissues. For pathogens such as Bovine Viral Diarrhea
152 Virus (*BVDv*), *Mycoplasma bovis,* and *Bovine alphaherpesvirus-1 (BoHV-1)*, positive samples
153 associated with abortions were not available. Therefore, positive samples submitted for
154 respiratory disease investigation were included to ensure representation of these pathogens in the
155 experiment. Positive clinical samples were not available for *Listeria monocytogenes, Leptospira*
156 spp. (3 different species), *Toxoplasma gondii, Neospora caninum,* and *Sarcocystis* spp. In these
157 cases, mock test samples were created by spiking 10 µl of ATCC culture (*Listeria*
158 *monocytogenes* (ATCC 7644), *Leptospira* spp. (3 different species ATCC 23580, ATCC 23606,
159 ATCC 23469), *Toxoplasma gondii* (ATCC 50611), *Neospora caninum* (ATCC 50977), and
160 *Sarcocystis* spp. (ATCC PRA-420)) of the pathogen into 90 µl of placenta matrix (clarified tissue
161 homogenate) that was confirmed negative for BovReproSeq pathogens using the current
162 diagnostic methods in Table 1.

163 **BovReproSeq workflow**

164 The complete BovReproSeq laboratory workflow includes nucleic acid extraction and reverse
165 transcription, multiplex PCR and library preparation, sequencing and bioinformatic analysis
166 (Figure 1).

167 ***Total nucleic acid extraction***

168 Applied Biosystems MagMAX CORE Nucleic Acid Purification Kit (Cat no. A32700) was used
169 for extracting total nucleic acid from various sample types, including placenta, amniotic fluid,
170 fetal tissues, and InPouch medium. For placenta and tissue samples, approximately 0.25 g of
171 tissue was added to1 ml of PBS in a beater tube with glass beads. The tissue was then lysed by
172 beating at 30 Hz for 2 minutes in Qiagen TissueLyser II. The lysate was then centrifuged at 1000
173 × g for 1 minute, and 100 µl of the clarified lysate (matrix) was used for extraction. In the case of
174 amniotic fluid (200 µl) and InPouch medium (300 µl), the samples were used directly without
175 any pre-processing. Total nucleic acid was extracted from the prepared samples on a KingFisher
176 Flex instrument using the MagMAX CORE kit according to the manufacturer's instructions.

177 ***First strand cDNA synthesis***

178 Reverse transcription was carried out with 8 µl of total nucleic acid added to 2 µl of NEB
179 Lunascript RT Supermix (New England Biolabs Cat no. E3010). Nuclease free water was used
180 as a no template control (NTC) and pooled plasmids containing the synthetic controls was used
181 as a positive control. The primers were annealed at 25°C for two minutes, followed by cDNA
182 synthesis at 55°C for 10 minutes and heat inactivation at 95°C for 1 minute. cDNA synthesized
183 in this step was used as a template for the multiplex PCR.

184 ***Multiplex PCR***

185 The mastermix was prepared by combining 12.5 µl of NEB Q5 Hot Start High-Fidelity 2 ×
186 Master Mix (New England Biolabs Cat no. M0492L), 9.5 µl of pooled primers (final
187 concentration: 0.5 µM for each primer), 0.9 µl of nuclease-free water, and 0.1 µl of IC-EGFP

188 (150 copies/µl). For each PCR reaction, 2 µl of cDNA/DNA was added to 23 µl of the master
189 mix. The PCR was carried out with the following cycling conditions: initial denaturation at 98°C
190 for 30 seconds, followed by 35 cycles of denaturation at 98°C for 10 seconds, annealing at 57°C
191 for 10 seconds, extension at 72°C for 30 seconds, and a final elongation step at 72°C for 2
192 minutes. The reaction was then held at 4°C.

193 ***Magnetic bead clean-up***

194 Agencourt Ampure XP beads (Cat No. Beckman Coulter A63880) (25 µl) were added to each
195 PCR product and mixed thoroughly. The mixture was incubated for 5 minutes at room
196 temperature and then placed on a magnetic rack to separate the beads. Once the solution became
197 clear, it was aspirated and discarded, leaving the pellet. Subsequently, 200 µl of 70% ethanol was
198 added to each sample without disturbing the pellet. The ethanol was aspirated and discarded.
199 Ethanol washing step was repeated once more. The pellet was allowed to dry for 2 minutes and
200 was then removed from the magnetic rack. The pellet was suspended in 50 µl of nuclease-free
201 water and incubated at room temperature for 5 minutes, followed by magnetic separation. Once
202 clear separation was achieved, the liquid was aspirated and placed into a labeled tube or a 96-
203 well PCR plate. The eluate was then quantified using the Qubit Invitrogen Qubit 4 Fluorometer
204 (Qubit 1x dsDNA HS Assay Kit (Thermofisher, Cat no. Q33231)).

205 ***Library preparation***

206 The sequencing library was prepared using the ligation sequencing amplicons protocol (Version:
207 NBA_9170_v114_revL_15Sep2022). The NEBNext Ultra II End repair/dA-tailing Module
208 (NEB, Cat no. E7546) was used to add dA-tails to the ends of the amplicons. Subsequently,
209 unique barcodes from the Native Barcoding Kit 96 V14 (ONT, Cat No. SQKNBD114.96) were
210 ligated to the end-prepped amplicons using NEB Blunt/TA Ligase Master Mix (NEB, Cat no.
211 M0367). After barcoding, samples with unique barcodes were pooled together. Sequencing
212 adapters were then ligated to the pooled samples using the NEBNext Quick Ligation Module
213 (NEB, Cat no. E6056). Finally, the resulting library was quantified using Invitrogen Qubit 4
214 Fluorometer with the (Qubit 1x dsDNA HS Assay Kit (Thermofisher, Cat no. Q33231)).

Sequencing

The prepared library was loaded onto a ONT R10.4.1 flowcell and sequenced on a GridION platform for a minimum duration of 15 hours. Super-accurate basecalling model was used for basecalling. Only reads with barcodes on both ends were used and barcodes were trimmed while sequencing using the trim barcodes option in MinKNOW software.

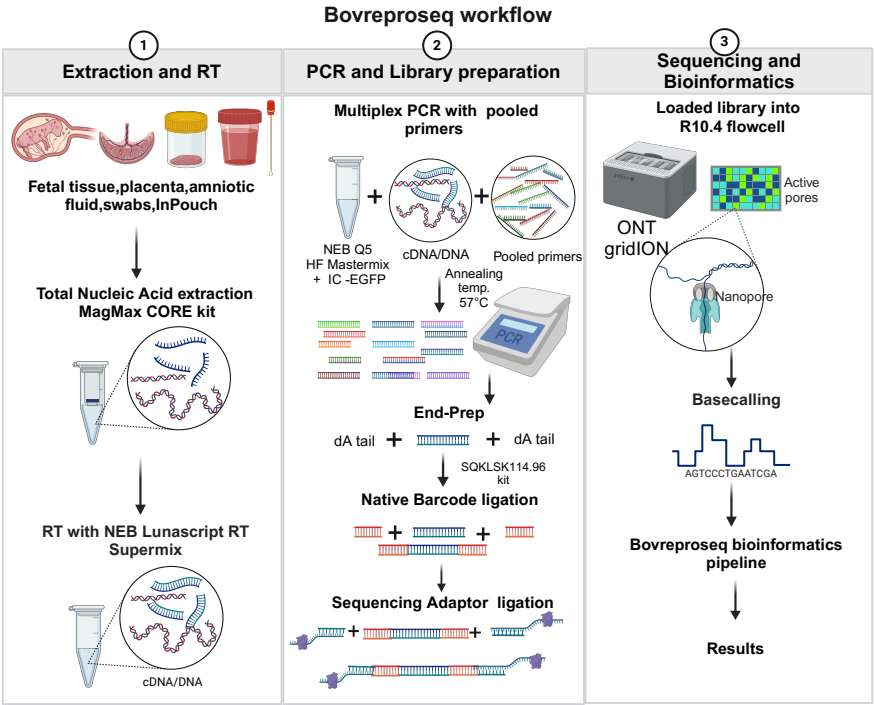

Figure 1: BovReproSeq workflow. Total nucleic acid was extracted from samples using the MagMAX CORE kit, followed by a two-step RT PCR with pooled primers. The generated amplicons were then prepared for sequencing using the Oxford Nanopore Ligation Sequencing Kit. The prepared library was loaded onto the R10.4 flowcell and sequenced on the ONT GridION sequencer. The resulting sequencing data were analyzed using a custom bioinformatics pipeline, and a comprehensive results report was generated.

