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 (bacteria, virus and protozoa) 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 clinical sensitivity of the assay was approximately 82%, while clinical specificity was 100%. The overall accuracy of the assay was 98.8%. Matthew's Correlation Coefficient (correlation coefficient of binary classification) 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.

Importance

Bovine reproductive failure causes substantial economic losses to beef and milk producers, and infectious disease contributes significantly to this syndrome. Etiologic diagnosis is complicated since multiple pathogens can be involved and infections with some pathogens are asymptomatic or cause similar clinical signs. A stepwise approach to diagnostic testing is time-consuming and increases the risk of missing the correct diagnosis. BovReproSeq is a next-generation sequencing based diagnostic panel that allows detection of 17 reproductive failure pathogens simultaneously.

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 $2.8 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 this condition also pose a concern to humans as some of them belong to risk group-3 or are zoonotic.

Multiple pathogens are associated with bovine reproductive failure including viruses, bacteria, fungi, and protozoa (Table-1). Infections with some of these pathogens are asymptomatic or cause similar clinical signs and differentiation between these pathogens requires a variety of laboratory tests (Table-1) . There is no recent data available on the diagnostic success rate in Western Canada, however, based on 338 aborted fetuses submitted to Prairie Diagnostic Services Inc. (PDS) in 2010, 52% had no diagnosis, 41% were attributed to an infectious disease, and 7% to a developmental anomaly (2). The success rate for diagnosing bovine reproductive failure is typically low as there is a significant delay between infection and the occurrence of abortion (3). The diagnostic success rate also depends on the availability of appropriate samples (e.g., fresh fetus, placenta, amniotic fluid etc.), and this is not always possible as aborted fetuses are often autolyzed (4). Veterinarians and producers usually focus on testing the most likely cause and if the initial test fails to confirm the suspected cause, subsequent tests are requested. This stepwise approach to diagnostic testing is time-consuming and increases the risk of missing the correct diagnosis as veterinarians and producers are less likely to follow through a diagnostic plan if the initial tests are negative. A syndromic panel approach in which multiple pathogens are tested for simultaneously, would increase the chance of successful diagnosis and significantly reduce the turnaround time for diagnosis (5). Currently, there is no syndromic panel test available in Canada for bovine reproductive pathogens.

65 PCR is the most preferred method for detection of reproductive failure pathogens due to its high
66 analytical sensitivity and specificity. Even though multiple targets can be detected using PCR,
67 there is a limit to the number of targets that can be included in multiplex assays (6).
68 Metagenomic next-generation sequencing (mNGS) can overcome this limitation by either
69 sequence-independent amplification of the genetic material in the sample or by directly
70 sequencing the genetic material, thereby providing comprehensive microbial analysis (7). The
71 sensitivity of this method, however, is lower compared to species specific real-time PCR due to
72 the abundance of host genetic material in the sample (8). Sensitivity of mNGS could be
73 improved with higher throughput sequencers and various enrichment methods, such as DNase
74 treatment, rRNA removal, or probe capture; however, this would significantly increase both
75 costs and turnaround times. These limitations can be overcome by combining PCR with high-
76 throughput next-generation sequencing. With this targeted approach, multiple pathogen-specific
77 targets are first amplified in an ultra-multiplex PCR and amplified fragments are then sequenced
78 and analysed (9, 10). While there are several high-throughput sequencing platforms available,
79 the Oxford Nanopore platform offers real-time sequence data and analysis, enabling rapid and
80 timely diagnosis. Multiple samples can be processed in a single sequencing run and diagnostic
81 costs can be significantly reduced. In addition to detecting pathogens of interest, targeted
82 metagenomic sequencing offers the ability to determine strain types (*in silico* serotyping,
83 virulence typing, multi-locus sequence typing (MLST), etc.) and identify variants, which could
84 be useful in differentiating closely related pathogens such as *Campylobacter fetus* subsp. *fetus*
85 and *Campylobacter fetus* subsp. *venerealis* (11). There is also potential to detect emerging or
86 novel pathogens with this approach using random primers along with the targeted primers (12).

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

93 **Materials and methods**

94 **Target pathogen selection**

95 Seventeen pathogens including bacteria, protozoa and viruses (both DNA and RNA viruses)
96 were selected for inclusion in the panel (Table 1). Pathogens were selected based on the results
97 of reproductive failure cases that were submitted to Prairie Diagnostic Services since 2014 and
98 additional advice from veterinarians and diagnostic pathologists. Originally 12 pathogens were
99 included and after successful results from initial experiments, five more pathogens were added to
100 the panel bringing the final count to 17 pathogens.

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

102 Multiple PCR assays for each pathogen were identified from published papers. The complete
103 RefSeq genomes for each pathogen were downloaded from NCBI RefSeq database. If complete
104 genomes were not available in RefSeq database, sequences of the target genes or genome contigs
105 of the pathogen were obtained from the NCBI GenBank database. *In silico* PCR was performed
106 in UGENE (version 48) using primer sets from published papers and the downloaded sequences
107 as templates. *In silico* PCR results were evaluated and the primer set that consistently produced
108 the expected amplicon across the highest proportion of representative genomes was chosen for
109 the BovReproSeq panel. A total of 37 primer sets were included based on the *in silico* PCR
110 results, typically two primer pairs targeting each pathogen (Table Sl). For Coccidian parasites
111 (*Neospora caninum*, *Toxoplasma gondii, Sarcocystis* spp.), one specific primer set for each
112 pathogen, and a pan-Coccidia specific primer set were added. An additional seven MLST primer
113 sets were included for *Campylobacter fetus* subspecies differentiation. A partial sequence
114 encoding Enhanced Green Fluorescent Protein (EGFP) was added as an internal amplification
115 control (IC-EGFP) to rule out false negatives due to presence of PCR inhibitors in the clinical
116 samples. Primers for IC-EGFP were designed using primer3 (13). Primers were synthesized by
117 IDT (Integrated DNA Technologies), and 100 µM stocks were prepared for each primer by
118 suspending the lyophilized primer in nuclease free water. Aliquots (5 µl) of each primer stock
119 (100 µM) were pooled together to create a single primer pool.

120 **Synthetic controls**

121 Synthetic controls were created and used to optimise and validate the targeted next-generation
122 sequencing approach. These controls were generated by flanking random sequences with target-
123 specific primer binding sites. Random sequences with equal nucleotide frequency were generated
124 using the RSAT random sequence web tool (http://rsat.sb-roscoff.fr/random-seq_form.cgi) and
125 then flanked with target-specific primer binding sites (14). The size of each synthetic control was
126 designed to match the actual size of the target (Supplemental File S3). Two to four of these
127 synthetic controls were integrated into a pUCIDT plasmid containing an ampicillin resistance
128 gene, creating a final total of 18 plasmids. *E. coli* DH5-alpha was transformed with these
129 plasmids, and control stocks were stored at -80 °C. Subsequently, plasmids were extracted from
130 overnight *E. coli* cultures in lysogeny broth (LB) containing 100 µg/ml ampicillin using Qiagen
131 QIAprep Spin Miniprep Kit (Qiagen, Cat no. 27104) and plasmid DNA extractions were
132 normalized to 5 ng/µl for further use. Synthetic controls were verified by conventional PCR
133 using both individual and pooled primers. Equimolar amounts of each plasmid were then pooled
134 to make a single pooled synthetic control, ensuring that equal numbers of each synthetic target
135 were in the pooled control. This pooled synthetic control was used as a template for PCR
136 optimisation and as a positive control in clinical sample analysis.

137 **PCR optimisation with synthetic controls**

138 The aim of this experiment was to determine the optimal multiplex PCR kit, annealing
139 temperature, and primer concentration for the multiplex PCR.Two high-fidelity DNA
140 polymerase kits (Invitrogen Platinum superfi II – Themofisher Cat no. 12361010 – (TF_Plati),
141 and NEB Q5 High-Fidelity 2X Master Mix – NEB Cat no. M0492L (NEB_Q5)) and three
142 annealing temperatures (57 °C, 60 °C, 63 °C) were evaluated. Multiplex PCR was performed
143 with six different combinations of master mixes and annealing temperatures and a primer
144 concentration of 0.015 µM. The synthetic control pool (2 µl) was used as a template. Resulting
145 PCR products (100 ng each sample) from each combination was barcoded with ONT Native
146 barcoding kit (SQK.NBD114.96) and the resulting pooled library was then sequenced on a
147 GridION sequencer using an R10.4 flowcell.

148 **Reverse-transcription**

149 Following the results from the PCR optimisation experiment, a reverse transcription step was
150 added to the protocol prior to the multiplex PCR with pooled primers to enable the detection of
151 *BVDv* (an RNA virus). Three different reverse transcriptase kits (NEB LunaScript Multiplex
152 One-Step RT-PCR Kit (Cat no. E1555S), NEB LunaScript RT SuperMix Kit (Cat no. E3010L)
153 and NEB LunaScript RT Master Mix Kit Primer-free (Cat no. E3025L) with *BVDv* specific
154 primers) were tested using manufacturer's instructions with two *BVDv* positive clinical samples.

155 **Clinical samples**

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

175 **BovReproSeq workflow**

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

179 ***Total nucleic acid extraction***

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

189 ***First strand cDNA synthesis***

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

196 ***Multiplex PCR***

197 The mastermix was prepared by combining 12.5 µl of NEB Q5 Hot Start High-Fidelity 2 ×
198 Master Mix (New England Biolabs Cat no. M0492L), 9.5 µl of pooled primers (37 primer sets
199 targeting 17 pathogens, final concentration: 0.5 µM for each primer), 0.9 µl of nuclease-free
200 water, and 0.1 µl of IC-EGFP (150 copies/µl). For each PCR reaction, 2 µl of cDNA/DNA was
201 added to 23 µl of the master mix. The PCR was carried out with the following cycling
202 conditions: initial denaturation at 98°C for 30 seconds, followed by 35 cycles of denaturation at

203 98°C for 10 seconds, annealing at 57°C for 10 seconds, extension at 72°C for 30 seconds, and a
204 final elongation step at 72°C for 2 minutes. The reaction was then held at 4°C.