Bioinformatics analysis

A Nextflow bioinformatics pipeline was developed for the analysis of sequencing data and interpretation of results (Figure 2). The path to a directory containing subdirectories with fastq files and the path to a Kraken2 Plus-PF-8 database (14) were used as the inputs for the pipeline. Individual fastq files were merged followed by quality filtering (q20) for each sample using

232 *nanoq* (15). The resulting merged file was then mapped to a reference sequence containing the
233 amplicon sequences using *minimap2* (16). *Samtools* was used to generate read statistics and split
234 reads based on mapped amplicons (17). Primer trimming was performed on both ends of the
235 sequences using *ampliconclip* option in *samtools*. Consensus sequences were generated on the
236 split reads with a minimum mapping quality of 30 and a base quality of 10 using *samtools* (17).
237 These generated consensus sequences were then polished using *medaka* and compared against a
238 custom Blast database containing the BovReproSeq amplicon sequences using *ABRicate* (18) to
239 determine the presence or absence of the pathogen and internal amplification control (IC-EGFP).
240 Concurrently, taxonomic classification was performed on the raw reads using *Kraken2* (14), to
241 identify the unmapped reads and to support the results from *ABRicate*. The results from *kraken2*
242 were visualized as an interactive pie chart using *krona* (19). Output data from these analyses are
243 collated into a HTML report. All the quality metrics from *samtools* and *Nanostat* were captured
244 and visualised using *multiQC* (20, 21). The pipeline, along with detailed information, is
245 accessible on GitHub at https://github.com/dhineshp565/BovReproSeq.

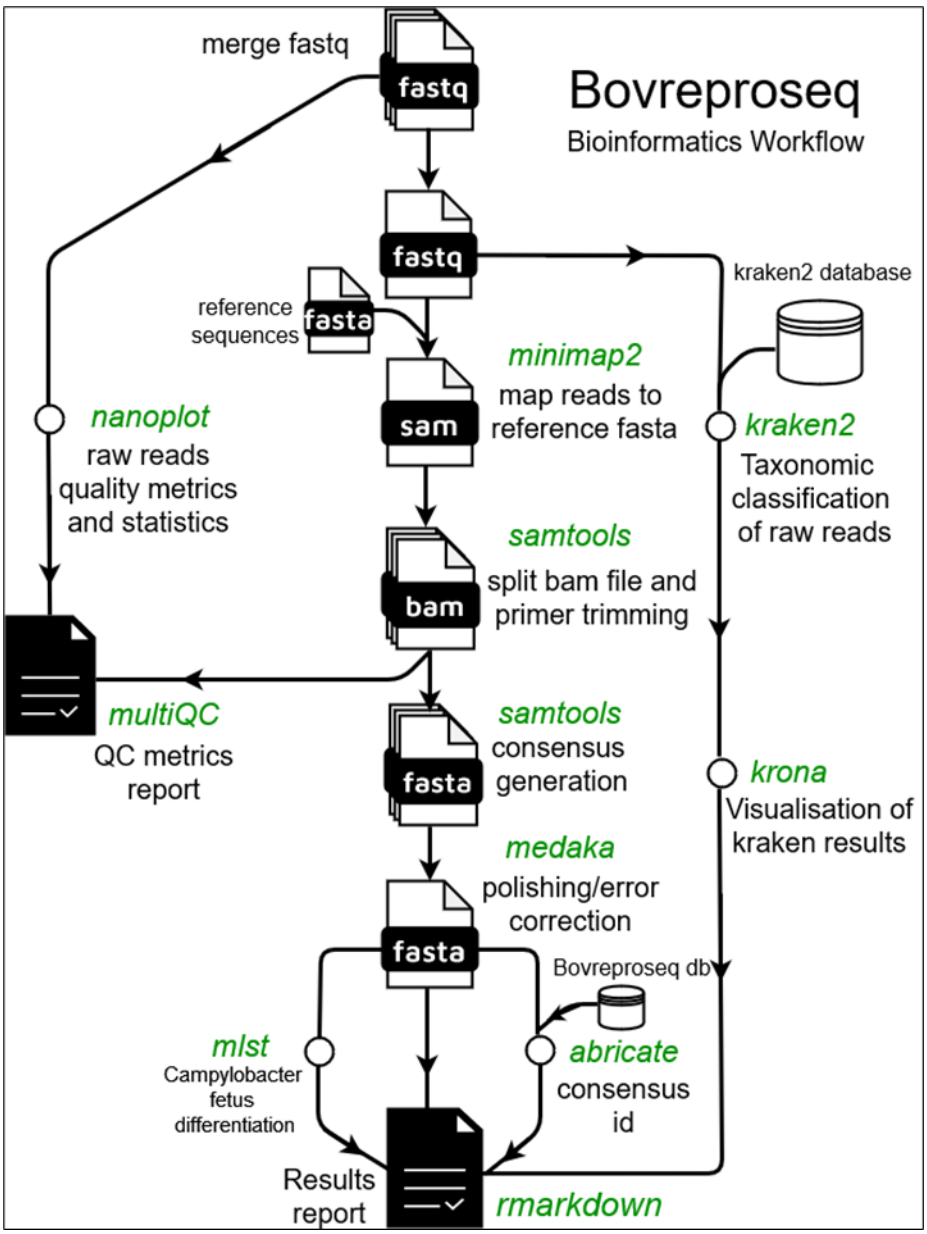

Figure 2: BovReproSeq Bioinformatics workflow. Multiple FASTQ files generated for each sample were merged into one and renamed. The merged file was then mapped against a reference using minimap2. The generated SAM file was subsequently used to generate consensus and read statistics using Samtools. The consensus sequence was further analyzed using MLST and Abricate to determine the type and identity of each consensus. In parallel, merged reads were taxonomically classified using kraken2. Finally, all the results were stitched together to create a comprehensive results report.

Real-time PCR

254 Total nucleic acid was extracted using Applied Biosystems MagMAX CORE Nucleic Acid
255 Purification Kit (Cat no. A32700) and used as the template for real-time PCR. For DNA-based
256 assays, real-time PCR was performed using Applied Biosystems TaqMan Fast Advanced Master
257 Mix (Cat No: 4444965), with specific primers and TaqMan probes tailored to each pathogen.
258 Detection of Bovine Viral Diarrhea Virus (*BVDv*) was carried out using the ThermoScientific
259 VetMAX-Gold BVDV PI Detection Kit (Cat no: 4413938).

260 **Evaluation of read-count threshold and overall performance**

261 Performance metrics, including sensitivity, specificity, accuracy, and precision were calculated
262 for each read-count threshold. Precision-recall curves were plotted to evaluate the performance
263 of the BovReproSeq method across different thresholds. The F1 score (harmonic mean of
264 precision and recall) was used to determine best read-count threshold. An F1 score of 1
265 represents perfect precision and recall, indicating optimal model performance, and a F1 score of
266 0 occurs when either precision or recall (or both) is zero, indicating poor model performance
267 (22). Matthew's correlation coefficient (MCC) was calculated to determine the overall
268 performance of the assay. MCC is a specialized application of Pearson's correlation coefficient,
269 designed for imbalanced binary classification scenarios (22, 23). It incorporates all four
270 categories of the contingency matrix (True Positives, False Positives, True Negatives, and False
271 Negatives) to provide a comprehensive score of overall performance (22, 23). The contingency
272 table, Precision-Recall curves, F1 score and MCC were generated using custom Python scripts
273 with the following libraries: pandas for data manipulation, scikit-learn for performance metric
274 calculations, and matplotlib for visualization.

275 # Results

276 **PCR optimization with synthetic controls**

277 The BovReproSeq method was developed and optimized using synthetic controls before testing
278 the approach on clinical samples. Pooled plasmids containing synthetic control sequences for 12
279 pathogens and one internal amplification control (IC-EGFP) were used as a template to test two

280 different multiplex PCR kits (NEB_Q5 and TF_Plati), and three different annealing
281 temperatures. The total mapped reads obtained from each amplification protocol were as follows:
282 NEB_Q5_57°C: 439,971, TF_Plati_57°C: 409,795, NEB_Q5_60°C: 470,324, TF_Plati_60°C:
283 414,916, NEB_Q5_63°C: 458,083, and TF_Plati_63°C: 486,733. The sequencing depth of each
284 amplicon varied significantly in each combination (Figure 3). The NEB_Q5 with an annealing
285 temperature of 57°C yielded a higher read depth for most of the amplicons (17 out of 29) and a
286 balanced performance, with reads evenly distributed across different amplicons compared to
287 other treatments, which showed skewed distributions favoring specific amplicons (Figure 3). As
288 a result, the NEB Q5 High-Fidelity 2x Master Mix and an annealing temperature of 57°C were
289 used in subsequent experiments. The read depths for *Campylobacter fetus* targets glyA, aspA,
290 and *Ureaplasma diversum* 16S rRNA were particularly low (<500 reads) compared to other
291 targets in all six combinations of annealing temperatures and polymerases. To address this,
292 primer concentration of each primer was increased to 0.5 µM and new primer pool was made.
293 The multiplex PCR was then repeated with NEB Q5 High-Fidelity 2x Master Mix with an
294 annealing temperature of 57°C and new pooled primers (0.5 µM), which resulted in a significant
295 increase in the number of reads mapping to *Campylobacter fetus* glyA (40 reads to 1047 reads),
296 aspA (137 reads to 2670 reads), and *Ureaplasma diversum* 16S rRNA (44 reads to 1214 reads)
297 (Figure 4). As a result, primer concentration of 0.5 µM was used in subsequent experiments.
298 Following these optimizations, five additional pathogens (*Toxoplasma gondii*, *Sarcocystis* spp.,
299 *Bacillus licheniformis, Trueperella pyogenes and Mycoplasma bovis)* were added to the panel,
300 and new synthetic controls were created for these pathogens. The updated pooled synthetic
301 control was tested, and all 37 synthetic targets were detected.