205 ***Magnetic bead clean-up***

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

217 ***Library preparation***

218 The sequencing library was prepared using the ligation sequencing amplicons protocol (Version:
219 NBA_9170_v114_revL_15Sep2022). The cleaned-up PCR product was then normalised to 62.5
220 ng (200 fmol based on 500 bp fragment size) and taken to end-preparation. The NEBNext Ultra
221 II End repair/dA-tailing Module (NEB, Cat no. E7546) was used to add dA-tails to the ends of
222 the amplicons, followed by magnetic bead clean up and elution with 12 µl of nuclease free water.
223 Subsequently, unique barcodes from the Native Barcoding Kit 96 V14 (ONT, Cat No.
224 SQKNBD114.96) were ligated to the end-prepped amplicons using NEB Blunt/TA Ligase
225 Master Mix (NEB, Cat no. M0367). After barcoding, samples with unique barcodes were pooled
226 together and cleaned up using magnetic beads. Sequencing adapters were then ligated to the
227 pooled samples using the NEBNext Quick Ligation Module (NEB, Cat no. E6056), followed by
228 magnetic bead clean up. Finally, the resulting library was quantified using Invitrogen Qubit 4
229 Fluorometer with the (Qubit 1x dsDNA HS Assay Kit (Thermofisher, Cat no. Q33231)).

230 ***Sequencing***

231 The prepared library (20 ng (~ 65 fmol)) was loaded onto a ONT R10.4.1 flowcell and
232 sequenced on a GridION platform for a minimum duration of 15 hours. Super-accurate
233 basecalling option in MinKNOW was used for basecalling. Only reads with barcodes on both
234 ends were used and barcodes were trimmed while sequencing using the trim barcodes option in
235 MinKNOW software (Versions used: 23.04.5, 23.04.6, 23.07.5, 23.07.12, 23.11.3)

236 ***Bioinformatics analysis***

237 A Nextflow bioinformatics pipeline was developed for the analysis of sequencing data and
238 interpretation of results (Figure 2). The path to a directory containing subdirectories with fastq
239 files and the path to a Kraken2 Plus-PF-8 database (k2_pluspf_08gb_20230605) (15) were used
240 as the inputs for the pipeline. Individual fastq files were merged followed by quality filtering
241 (q20) for each sample using nanoq (version- 0.10.0) (16). The resulting merged file was then
242 mapped to a reference sequence containing the amplicon sequences using minimap2 (version
243 2.25)(17). SAMtools (version-1.17) was used to generate read statistics and split the bam file
244 with aligned reads into amplicon-specific bam files (18). Primer trimming was performed on
245 both ends of the sequences using ampliconclip option in SAMtools (version-1.17). Consensus
246 sequences were generated on the split bam files with a minimum mapping quality of 30 (19) and
247 a base quality of 10 using SAMtools (version-1.17) (18). These generated consensus sequences
248 were then polished using medaka (version-1.11.3) (20) and compared against a custom Blast
249 database containing the BovReproSeq amplicon sequences using ABRicate (version-1.0.1) (21,
250 22) to determine the presence or absence of the pathogen and internal amplification control (IC-
251 EGFP). Concurrently, taxonomic classification was performed on the raw reads using Kraken2
252 (version-2.1.3) (15), to identify the unmapped reads and to support the results from ABRicate
253 (version-1.0.1). The results from kraken2 (version-2.1.3) were visualized as an interactive pie
254 chart using krona (version-2.7.1) (23). Output data from these analyses are collated into a HTML
255 report. All the quality metrics from SAMtools and Nanoplot (version-1.42.0) were captured and
256 visualised using multiQC (version-1.14) (24, 25). The pipeline, along with detailed information,
257 is accessible on GitHub at https://github.com/dhineshp565/BovReproSeq.

258 **Real-time PCR**

259 Real-time PCR was performed on clinical samples to verify results from BovReproSeq. Total
260 nucleic acid was extracted using Applied Biosystems MagMAX CORE Nucleic Acid
261 Purification Kit (Cat no. A32700) and used as the template for real-time PCR. For DNA-based
262 assays, real-time PCR was performed using Applied Biosystems TaqMan Fast Advanced Master
263 Mix (Cat No: 4444965), with specific primers and TaqMan probes tailored to each pathogen.
264 Detection of Bovine Viral Diarrhea Virus (*BVDv*) was carried out using the ThermoScientific
265 VetMAX-Gold BVDV PI Detection Kit (Cat no: 4413938).

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

267 Read-count thresholds represent the minimum number of reads required to classify a sample as
268 positive or negative. To determine whether a read-count threshold of 10 reads is optimal for
269 classifying a sample as positive for a target, sequence data from clinical samples were analyzed
270 using the BovReproSeq pipeline with 12 different thresholds (ranging from 1 read to 100 reads).
271 The results from each threshold were then compared to the available real-time PCR results. For
272 each read-count threshold, a contingency matrix (2x2 table) was created to check the agreement
273 of results between the two methods. Due to the imbalance between positive and negative results,
274 a precision-recall curve was used instead of a receiver operator characteristic curve to determine
275 the optimal read-count threshold (26, 27). Performance metrics, including sensitivity, specificity,
276 accuracy, and precision were calculated for each read-count threshold using the data from 2 x 2
277 contingency table. Precision-recall curves were plotted to evaluate the performance of the
278 BovReproSeq method across different thresholds. The F1 score (harmonic mean of precision and
279 recall) was used to determine best read-count threshold. An F1 score of 1 represents perfect
280 precision and recall, indicating optimal model performance, and a F1 score of 0 occurs when
281 either precision or recall (or both) is zero, indicating poor model performance (28). Matthew's
282 correlation coefficient (MCC) was calculated to determine the overall performance of the assay.
283 MCC is a specialized application of Pearson's correlation coefficient, designed for imbalanced
284 binary classification scenarios (28, 29). It incorporates all four categories of the contingency
285 matrix (True Positives, False Positives, True Negatives, and False Negatives) to provide a
286 comprehensive score of overall performance (28, 29). The contingency table, Precision-Recall
287 curves, F1 score and MCC were generated using custom Python scripts with the following

288 libraries: pandas for data manipulation (30), scikit-learn for performance metric calculations
289 (31), and matplotlib for visualization (32).