302 A reverse transcriptase step was then added to the multiplex PCR step to enable the detection of
303 *BVDv* (RNA virus), and three different reverse transcriptase kits were tested on two *BVDv*
304 positive clinical samples. Gel electrophoresis results from the NEB LunaScript Multiplex One-
305 Step RT-PCR Kit (Cat no. E1555S) indicated a significant amount of short non-specific
306 amplifications, so the PCR product from this kit was not sequenced. PCR products from the NEB
307 LunaScript RT SuperMix Kit (Cat no. E3010L) and the NEB LunaScript RT Master Mix Kit
308 (Primer-free, Cat no. E3025L) with *BVDv*-specific primers were sequenced. The NEB
309 LunaScript RT Master Mix Kit (Primer-free) yielded 10,598 *BVDv*-specific reads in sample-1

310 and 7,385 *BVDv*-specific reads in sample-2. In contrast, the NEB LunaScript RT SuperMix Kit
311 with random hexamers produced 1,170 *BVDv*-specific reads in sample-1 and 2,567 *BVDv*-
312 specific reads in sample-2 (Table S4). A control experiment without the RT step showed no
313 detection of *BVDv* in either sample. Despite the lower read depth for *BVDv*, the NEB LunaScript
314 RT SuperMix Kit (Cat no. E3010L) was selected because it successfully amplified *BVDv* RNA
315 in clinical samples and allows for the potential detection of additional pathogens not included in
316 the panel, due to the presence of random hexamers in the mix.

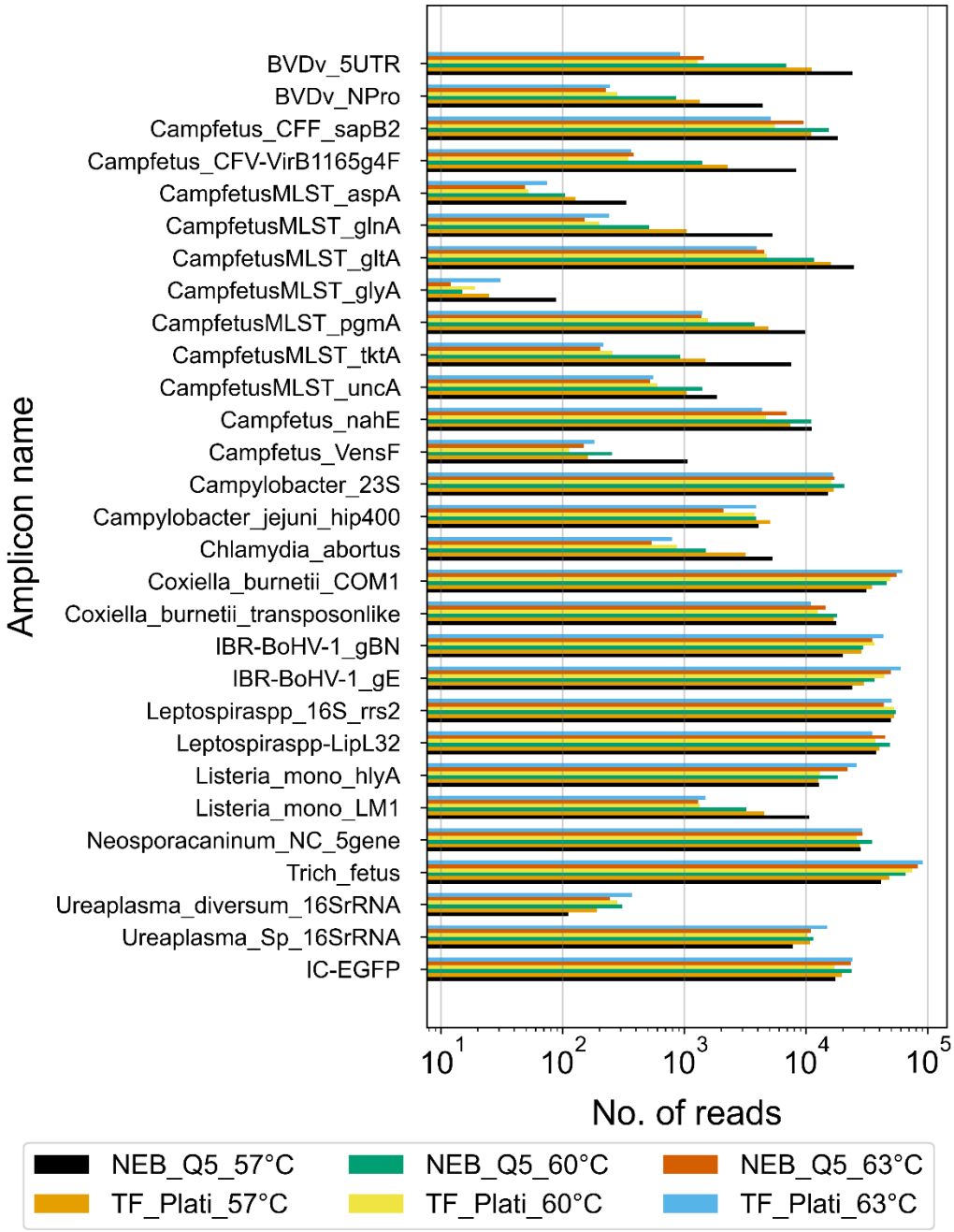

Figure 3: Comparison of sequencing depth of amplicons produced with NEB Q5 High-Fidelity 2x Master Mix and TF Invitrogen Platinum superfi ii polymerase, annealing temperatures of 57°C, 60°C and 63°C, and primer concentration of 0.015 µM. NEB Q5 High-Fidelity 2x Master Mix at 57°C consistently produced higher read depths across most amplicons.

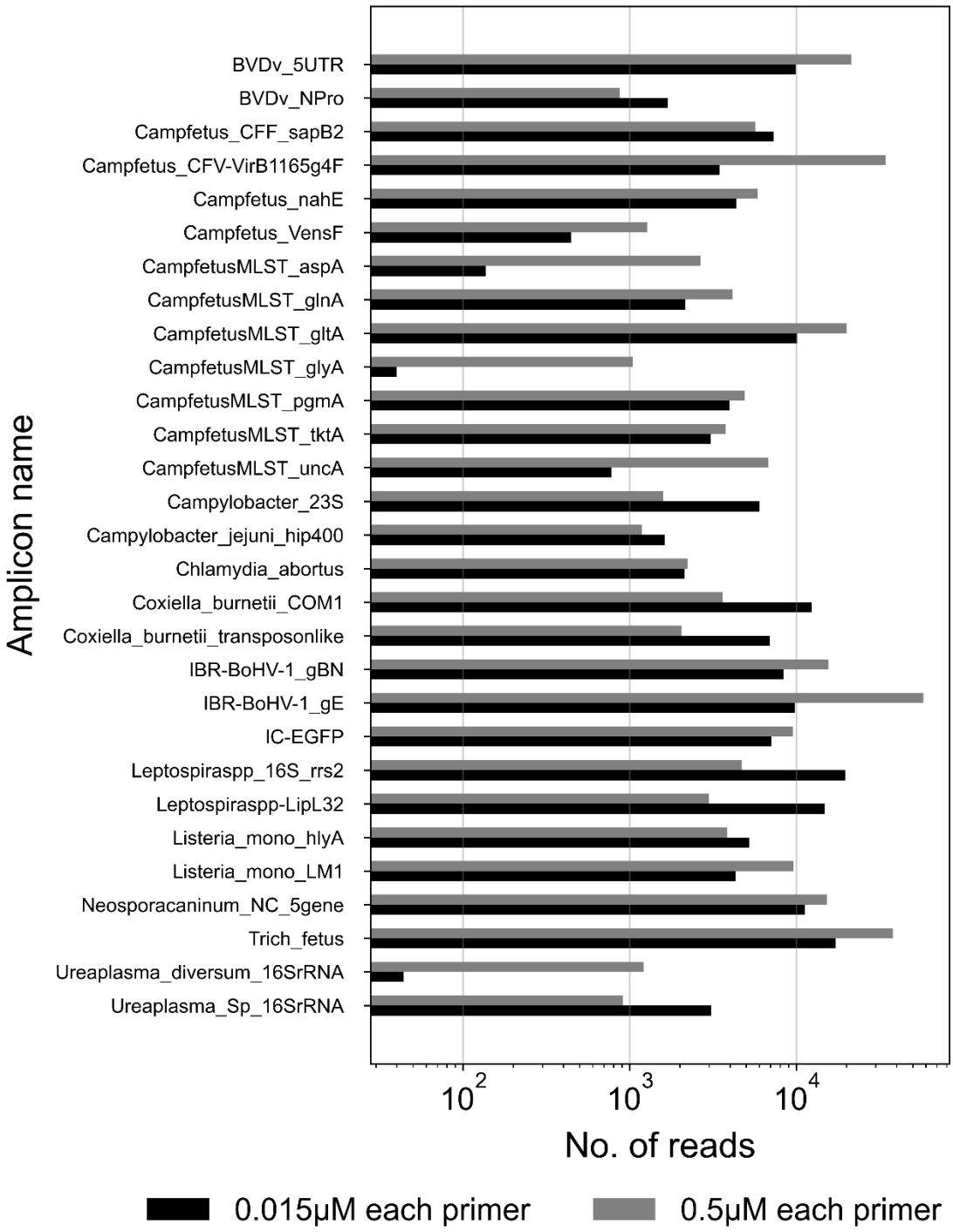

Figure 4: Comparison of sequencing depth for amplicons produced with two different primer concentrations (0.015 µM and 0.5 µM) using NEB Q5 High-Fidelity 2x Master Mix and an annealing temperature of 57°C. The higher primer concentration of 0.5 µM consistently yielded greater read depths across most amplicons compared to the lower concentration of 0.015 µM.

329 **IC-EGFP titration and Repeatability**

330 An internal amplification control was included in the BovReproSeq panel to rule out false
331 negatives caused by inhibiting factors such as humic acid and urea in clinical samples. The
332 amount of IC-EGFP to be spiked into each reaction was determined experimentally, as excessive
333 amounts could interfere with the detection of targets of interest, while insufficient amounts might
334 go undetected.

335 To determine the optimal quantity of IC-EGFP to be spiked into each sample, a titration
336 experiment was conducted using two concentrations of IC-EGFP: 150 copies/reaction and 15
337 copies/reaction. A mock *Neospora caninum* positive sample was prepared by spiking 10 µl of
338 *Neospora caninum* ATCC 50845 culture into 90 µl of placenta matrix (clarified tissue
339 homogenate) confirmed negative by current diagnostic methods (Table 1). This mock sample
340 was serially diluted (10-fold dilution) using placenta matrix (clarified tissue homogenate) as the
341 diluent to maintain consistent matrix composition across dilutions. Total nucleic acid
342 (DNA/RNA) was extracted from the dilutions, and BovReproSeq PCR was performed in
343 triplicate for each of the two IC-EGFP concentrations. For each concentration, a sequencing
344 library was prepared and sequenced on separate flow cells. Real-time PCR (*Neospora caninum)*
345 was performed in triplicate to determine the Ct value of each dilution.

346 Both pan-Coccidia and *Neospora caninum*-pNC-5 gene targets were consistently detected in all
347 replicates and dilutions up to a Ct value of 32 (10^{-5} dilution) in both experiments demonstrating
348 the repeatability of the assay (Figure 5A, 5B). However, for dilutions with Ct values above 32,
349 detection of both *Neospora caninum* targets was inconsistent, occurring only in one or two of the
350 replicates. As the concentration of *Neospora caninum* decreased in the dilutions, the number of
351 IC-EGFP reads increased in both experiments. The IC-EGFP reads were at their lowest when the
352 number of reads and concentration of *Neospora caninum* were at their highest (Ct 18.25).
353 Specifically, in the 15 copies/reaction experiment, there were 23 IC-EGFP reads out of 42,275
354 total reads, while in the 150 copies/reaction experiment, there were 2,081 IC-EGFP reads out of
355 51,175 total reads. In both experiments, IC-EGFP reads stabilized around a dilution of 10^{-5}. The
356 experiment with 15 copies of IC-EGFP per reaction demonstrated better read depth for *Neospora*

caninum, particularly in dilutions with lower pathogen load (Ct 36.56). At this dilution, the 15 copies/reaction experiment yielded 4,021 *Neospora caninum* reads out of 12,932 total reads, compared to only 21 *Neospora caninum* reads out of 85,252 total reads in the 150 copies/reaction experiment. Moreover, *Neospora caninum* was detected by BovReproSeq in dilutions 10^{-7} and 10^{-8} which were not detected with real-time PCR. Although these results suggest that the BovReproSeq method is more sensitive than real-time PCR, it is important to note that these experiments used samples spiked with live organisms. The sensitivity may vary in actual clinical samples, where factors such as sample type, sample matrix and presence of inhibitors could influence detection limits. Based on these results, 15 copies of IC-EGFP per reaction was selected as the optimal concentration for subsequent experiments, as it provided a better balance between internal amplification control and pathogen detection.

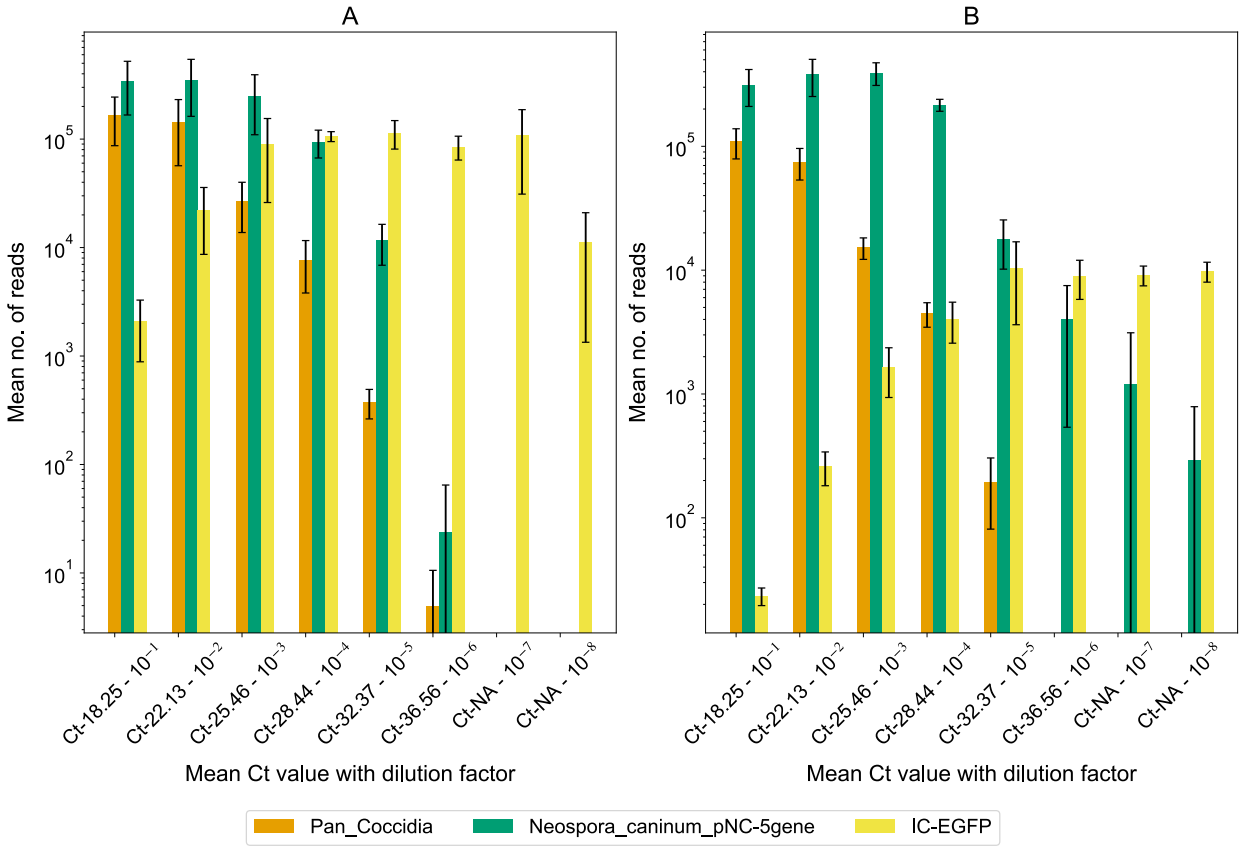

Figure 5: Comparison of mapped reads in serially diluted mock *Neospora caninum* positive samples spiked with 150 copies (A) and 15 copies (B) of IC-EGFP per reaction. Each dilution was tested in triplicate. The x-axis shows mean Ct values from real-time PCR of serial dilutions of the mock sample. Error bars represent the standard deviation of read counts across triplicates. *Neospora caninum* was

373 consistently detected in all replicates up to Ct of 32, with inconsistent detection in subsequent dilutions.
374 15 copies of IC-EGFP per reaction yielded higher read depth for *Neospora caninum* in higher dilutions
375 than 150 copies per reaction.

376 **Specificity evaluation**

377 A total of 29 clinical samples, both positive and negative for BovReproSeq targets, along with
378 no-template controls (NTC) and pooled BovReproSeq synthetic positive controls, were tested to
379 determine the specificity of the assay. Results from the BovReproSeq method matched current
380 diagnostic methods for all samples. However, most of the samples showed a high abundance of
381 reads mapping to the *Campylobacter jejuni* amplicon (cjejuni23s). Manual analysis of the
382 consensus sequence using BLAST revealed that these reads did not correspond to
383 *Campylobacter jejuni*. Instead, they matched other *Campylobacter* species and *Campylobacter*-
384 like organisms. Similar non-specificity issues were observed with the Tpyo-cpn60 target, as
385 reads mapping to this target did not match *Trueperella pyogenes* reference sequences and
386 resulted in poor consensus sequences. Additionally, in *Mycoplasma bovis* (Mbovis) positive
387 clinical samples, the Mbovis gltX target was consistently undetected, and the gspA target was
388 detected with only a very low number of reads. Due to the non-specificity of cjejuni23s, Tpyo-
389 cpn60, and the poor performance of both Mbovis targets, these primers were replaced with more
390 specific targets to improve the overall assay performance. Synthetic controls were developed for
391 the newly added targets, and a new set of pooled controls was made (Table Sl). This updated
392 pooled synthetic control and the clinical samples were then tested with BovReproSeq method.
393 All 37 synthetic controls including the replacements were detected from a single PCR reaction
394 (Figure 6).