290 Results

291 PCR optimization with synthetic controls

292 The BovReproSeq method was developed and optimized using synthetic controls before testing
293 the approach on clinical samples. Pooled plasmids containing synthetic control sequences for 12
294 pathogens and one internal amplification control (IC-EGFP) were used as a template to test two
295 different multiplex PCR kits (NEB_Q5 and TF_Plati), and three different annealing
296 temperatures. The total mapped reads obtained from each amplification protocol were as follows:
297 NEB_Q5_57°C: 439,971, TF_Plati_57°C: 409,795, NEB_Q5_60°C: 470,324, TF_Plati_60°C:
298 414,916, NEB_Q5_63°C: 458,083, and TF_Plati_63°C: 486,733. The sequencing depth of each
299 amplicon varied significantly in each combination (Figure 3). The NEB_Q5 with an annealing
300 temperature of 57°C yielded a higher read depth for most of the amplicons (17 out of 29) and a
301 balanced performance, with reads evenly distributed across different amplicons compared to
302 other treatments, which showed skewed distributions favoring specific amplicons (Figure 3). As
303 a result, the NEB Q5 High-Fidelity 2x Master Mix and an annealing temperature of 57°C were
304 used in subsequent experiments. The read depths for *Campylobacter fetus* targets *glyA, aspA*, and
305 *Ureaplasma diversum 16S rRNA* were particularly low (<500 reads) compared to other targets in
306 all six combinations of annealing temperatures and polymerases. To address this, primer
307 concentration of each primer was increased to 0.5 µM and new primer pool was made. The
308 multiplex PCR was then repeated with NEB Q5 High-Fidelity 2x Master Mix with an annealing
309 temperature of 57°C and new pooled primers (0.5 µM), which resulted in a significant increase
310 in the number of reads mapping to *Campylobacter fetus glyA* (40 reads to 1047 reads), *aspA* (137
311 reads to 2670 reads), and *Ureaplasma diversum 16S rRNA* (44 reads to 1214 reads) (Figure 4).
312 As a result, primer concentration of 0.5 µM was used in subsequent experiments. Following
313 these optimizations, five additional pathogens (*Toxoplasma gondii*, *Sarcocystis* spp., *Bacillus*
314 *licheniformis, Trueperella pyogenes and Mycoplasma bovis)* were added to the panel, and new

315 synthetic controls were created for these pathogens. The updated pooled synthetic control was
316 tested, and all 37 synthetic targets were detected.

317 A reverse transcriptase step was then added to the multiplex PCR step to enable the detection of
318 *BVDv* (RNA virus), and three different reverse transcriptase kits were tested on two *BVDv*
319 positive clinical samples. Gel electrophoresis results from the NEB LunaScript Multiplex One-
320 Step RT-PCR Kit (Cat no. E1555S) indicated a significant amount of short non-specific
321 amplifications, so the PCR product from this kit was not sequenced. PCR products from the NEB
322 LunaScript RT SuperMix Kit (Cat no. E3010L) and the NEB LunaScript RT Master Mix Kit
323 (Primer-free, Cat no. E3025L) with *BVDv*-specific primers were sequenced. The NEB
324 LunaScript RT Master Mix Kit (Primer-free) yielded 10,598 *BVDv*-specific reads in sample-1
325 and 7,385 *BVDv*-specific reads in sample-2. In contrast, the NEB LunaScript RT SuperMix Kit
326 with random hexamers produced 1,170 *BVDv*-specific reads in sample-1 and 2,567 *BVDv*-
327 specific reads in sample-2 (Table S4). A control experiment without the RT step showed no
328 detection of *BVDv* in either sample. Despite the lower read depth for *BVDv*, the NEB LunaScript
329 RT SuperMix Kit (Cat no. E3010L) was selected because it successfully amplified *BVDv* RNA
330 in clinical samples and allows for the potential detection of additional pathogens not included in
331 the panel, due to the presence of random hexamers in the mix.

332 **IC-EGFP titration and repeatability**

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

338 To determine the optimal quantity of IC-EGFP to be spiked into each sample, a titration
339 experiment was conducted using two concentrations of IC-EGFP: 150 copies/reaction and 15
340 copies/reaction. A mock *Neospora caninum* positive sample was prepared by spiking 10 µl of
341 *Neospora caninum* ATCC 50845 culture into 90 µl of placenta matrix (clarified tissue
342 homogenate) confirmed negative by current diagnostic methods (Table 1). This mock sample

343 was serially diluted (10-fold dilution) using placenta matrix (clarified tissue homogenate) as the
344 diluent to maintain consistent matrix composition across dilutions. Total nucleic acid
345 (DNA/RNA) was extracted from the dilutions, and BovReproSeq PCR was performed in
346 triplicate for each of the two IC-EGFP concentrations. For each concentration, a sequencing
347 library was prepared and sequenced on separate flow cells. *Neospora caninum* real-time PCR
348 was performed in triplicate to determine the Ct value of each dilution.