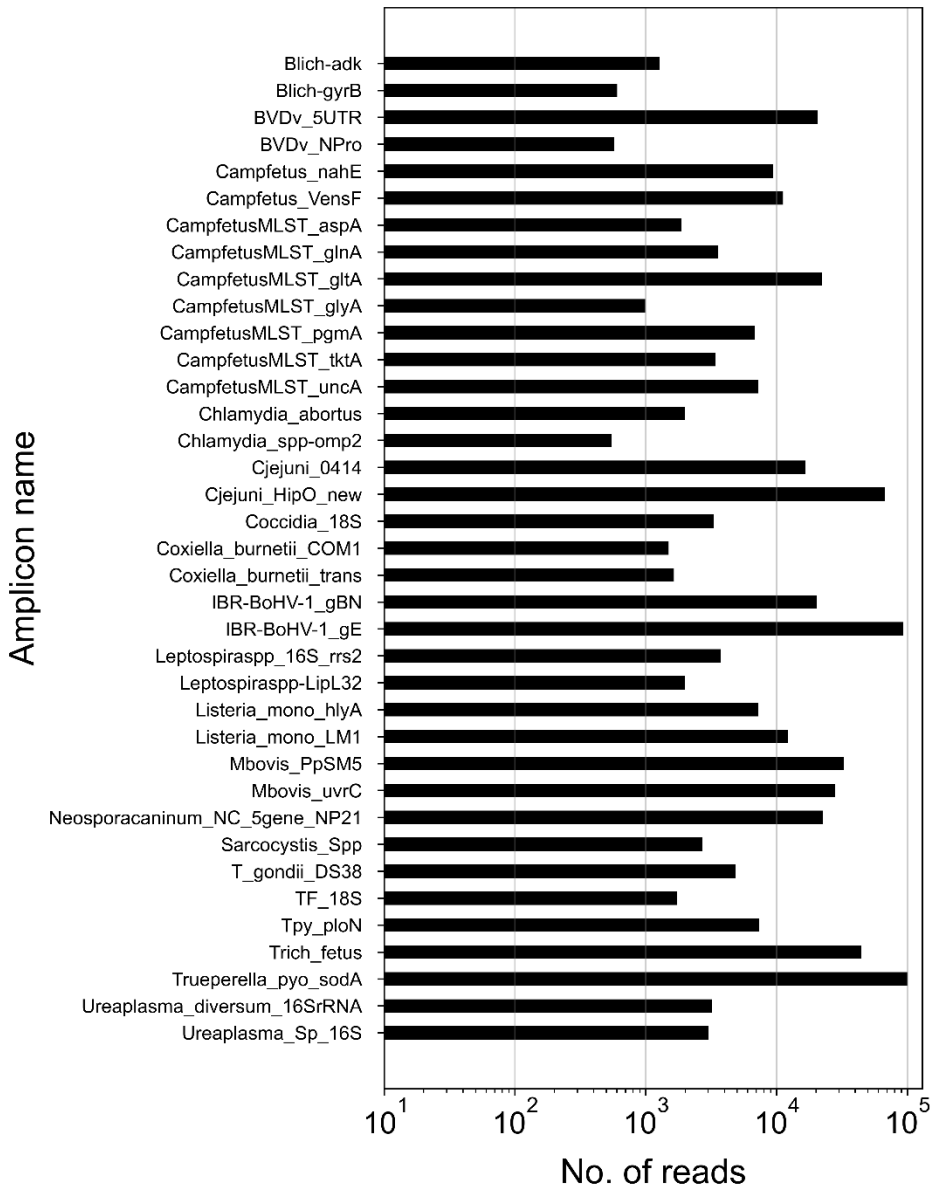

Figure 6: Comparison of read counts for each synthetic control from a single PCR reaction, where the final version of the pooled control was used as a template. All 37 synthetic targets were detected from a single PCR reaction.

Performance on clinical samples

To evaluate the performance of the BovReproSeq method against current diagnostic methods (Table 1), 116 clinical samples with existing results (for tests requested by veterinarians and pathologists) were tested using the BovReproSeq approach. BovReproSeq PCR was performed

403 and PCR amplicons from these samples were pooled into four libraries, loaded onto four R10.4
404 flow cells, and sequenced. The data was analyzed using the BovReproSeq bioinformatics
405 pipeline. The number of raw reads per sample varied significantly, ranging from 6,636 to
406 642,695 reads, with a median of 110,509 reads. The median total bases generated per sample was
407 105.5 Mb, and the median raw read quality score was 19.6. Positive controls had reads and
408 consensus sequences mapping to all 37 synthetic controls, while negative controls (NTCs) were
409 negative for all pathogens. Two samples lacked IC-EGFP reads, indicating potential PCR
410 inhibition, and were omitted from further analysis. The results from the BovReproSeq method
411 matched the existing results from current methods for approximately 93% of the samples
412 (139/150, Table 2). The kappa value of 0.85 indicated almost perfect agreement between the two
413 methods, however, there were 11 false negatives. The Ct values from the current diagnostic
414 PCRs for positive samples ranged from 12 to 37, with all false negatives having Ct values
415 between 33 and 37, which is consistent with low pathogen load (Figure 7). For the six samples
416 submitted for *Campylobacter* spp. PCR, the BovReproSeq method successfully differentiated
417 *Campylobacter fetus* subspecies in all cases and identified the MLST type in five of them.
418 Additionally, BovReproSeq identified *BVDv*-type 2 in two samples among those tested for
419 *BVDv*.

420 Table 2: Comparison of results between current laboratory methods and BovReproSeq. It includes
421 only the results for which specific testing was requested.

		BovReproSeq results		
		Positive	Negative	Total
Current lab methods	Positive	52	11	63
	Negative	0	87	87
	Total	52	98	150

422

423

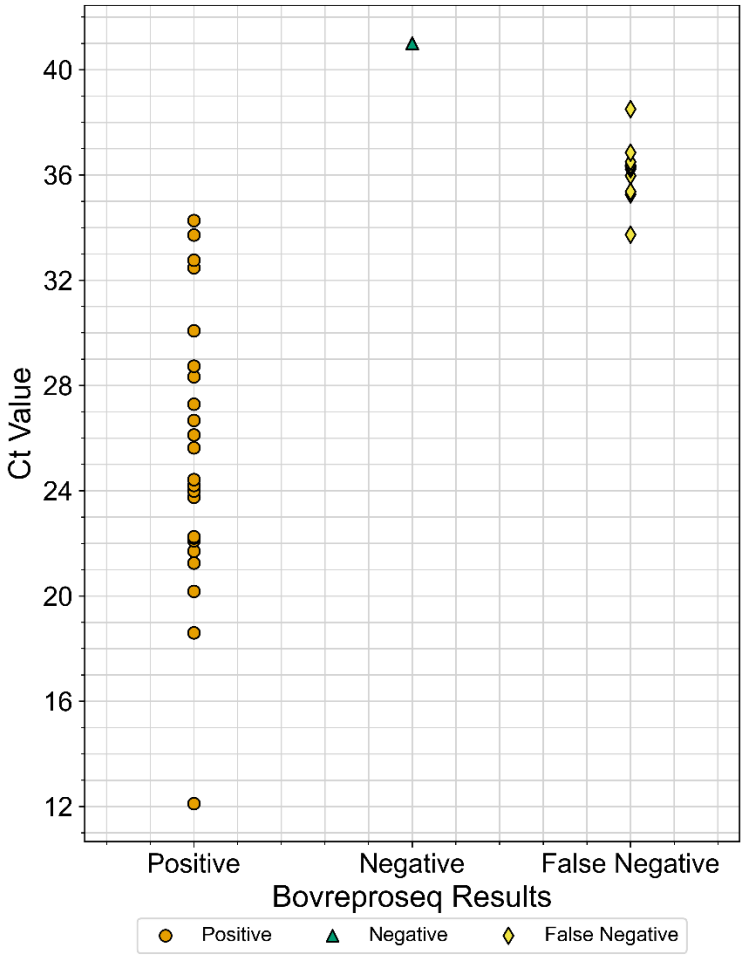

Figure 7: Real-time PCR Ct values of the samples and the corresponding results from the BovReproSeq method. Ct values for the positive samples ranged from 12 to 36, while all the false negatives had a Ct value close to 35 or above indicating that these samples had a very low pathogen load. Additionally, some of these false negative samples had a high abundance of reads mapping to pathogens such as *Ureaplasma diversum* and *Trueperella pyogenes*.