349 Both pan-Coccidia and *Neospora caninum*-pNC-5 gene targets were consistently detected in all
350 replicates and dilutions up to a Ct value of 32 (10^{-5} dilution) in both experiments demonstrating
351 the repeatability of the assay (Figure 5A, 5B). However, for dilutions with Ct values above 32,
352 detection of both *Neospora caninum* targets was inconsistent, occurring only in one or two of the
353 replicates. As the concentration of *Neospora caninum* decreased in the dilutions, the number of
354 IC-EGFP reads increased in both experiments. The IC-EGFP reads were at their lowest when the
355 number of reads and concentration of *Neospora caninum* were at their highest (Ct 18.25).
356 Specifically, in the 15 copies/reaction experiment, there were 23 IC-EGFP reads out of 42,275
357 total reads, while in the 150 copies/reaction experiment, there were 2,081 IC-EGFP reads out of
358 51,175 total reads. In both experiments, IC-EGFP reads stabilized around a dilution of 10^{-5}. The
359 experiment with 15 copies of IC-EGFP per reaction demonstrated better read depth for *Neospora*
360 *caninum*, particularly in dilutions with lower pathogen load (Ct 36.56). At this dilution, the 15
361 copies/reaction experiment yielded 4,021 *Neospora caninum* reads out of 12,932 total reads,
362 compared to only 21 *Neospora caninum* reads out of 85,252 total reads in the 150 copies/reaction
363 experiment. Moreover, *Neospora caninum* was detected by BovReproSeq in dilutions 10^{-7} and
364 10^{-8} which were not detected with real-time PCR. Although these results suggest that the
365 BovReproSeq method is more sensitive than real-time PCR, it is important to note that these
366 experiments used samples spiked with live organisms. The sensitivity may vary in actual clinical
367 samples, where factors such as sample type, sample matrix and presence of inhibitors could
368 influence detection limits. Based on these results, 15 copies of IC-EGFP per reaction was
369 selected as the optimal concentration for subsequent experiments, as it provided a better balance
370 between internal amplification control and pathogen detection.

371 **Specificity evaluation**

372 A total of 29 clinical samples, both positive and negative for BovReproSeq targets, along with
373 no-template controls (NTC) and pooled BovReproSeq synthetic positive controls, were tested to
374 determine the specificity of the assay. Results from the BovReproSeq method matched current
375 diagnostic methods for all samples (Table S5). However, most of the samples showed a high
376 abundance of reads mapping to the *Campylobacter jejuni* amplicon (cjejuni23s) (Table S5).
377 Manual analysis of the consensus sequence using BLAST revealed that these reads did not
378 correspond to *Campylobacter jejuni*. Instead, they matched other *Campylobacter* species and
379 *Campylobacter*-like organisms. Similar non-specificity issues were observed with the Tpyo-
380 cpn60 target, as reads mapping to this target did not match *Trueperella pyogenes* reference
381 sequences and resulted in poor consensus sequences. Additionally, in *Mycoplasma bovis*
382 (Mbovis) positive clinical samples, the Mbovis gltX target was consistently undetected, and the
383 gspA target was detected with only a very low number of reads. Due to the non-specificity of
384 cjejuni23s, Tpyo-cpn60, and the poor performance of both Mbovis targets, these primers were
385 replaced with more specific targets to improve the overall assay performance. Synthetic controls
386 were developed for the newly added targets, and a new set of pooled controls was made (Table
387 Sl). This updated pooled synthetic control and the clinical samples were then tested with
388 BovReproSeq method. All 37 synthetic controls including the replacements were detected from a
389 single PCR reaction (Figure 6).

390 **Performance on clinical samples**

391 To evaluate the performance of the BovReproSeq method against current diagnostic methods
392 (Table 1), 116 clinical samples with existing results (for tests requested by veterinarians and
393 pathologists) were tested using the BovReproSeq approach. BovReproSeq PCR was performed
394 and PCR amplicons from these samples were pooled into four libraries (~30 samples per library),
395 each library loaded onto a ONT R10.4.1 flow cell, and sequenced. The data was analyzed using
396 the BovReproSeq bioinformatics pipeline. The number of raw reads per sample varied
397 significantly, ranging from 6,636 to 642,695 reads, with a median of 110,509 reads. The median
398 total bases generated per sample was 105.5 Mb, and the median raw read quality score was 19.6.
399 Positive controls had reads and consensus sequences mapping to all 37 synthetic controls, while
400 negative controls (NTCs) were negative for all pathogens. Two samples lacked IC-EGFP reads,

401 indicating potential PCR inhibition, and were omitted from further analysis. The results from the
402 BovReproSeq method matched the existing results from current methods for approximately 93%
403 of the samples (139/150, Table 2). The kappa value of 0.85 indicated almost perfect agreement
404 between the two methods, however, there were 11 false negatives. The Ct values from the
405 current diagnostic PCRs for positive samples ranged from 12 to 37, with all false negatives
406 having Ct values between 33 and 37, which is consistent with low pathogen load (Figure 7).
407 BovReproSeq method successfully differentiated *Campylobacter fetus* subspecies in 6 cases and
408 identified the MLST type in five of them (Table S2). Additionally, BovReproSeq identified
409 *Pestivirus tauri (BVDv*-type 2) in two samples among those tested for *BVDv* (Table S2).