Notably, in 43 samples, additional pathogens were detected using the BovReproSeq method that were not part of the initial testing request (Table S2). In most cases, these additional pathogens were either *Ureaplasma diversum* or *Trueperella pyogenes*. However, in seven cases where the initial test requested was negative, abortifacient pathogens such as *Coxiella burnetii, Campylobacter jejuni*, or *Campylobacter fetus* were identified (Table 3). Subsequent real-time PCR results confirmed the presence of these additional pathogens.

436 Table 3: Comparison of results from initial test request and BovReproSeq method.

Tests requested by veterinarians	**Results from initial request**	**Results BovReproSeq**
Chlamydia abortus, Coxiella burnetii	Negative	*Campylobacter jejuni*
Chlamydia abortus, Coxiella burnetii	Negative	*Campylobacter jejuni*
Chlamydia abortus	Negative	*Campylobacter jejuni*
BVDv, Campylobacter fetus subsp. *venerealis, Coxiella burnetii, BoHV-1, Tritrichomonas fetus*	Negative	*Trueperella pyogenes*
Coxiella burnetii	Negative	*Chlamydia abortus, Trueperella pyogenes*
Coxiella burnetii	Negative	*Campylobacter fetus* subsp. *fetus*
Ureaplasma diversum	Negative	*Coxiella burnetii*
Bovine alphaherpesvirus 1 (BoHV-1)	Positive	*Mycoplasma bovis, Coxiella burnetii, Trueperella pyogenes*
Bovine alphaherpesvirus 1 (BoHV-1)	Positive	*Ureaplasma diversum, Mycoplasma bovis, Trueperella pyogenes*

437

438 The presence of additional pathogens and the detection of pathogens in samples where the initial
439 test request was negative indicated that to comprehensively evaluate the BovReproSeq method,
440 each sample had to be tested for all 17 pathogens using the conventional methods. This would
441 provide a true comparison between the current methods and the BovReproSeq method. This was
442 not possible as some of the samples used were archived nucleic acid samples collected over a
443 long period, limiting our ability to apply methods other than PCR. Real-time PCR assays were
444 available for only 10 BovReproSeq pathogens (*BVDv, Campylobacter jejuni, Campylobacter*
445 *fetus* subsp. *venerealis, Neospora caninum, Toxoplasma gondii, Coxiella burnetii, Chlamydia*
446 *abortus, Tritrichomonas fetus, Ureaplasma diversum, Mycoplasma bovis*). Therefore, real-time
447 PCR was performed on samples (109 samples) that had enough nucleic acid to detect the
448 presence of these 10 pathogens, providing 1090 results. Results from each PCR were collated
449 into a single results table (1090 PCR results). The compiled real-time PCR results for 10
450 pathogens across 109 clinical samples showed 1016 negative and 74 positive outcomes. These
451 real-time PCR results were then compared to the results from the BovReproSeq method. The
452 results from the BovReproSeq method matched the current methods for 98.8% of the samples
453 (1077/1090, Table 4). There were 13 false negative results. All false negative results occurred in
454 samples with Ct values between 33 and 37 except for one sample with a Ct value of 29 for
455 *BVDv.* In this case, four other pathogens (*BoHV-1, Mycoplasma bovis, Coxiella burnetii* and

456 *Trueperella pyogenes*) were detected in high abundance, and the pathologist had reported lesions
457 characteristic of *BoHV-1*. The kappa value of 0.90 indicated almost perfect agreement between
458 the two methods. Due to the imbalance between positive and negative results, the kappa value
459 may not accurately reflect the performance of the assay *(24)*. Therefore, Matthew's Correlation
460 Coefficient (MCC) was calculated. MCC value of approximately 0.90 indicated excellent overall
461 performance of the assay.

462 Table 4: Comparison of results between Real-time PCR assay of 10 pathogens and BovReproSeq.

		BovReproSeq results		
		Positive	Negative	Total
Real-time PCR results	Positive	61	13	74
	Negative	0	1016	1016
	Total	61	1029	1090

463

464 **Evaluation of read-count threshold and assay performance**

465 Read-count thresholds represent the minimum number of reads required to classify a sample as
466 positive or negative. To determine whether a read-count threshold of 10 reads is optimal for
467 classifying a sample as positive for a target, sequence data from clinical samples were analyzed
468 using the BovReproSeq pipeline with 12 different thresholds (ranging from 1 read to 100 reads).
469 The results from each threshold were then compared to the available real-time PCR results. For
470 each read-count threshold, a contingency matrix (2x2 table) was created to check the agreement
471 of results between the two methods. Due to the imbalance between positive and negative results,
472 a precision-recall curve was used instead of a receiver operator characteristic curve to determine
473 the optimal read-count threshold (25, 26). Precision (positive predictive value) and recall
474 (sensitivity) were calculated for each read-count threshold, and a precision-recall curve was
475 plotted (Figure 8). At the lowest threshold of one read, precision was at its lowest around 48%,
476 while sensitivity peaked at 95%. Increasing the read-count threshold to five reads resulted in a
477 substantial improvement in precision (86%) but a marked reduction in sensitivity (85%).

Precision reached 100% at a threshold of 10 reads and remained consistent across remaining thresholds, but sensitivity declined with increasing read-count thresholds and was lowest (70%) at a threshold of 100 reads. The threshold of 10 reads yielded the optimal balance between precision and recall with the maximum F1 score of 0.90. At the read-count threshold of 10 reads, the sensitivity of the assay was approximately 82%, while specificity was 100%. The positive predictive value was 100%, and the negative predictive value was 98.7%. The overall accuracy of the assay was 98.8%. Based on these results, a read count threshold of 10 was determined to be optimal for the BovReproSeq method.

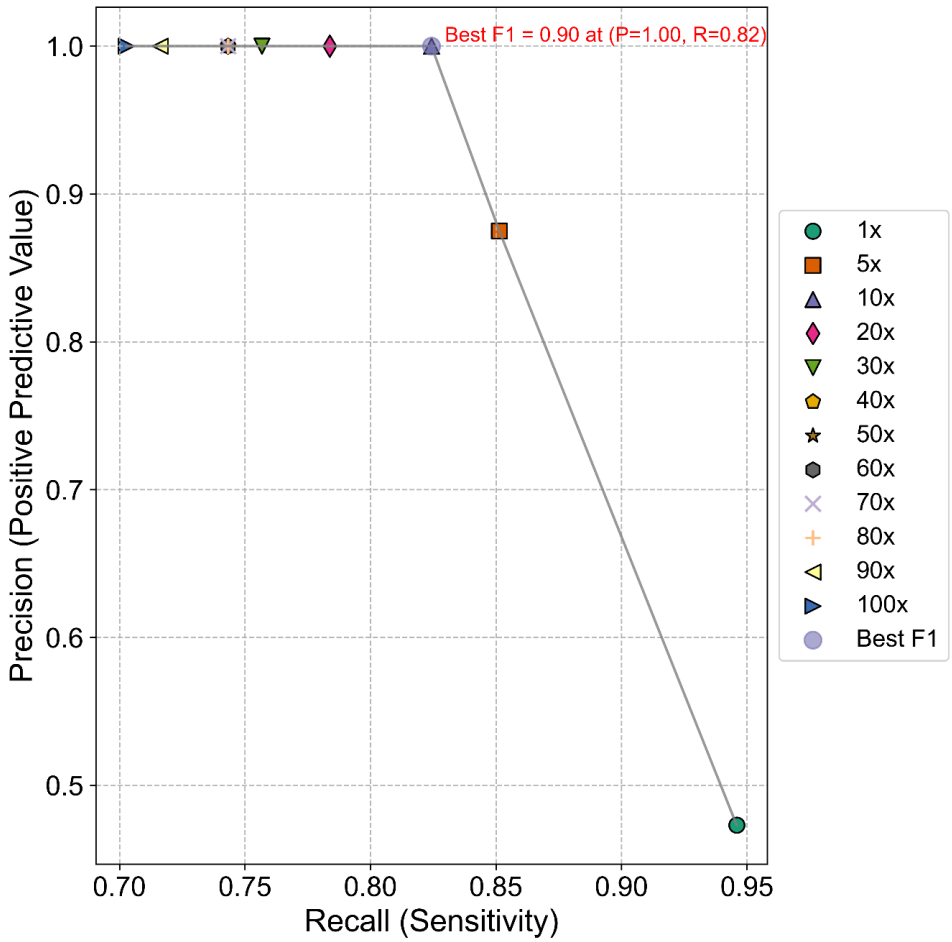

Figure 8: Precision-Recall curve. The graph shows the trade-off between precision (positive predictive value) and recall (sensitivity) at various read-count thresholds. The grey line represents the precision-recall curve, with data points corresponding to different read-count threshold (1x to 100x). The optimal performance point (10x) is highlighted, showing a best F1 score of 0.90 with a precision of 1.00 and recall of 0.82.