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

416 The presence of additional pathogens and the detection of pathogens in samples where the initial
417 test request was negative indicated that to comprehensively evaluate the BovReproSeq method,
418 each sample had to be tested for all 17 pathogens using conventional methods. This would
419 provide a true comparison between the current methods and the BovReproSeq method. This was
420 not possible as some of the samples used were archived nucleic acid samples collected over a
421 long period, limiting our ability to apply methods other than PCR. Real-time PCR assays were
422 available for only 10 BovReproSeq pathogens (*BVDv, Campylobacter jejuni, Campylobacter*
423 *fetus* subsp. *venerealis, Neospora caninum, Toxoplasma gondii, Coxiella burnetii, Chlamydia*
424 *abortus, Tritrichomonas fetus, Ureaplasma diversum, Mycoplasma bovis*). Therefore, real-time
425 PCR was performed on samples (109 samples) that had enough nucleic acid to detect the
426 presence of these 10 pathogens, providing 1090 results. Results from each PCR were collated
427 into a single results table (1090 PCR results). The compiled real-time PCR results for 10
428 pathogens across 109 clinical samples showed 1016 negative and 74 positive outcomes. These
429 real-time PCR results were then compared to the results from the BovReproSeq method. The

430 results from the BovReproSeq method matched the current methods for 98.8% of the samples
431 (1077/1090, Table 4). There were 13 false negative results. All false negative results occurred in
432 samples with Ct values between 33 and 37 except for one sample with a Ct value of 29 for
433 *BVDv.* In this case, four other pathogens (*BoHV-1, Mycoplasma bovis, Coxiella burnetii* and
434 *Trueperella pyogenes*) were detected in high abundance, and the pathologist had reported lesions
435 characteristic of *BoHV-1*. The kappa value of 0.90 indicated almost perfect agreement between
436 the two methods. Due to the imbalance between positive and negative results, the kappa value
437 may not accurately reflect the performance of the assay *(33)*. Therefore, Matthew's Correlation
438 Coefficient (MCC) was calculated. MCC value of approximately 0.90 indicated excellent overall
439 performance of the assay.

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

441 Precision (positive predictive value) and recall (sensitivity) were calculated for read-count
442 thresholds from 1 to 100 reads as described in the Methods, and a precision-recall curve was
443 plotted (Figure 8). At the lowest threshold of one read, precision was at its lowest around 48%,
444 while sensitivity peaked at 95%. Increasing the read-count threshold to five reads resulted in a
445 substantial improvement in precision (86%) but a marked reduction in sensitivity (85%).
446 Precision reached 100% at a threshold of 10 reads and remained consistent across remaining
447 thresholds, but sensitivity declined with increasing read-count thresholds and was lowest (70%)
448 at a threshold of 100 reads. The threshold of 10 reads yielded the optimal balance between
449 precision and recall with the maximum F1 score of 0.90. At the read-count threshold of 10 reads,
450 the sensitivity of the assay was approximately 82%, while specificity was 100%. The positive
451 predictive value was 100%, and the negative predictive value was 98.7%. The overall accuracy
452 of the assay was 98.8%. Based on these results, a read count threshold of 10 was determined to
453 be optimal for the BovReproSeq method.

454 # Discussion

455 Our study demonstrates the development and validation of a syndromic next-generation
456 sequencing panel (BovReproSeq) designed to simultaneously detect 17 pathogens associated

457 with bovine reproductive failure. Initial testing with a pooled synthetic control successfully
458 identified all 37 synthetic targets in the products of a single PCR reaction, demonstrating the
459 multiplexing ability of the assay. The results from BovReproSeq testing of 116 clinical samples
460 matched with existing results from current diagnostic methods for 93% of samples. Additionally,
461 the BovReproSeq method detected pathogens not originally requested for testing, indicating
462 possible mixed infections (Table S2). This ability of the BovReproSeq to identify additional
463 pathogens highlights its advantage over current single-plex diagnostic methods by providing
464 broader detection without iterative testing.

465 In addition to detection of 17 pathogens, BovReproSeq method enables the typing of *BVDv*
466 (*Pestivirus bovis* (*BVDv* type-1) and *Pestivirus tauri (BVDv* type-2)) based on differences in 5′-
467 UTR sequences of *Pestivirus*. The 5′-UTR sequences could also be used to detect *Pestivirus*
468 *braziliense* and *Border disease virus* in sheep and goats. With *Campylobacter fetus*, in addition
469 to detection and differentiation of subspecies, we are also able to identify the sequence type
470 using MLST (Table S2), which aids in epidemiological studies. Since this assay involves random
471 hexamers and multiple primers, there is potential to detect pathogens not included in the panel,
472 further broadening the scope of the assay. The entire BovReproSeq workflow (Figure 1) can be
473 completed within 48 hours of receiving the sample and can be cost-effective if multiple samples
474 are processed within a single flow cell.