492 # Discussion

493 Our study demonstrates the development and validation of a syndromic next-generation
494 sequencing panel (BovReproSeq) designed to simultaneously detect 17 pathogens associated
495 with bovine reproductive failure. Initial testing with a pooled synthetic control successfully
496 identified all 37 synthetic targets in the products of a single PCR reaction, demonstrating the
497 multiplexing ability of the assay. The results from BovReproSeq testing of 116 clinical samples
498 matched with existing results from current diagnostic methods for 93% of samples. Additionally,
499 the BovReproSeq method detected pathogens not originally requested for testing, indicating
500 possible mixed infections (Table 3, S2). This ability of the BovReproSeq to identify additional
501 pathogens highlights its advantage over current single-plex diagnostic methods by providing
502 broader detection without iterative testing.

503 In addition to detection of 17 pathogens, BovReproSeq method enables the typing of *BVDv*
504 (type-1 and type-2) and *Campylobacter fetus* (MLST), aiding epidemiological studies. Since this
505 assay involves random hexamers and multiple primers, there is potential to detect pathogens not
506 included in the panel, further broadening the scope of the assay. The entire BovReproSeq
507 workflow (Figure 1) can be completed within 48 hours of receiving the sample and can be cost-
508 effective if multiple samples are processed within a single flow cell.

509 The BovReproSeq method has some limitations, particularly in samples with high Ct values
510 (close to or above 35) indicating that very low pathogen concentrations in the sample may not be
511 detected (Figure 7). This finding is consistent with a similar study in which a targeted next-
512 generation sequencing method was not able to detect vector-borne pathogens in canine samples
513 with Ct values greater than 35 (9). It is worth noting that PCR artifacts and cross-contamination
514 can sometimes lead to amplification at Ct values >35, and samples with Ct values >35 are often
515 considered "suspect" rather than "positive" in diagnostic laboratories. Moreover, pathogens like
516 *Coxiella burnetii* and *Chlamydia abortus* are known to occur in healthy animals in low numbers
517 and their presence in low quantities without associated lesions is generally not considered as the
518 cause of the abortion (27). In most infectious abortion cases, high pathogen loads with
519 characteristic lesions or signs are typical. Thus, the failure of BovReproSeq to detect very low

520 abundance pathogens may not be a clinically significant limitation. Notably, we had only one
521 sample with a Ct value of 29 for *BVDv* where BovReproSeq failed to detect the virus. In this
522 case, four other pathogens in the BovReproSeq panel (BoHV-1, *Mycoplasma bovis, Coxiella*
523 *burnetii* and *Trueperella pyogenes*) were detected in high abundance, and the pathologist had
524 reported lesions characteristic of BoHV-1 (Table 3). Similarly, we frequently detected pathogens
525 such as *Ureaplasma diversum* and *Trueperella pyogenes*, and in a couple of cases, their
526 abundance had a negative impact on the detection of the causative pathogen. This suggests that
527 multiple pathogens in high abundance in a sample may impede the detection of pathogens in low
528 abundance.

529 The analytical sensitivity of our BovReproSeq method is lower compared to PCR assays and it
530 could be improved by reducing the number of samples per flow cell and extending sequencing
531 time to improve the sequencing depth of low abundance targets. The one-pot multiplex PCR
532 could also be split into multi-pot PCR with smaller primer pools to improve sensitivity, and the
533 number of PCR cycles could be increased from 35 to 40 to increase the number of copies of the
534 amplicons. Regarding primers and PCR target selection, selecting specific genes of pathogens
535 rather than conserved housekeeping genes was more successful, as the primers targeting
536 conserved genes (*T. pyogenes* cpn60 and *C. jejuni* 23S rRNA gene) were more likely to amplify
537 sequences from related organisms present in the sample and obscure the targeted pathogens of
538 interest. This observation should be factored into any future additions to or modification of the
539 panel.

540 Similar studies have been published recently, but our research presents a few methodological
541 differences and advancements (Anis et al., 2018; Kattoor et al., 2022). Anis et al. developed a
542 targeted NGS method for the simultaneous detection of 43 common bovine and small-ruminant
543 pathogens. Their primer design was a collaboration with the White Glove Team (Ion Torrent;
544 Thermofisher Scientific) using Ion Ampliseq Designer (Ion Torrent; Thermofisher Scientific,
545 Waltham, MA). Kattoor et al. focused on the comprehensive detection of 17 vector-borne
546 pathogens in canines. The primers in their study were designed with Thermofisher Scientific's
547 AgriSeq Bioinformatics team, and the primer sets are now proprietary items of Thermofisher
548 (Kattoor et al., 2022). Both studies used Ion Torrent Suite and Geneious software for

549 bioinformatics and neither includes an internal amplification control. In our study, we have used
550 previously published primers and further evaluated them based on *in silico* PCR results.
551 Additionally, in the BovReproSeq bioinformatics pipeline, we have used open-source tools and
552 packaged them as one Nextflow pipeline that produces a readable results report and the pipeline
553 can be scaled up to run on any platform (Paolo Di Tommaso et al., 2017). Our pipeline is
554 currently available on GitHub, and the primer information is included in the supplementary
555 section (Table S1) to ensure open access to our methods and facilitate reproducibility.

556 The previously published bovine and canine pathogen sequencing panels used Thermofisher Ion
557 Torrent technology with IonChef for sequencing multiplex PCR products. This second-
558 generation sequencing technology has limitations including short reads (<400 bp), low
559 throughput and poor accuracy in homopolymer regions. BovReproSeq uses Oxford Nanopore
560 technology (ONT), a long-read approach that enables sequencing of longer gene segments than
561 the Ion Torrent platform and allows for real-time sequencing and analysis, resulting in faster
562 turnaround times. Additionally, BovReproSeq uses only full-length reads to generate the
563 consensus sequences resulting highly accurate results. The use of long reads allows us to perform
564 MLST typing, as most MLST fragments are around 600 bp, which would be challenging with
565 Ion Torrent technology. Although the raw read accuracy with ONT was known to be lower
566 compared to other platforms, the raw read quality has significantly improved with the newest
567 version of the R10 nanopore and the associated kit 14 chemistry (28, 29). In our BovReproSeq
568 pipeline, we exclusively used raw reads with a quality score of 20 or above (>99% accuracy) to
569 ensure that only highly accurate reads were used for downstream analysis.

570 The incorporation of an internal amplification control in the BovReproSeq assay allows the rule
571 out of false negatives due to PCR inhibition. The amount to be spiked into each reaction was
572 determined experimentally and to ensure repeatability, we performed this assay in triplicates on
573 serial dilutions with consistent results between the replicates (Figure 5A, 5B). Additionally, the
574 pooled synthetic positive control for the assay allows the detection of positive control
575 contamination since the sequences of the positive control amplicons are unrelated to the actual
576 biological targets. Finally, we analysed the sequence data generated from 109 clinical samples

using 12 different read-count thresholds to determine the optimal threshold and the overall performance of assay, significantly enhancing the reliability and robustness of our results.

Conclusion

In conclusion, our study presents a significant improvement to diagnosis of infectious causes of Bovine reproductive failure. The ability of the assay to detect 17 pathogens despite some limitations, makes BovReproSeq panel a promising tool for veterinary diagnostics. This approach could be applied to other syndromic diseases and could contribute to broader improvements in animal health diagnostics. However, as with all nucleic acid-based methods, the detection of a pathogen does not necessarily mean it is the cause of the disease and results from the BovReproSeq method must be carefully interpreted with pathological findings and case history.

Data availability

All synthetic control sequences, reference amplicon sequences, and the custom ABRicate database with Accession numbers for each sequence, along with the Nextflow pipeline, are available in the GitHub repository at https://github.com/dhineshp565/BovReproSeq.

Acknowledgements

We thank Saskatchewan Agriculture Development Fund (ADF) and the Beef Cattle Research Council (BCRC) for their financial support. We also extend our gratitude to the PDS Molecular Diagnostics section for providing the samples and assisting with follow-up real-time PCR, and Anju Tumber for her help in procuring reagents.

References

1. Bellows DS, Ott SL, Bellows RA. 2002. Review: Cost of Reproductive Diseases and Conditions in Cattle11Portions of this research were conducted under a cooperative

600 agreement between USDA, ARS and the Montana Agric. Exp. Stn. Mention of a
601 proprietary product does not constitute a guarantee or warranty of the product by USDA,
602 Montana Agric. Exp. Stn., or the authors and does not imply its approval to the exclusion of
603 other products that may also be suitable. USDA, ARS, Northern Plains Area, is an equal
604 opportunity/affirmative action employer, and all agency services are available without
605 discrimination. Prof Anim Sci 18:26–32.

606 2. Waldner C. 2010. Gross postmortem and histologic examination findings from abortion
607 losses and calf mortalities in western Canadian beef herds. Can Vet J 51:1227–1238.

608 3. Reichel MP, Wahl LC, Hill FI. 2018. Review of Diagnostic Procedures and Approaches to
609 Infectious Causes of Reproductive Failures of Cattle in Australia and New Zealand. Front
610 Vet Sci 5.

611 4. Bradley L. Njaa, Njaa BL. 2011. Kirkbride's diagnosis of abortion and neonatal loss in
612 animals. https://doi.org/10.1002/9781119949053.

613 5. Dumkow LE, Worden LJ, Rao SN. 2021. Syndromic diagnostic testing: a new way to
614 approach patient care in the treatment of infectious diseases. J Antimicrob Chemother
615 76:iii4–iii11.