475 The BovReproSeq method has some limitations, particularly in samples with high Ct values
476 (close to or above 35) indicating that very low pathogen concentrations in the sample may not be
477 detected (Figure 7). This finding is consistent with a similar study in which a targeted next-
478 generation sequencing method was not able to detect vector-borne pathogens in canine samples
479 with Ct values greater than 35 (10). It is worth noting that PCR artifacts and cross-contamination
480 can sometimes lead to amplification at Ct values >35, and samples with Ct values >35 are often
481 considered "suspect" rather than "positive" in diagnostic laboratories. Moreover, pathogens like
482 *Coxiella burnetii* and *Chlamydia abortus* are known to occur in healthy animals in low numbers
483 and their presence in low quantities without associated lesions is generally not considered as the
484 cause of the abortion (34). In most infectious abortion cases, high pathogen loads with
485 characteristic lesions or signs are typical. Thus, the failure of BovReproSeq to detect very low

486 abundance pathogens may not be a clinically significant limitation. Notably, we had only one
487 sample (Sample ID - 4122-2) with a Ct value of 29 for *BVDv* where BovReproSeq failed to
488 detect the virus. In this case, four other pathogens in the BovReproSeq panel (BoHV-1,
489 *Mycoplasma bovis, Coxiella burnetii* and *Trueperella pyogenes*) were detected in high
490 abundance, and the pathologist had reported lesions characteristic of BoHV-1 (Table 3).
491 Similarly, we frequently detected pathogens such as *Ureaplasma diversum* and *Trueperella*
492 *pyogenes*, and in a couple of cases, their abundance had a negative impact on the detection of the
493 causative pathogen. This suggests that multiple pathogens in high abundance in a sample may
494 impede the detection of pathogens in low abundance. This limitation could be overcome by
495 removing primers for *Ureaplasma diversum* and *Trueperella pyogenes* from the main pool and
496 placing them in a separate PCR reaction, which could reduce their impact on the detection of
497 other targeted pathogens.

498 The analytical sensitivity of our BovReproSeq method is lower compared to PCR assays. The
499 sensitivity of detection of specific pathogens in highly multiplexed assays are significantly
500 impacted by the amount and number of other targets that are in the sample (35). This limitation
501 could be addressed by reducing the number of samples per flow cell, extending sequencing time
502 to improve the sequencing depth of low abundance targets and by using high-throughput
503 sequencers such as ONT PromethION, Illumina MiSeq or NovaSeq. The one-pot multiplex PCR
504 could also be split into multi-pot PCR with smaller primer pools to improve sensitivity, and the
505 number of PCR cycles could be increased from 35 to 40 to increase the number of copies of the
506 amplicons. Regarding primers and PCR target selection, selecting specific genes of pathogens
507 rather than conserved housekeeping genes was more successful, as the primers targeting
508 conserved genes (*T. pyogenes* cpn60 and *C. jejuni* 23S rRNA gene) were more likely to amplify
509 sequences from related organisms present in the sample and obscure the targeted pathogens of
510 interest. This observation should be factored into any future additions to or modification of the
511 panel.

512 Similar studies have been published recently, but our research presents a few methodological
513 differences and advancements (Anis et al., 2018; Kattoor et al., 2022). Anis et al. developed a
514 targeted NGS method for the simultaneous detection of 43 common bovine and small-ruminant

515 pathogens. Their primer design was a collaboration with the White Glove Team (Ion Torrent;
516 Thermofisher Scientific) using Ion Ampliseq Designer (Ion Torrent; Thermofisher Scientific,
517 Waltham, MA). Kattoor et al. focused on the comprehensive detection of 17 vector-borne
518 pathogens in canines. The primers in their study were designed with Thermofisher Scientific's
519 AgriSeq Bioinformatics team, and the primer sets are now proprietary items of Thermofisher
520 (Kattoor et al., 2022). Both studies used Ion Torrent Suite and Geneious software for
521 bioinformatics and neither includes an internal amplification control. In our study, we have used
522 previously published primers and further evaluated them based on *in silico* PCR results.
523 Additionally, in the BovReproSeq bioinformatics pipeline, we have used open-source tools and
524 packaged them as one Nextflow pipeline that produces a readable results report and the pipeline
525 can be scaled up to run on any platform (36). Our pipeline is currently available on GitHub, and
526 the primer information is included in the supplementary section (Table S1) to ensure open access
527 to our methods and facilitate reproducibility.

528 The previously published bovine and canine pathogen sequencing panels used Thermofisher Ion
529 Torrent technology with IonChef for sequencing multiplex PCR products. This second-
530 generation sequencing technology has limitations including short reads (<400 bp), low
531 throughput and poor accuracy in homopolymer regions. BovReproSeq uses Oxford Nanopore
532 technology (ONT), a long-read approach that enables sequencing of longer gene segments than
533 the Ion Torrent platform and allows for real-time sequencing and analysis, resulting in faster
534 turnaround times. Additionally, BovReproSeq uses only full-length reads to generate the
535 consensus sequences resulting in highly accurate results. The use of long reads allows us to
536 perform MLST typing, as most MLST fragments are around 600 bp, which would be challenging
537 with Ion Torrent technology. Although the raw read accuracy with ONT was known to be lower
538 compared to other platforms, the raw read quality has significantly improved with the newest
539 version of the R10 nanopore and the associated kit 14 chemistry (37, 38). It is also important to
540 note that the raw read accuracy of long homopolymer regions is still poor with ONT; however,
541 this is not an issue with BovReproSeq, as the amplicons are relatively short and do not contain
542 long homopolymer regions. In our BovReproSeq pipeline, we exclusively used raw reads with a
543 quality score of 20 or above (>99% accuracy) to ensure that only highly accurate reads were used
544 for downstream analysis.