616 6. Laupland KB, Valiquette L. 2013. The changing culture of the microbiology laboratory.
617 Can J Infect Dis Med Microbiol 24:125–128.

618 7. Slinporn Prachayangprecha, Prachayangprecha S, Claudia M. E. Schapendonk,
619 Schapendonk CME, Marion Koopmans, Koopmans M, Albert D. M. E. Osterhaus,
620 Osterhaus ADME, Anita C. Schürch, Schürch AC, Suzan D. Pas, Pas SD, Annemiek A. van

621 der Eijk, van der Eijk AA, Yong Poovorawan, Poovorawan Y, Bart L. Haagmans,
622 Haagmans BL, Saskia L. Smits, Smits SL. 2014. Exploring the Potential of Next-
623 Generation Sequencing in Detection of Respiratory Viruses. J Clin Microbiol 52:3722–
624 3730.

625 8. Anis E, Hawkins IK, Ilha MRS, Woldemeskel M, Saliki JT, Wilkes RP. 2018. Evaluation
626 of Targeted Next-Generation Sequencing for Detection of Bovine Pathogens in Clinical
627 Samples. J Clin Microbiol 56.

628 9. Kattoor JJ, Jobin Jose Kattoor, Emma Nikolai, Emma Nikolai, Barbara Qurollo, Barbara A.
629 Qurollo, Rebecca P. Wilkes, Rebecca P. Wilkes. 2022. Targeted Next-Generation
630 Sequencing for Comprehensive Testing for Selected Vector-Borne Pathogens in Canines.
631 Pathogens 11:964–964.

632 10. Whitford W, Hawkins V, Moodley KS, Grant MJ, Lehnert K, Snell RG, Jacobsen JC. 2022.
633 Proof of concept for multiplex amplicon sequencing for mutation identification using the
634 MinION nanopore sequencer. 1. Sci Rep 12:8572.

635 11. Deng X, Achari A, Federman S, Yu G, Somasekar S, Bártolo I, Yagi S, Mbala-Kingebeni
636 P, Kapetshi J, Ahuka-Mundeke S, Muyembe-Tamfum J-J, Ahmed AA, Ganesh V,
637 Tamhankar M, Patterson JL, Ndembi N, Mbanya D, Kaptue L, McArthur C, Muñoz-
638 Medina JE, Gonzalez-Bonilla CR, López S, Arias CF, Arevalo S, Miller S, Stone M, Busch
639 M, Hsieh K, Messenger S, Wadford DA, Rodgers M, Cloherty G, Faria NR, Thézé J, Pybus
640 OG, Neto Z, Morais J, Taveira N, R. Hackett J, Chiu CY. 2020. Metagenomic sequencing
641 with spiked primer enrichment for viral diagnostics and genomic surveillance. 3. Nat
642 Microbiol 5:443–454.

643 12. Andreas Untergasser, Untergasser A, Ioana Cutcutache, Cutcutache I, Triinu Kõressaar,
644 Koressaar T, Jian Ye, Ye J, Brant C. Faircloth, Faircloth BC, Maido Remm, Remm M,
645 Rozen S, Steve Rozen, Rozen SG, Rozen SG. 2012. Primer3—new capabilities and
646 interfaces. Nucleic Acids Res 40.

647 13. Mark F. Kavlick, Mark F Kavlick, Kavlick MF. 2018. Development of a universal internal
648 positive control. BioTechniques 65:275–280.

649 14. Derrick E. Wood, Derrick E. Wood, Wood DE, Jennifer Lu, Lu J, Ben Langmead,
650 Langmead B. 2019. Improved metagenomic analysis with Kraken 2. Genome Biol 20:257–
651 257.

652 15. Steinig EJ, Eike J. Steinig, Coin L, Lachlan Coin. 2022. Nanoq: ultra-fast quality control
653 for nanopore reads. J Open Source Softw 7:2991–2991.

654 16. Heng Li, Li H. 2017. Minimap2: fast pairwise alignment for long nucleotide sequences.
655 Bioinform https://doi.org/10.1093/bioinformatics/bty191.

656 17. Petr Danecek, Danecek P, James K. Bonfield, Bonfield JK, Liddle J, Jennifer Liddle, Liddle
657 J, John Marshall, Marshall J, Valeriu Ohan, Ohan V, Martin Pollard, Pollard MO, Andrew
658 Whitwham, Whitwham A, Thomas M. Keane, Keane TM, Keane TM, Shane McCarthy,
659 McCarthy S, Davies SJ, Robert M. Davies, Davies RL, Li H, Heng Li, Li H. 2021. Twelve
660 years of SAMtools and BCFtools. GigaScience 10.

661 18. Seeman T. 2016. ABRicate.

662 19. Brian D. Ondov, Ondov BD, Nicholas H. Bergman, Bergman NH, Adam M. Phillippy,
663 Phillippy AM. 2011. Interactive metagenomic visualization in a Web browser. BMC
664 Bioinformatics 12:385–385.

665 20. Wouter De Coster, De Coster W, Svenn D'Hert, D'Hert S, Darrin T. Schultz, Schultz DT,
666 Marc Cruts, Cruts M, Marc Cruts, Marc Cruts, Christine Van Broeckhoven, Van
667 Broeckhoven C. 2018. NanoPack: visualizing and processing long read sequencing data.
668 bioRxiv 237180.

669 21. Philip Ewels, Ewels P, Måns Magnusson, Magnusson M, Sverker Lundin, Lundin S, Max
670 Käller, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples
671 in a single report. Bioinformatics 32:3047–3048.

672 22. Steven A. Hicks, Steven A Hicks, Inga Strümke, Inga Strümke, Vajira Thambawita, Vajira
673 Thambawita, Malek Hammou, Malek Hammou, Michael Riegler, Michael A Riegler, Pål
674 Halvorsen, Pål Halvorsen, Sravanthi Parasa, Sravanthi Parasa. 2022. On evaluation metrics
675 for medical applications of artificial intelligence. Sci Rep 12:5979–5979.

676 23. Davide Chicco, Chicco D, Niklas Tötsch, Tötsch N, Niklas Tötsch, Giuseppe Jurman,
677 Jurman G. 2021. The Matthews correlation coefficient (MCC) is more reliable than
678 balanced accuracy, bookmaker informedness, and markedness in two-class confusion
679 matrix evaluation. Biodata Min 14:13–13.

680 24. Rosario Delgado, Delgado R, Xavier-Andoni Tibau, Tibau X-A, Xavier-Andoni Tibau.
681 2019. Why Cohen's Kappa should be avoided as performance measure in classification.
682 PLOS ONE 14.

683 25. Deepak Dhamnetiya, Dhamnetiya D, Ravi Prakash Jha, Ravi Prakash Jha, Jha RP, Shalini
684 Shalini, Shalini Shalini, Krittika Bhattacharyya, Bhattacharyya K. 2021. How to Analyze
685 the Diagnostic Performance of a New Test? Explained with Illustrations. J Lab Physicians
686 https://doi.org/10.1055/s-0041-1734019.

687 26. Takaya Saito, Saito T, Marc Rehmsmeier, Rehmsmeier M. 2015. The Precision-Recall plot
688 is more informative than the ROC plot when evaluating binary classifiers on imbalanced
689 datasets. PLOS ONE 10:1–21.

690 27. Murray Hazlett, Hazlett M, Rebeccah McDowall, McDowall RM, McDowall R, J Delay,
691 DeLay J, Margaret Stalker, Stalker M, Beverly McEwen, McEwen B, Tony van Dreumel,
692 van Dreumel T, Maria Spinato, Spinato M, Brian Binnington, Brian Binnington, Binnington
693 B, Đurđa Slavić, Slavic D, Slavic D, Susy Carman, Carman S, Hugh Y. Cai, Cai HY. 2013.
694 A prospective study of sheep and goat abortion using real-time polymerase chain reaction
695 and cut point estimation shows Coxiella burnetii and Chlamydophila abortus infection
696 concurrently with other major pathogens. J Vet Diagn Invest 25:359–368.

697 28. Mantas Sereika, Mantas Sereika, Rasmus Hansen Kirkegaard, Kirkegaard RH, Søren
698 Michael Karst, Karst SM, Thomas Yssing Michaelsen, Thomas Yssing Michaelsen, Emil
699 A. Sørensen, Sørensen EA, Rasmus Dam Wollenberg, Rasmus Dam Wollenberg, Mads
700 Albertsen, Albertsen M. 2022. Oxford Nanopore R10.4 long-read sequencing enables the
701 generation of near-finished bacterial genomes from pure cultures and metagenomes without
702 short-read or reference polishing. Nat Methods https://doi.org/10.1038/s41592-022-01539-
703 7.

704 29. Tianyuan Zhang, Hanzhou Li, Silin Ma, Jian Cao, Hao Liao, Qiaoyun Huang, Wenli Chen.
705 2023. The newest Oxford Nanopore R10.4.1 full-length 16S rRNA sequencing enables the
706 accurate resolution of species-level microbial community profiling. Appl Environ
707 Microbiol 89.

708