545 The incorporation of an internal amplification control in the BovReproSeq assay allows the rule
546 out of false negatives due to PCR inhibition. The amount to be spiked into each reaction was
547 determined experimentally and to ensure repeatability, we performed this assay in triplicates on
548 serial dilutions with consistent results between the replicates (Figure 5A, 5B). Additionally, the
549 pooled synthetic positive control for the assay allows the detection of positive control
550 contamination since the sequences of the positive control amplicons are unrelated to the actual
551 biological targets. In contrast, using the actual pathogen sequence as a positive control could
552 complicate interpretation. If contamination occurs from the positive control, a positive result
553 could be attributed to either the contamination of the assay or the genuine presence of the
554 pathogen. Therefore, using a synthetic control that shares the same primer binding sites but has
555 different sequences in between will help identify contamination while also making sure that PCR
556 is working correctly. Finally, we analysed the sequence data generated from 109 clinical samples
557 using 12 different read-count thresholds to determine the optimal threshold and the overall
558 performance of assay, significantly enhancing the reliability and robustness of our results.

559 **Conclusion**

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

568 **Data availability**

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

572 **Acknowledgements**

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

577 **References**

578 1. Mercadante VRG, Nicholas W Dias, Dias NW, C. Timlin, Timlin CL, Stefania Pancini,
579 Pancini S. 2020. 375 Economic Consequences of Pregnancy Loss in Beef Cattle. J Anim
580 Sci 98:124–124.

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

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

586 4. Anderson ML. 2012. Disorders of Cattle, p. 13–48. *In* Kirkbride's Diagnosis of Abortion
587 and Neonatal Loss in Animals. John Wiley & Sons, Ltd.

588 5. Dumkow LE, Worden LJ, Rao SN. 2021. Syndromic diagnostic testing: a new way to
589 approach patient care in the treatment of infectious diseases. J Antimicrob Chemother
590 76:iii4–iii11.

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

593 7. Chiu CY, Miller SA. 2019. Clinical metagenomics. Nat Rev Genet 20:341–355.

594 8. Prachayangprecha S, Claudia M. E. Schapendonk, Schapendonk CME, Marion Koopmans,
595 Koopmans M, Albert D. M. E. Osterhaus, Osterhaus ADME, Anita C. Schürch, Schürch
596 AC, Suzan D. Pas, Pas SD, Annemiek A. van der Eijk, van der Eijk AA, Yong
597 Poovorawan, Poovorawan Y, Bart L. Haagmans, Haagmans BL, Saskia L. Smits, Smits SL.
598 2014. Exploring the Potential of Next-Generation Sequencing in Detection of Respiratory
599 Viruses. J Clin Microbiol 52:3722–3730.

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

603 10. Kattoor JJ, Emma Nikolai, Barbara Qurollo, Barbara A. Qurollo. 2022. Targeted Next-
604 Generation Sequencing for Comprehensive Testing for Selected Vector-Borne Pathogens in
605 Canines. Pathogens 11:964–964.

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

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

617 13. Untergasser A, Ioana Cutcutache, Cutcutache I, Triinu Kõressaar, Koressaar T, Jian Ye, Ye
618 J, Brant C. Faircloth, Faircloth BC, Maido Remm, Remm M, Rozen S, Steve Rozen, Rozen
619 SG, Rozen SG. 2012. Primer3—new capabilities and interfaces. Nucleic Acids Res 40.

620 14. Kavlick MF. 2018. Development of a universal internal positive control. BioTechniques
621 65:275–280.

622 15. Wood DE, Jennifer Lu, Lu J, Ben Langmead, Langmead B. 2019. Improved metagenomic
623 analysis with Kraken 2. Genome Biol 20:257–257.

624 16. Steinig EJ, Coin L, Lachlan Coin. 2022. Nanoq: ultra-fast quality control for nanopore
625 reads. J Open Source Softw 7:2991–2991.

626 17. Li H. 2017. Minimap2: fast pairwise alignment for long nucleotide sequences. Bioinform
627 https://doi.org/10.1093/bioinformatics/bty191.

628 18. Danecek P, James K. Bonfield, Bonfield JK, Liddle J, Jennifer Liddle, Liddle J, John
629 Marshall, Marshall J, Valeriu Ohan, Ohan V, Martin Pollard, Pollard MO, Andrew
630 Whitwham, Whitwham A, Thomas M. Keane, Keane TM, Keane TM, Shane McCarthy,
631 McCarthy S, Davies SJ, Robert M. Davies, Davies RL, Li H, Heng Li, Li H. 2021. Twelve
632 years of SAMtools and BCFtools. GigaScience 10.

633 19. Murphy SG, Smith C, Lapierre P, Shea J, Patel K, Halse TA, Dickinson M, Escuyer V,
634 Rowlinson MC, Musser KA. 2023. Direct detection of drug-resistant Mycobacterium
635 tuberculosis using targeted next generation sequencing. Front Public Health 11.

636 20. Wright C, Nicholls S, Rescheneder P. 2023. nanoporetech/medaka. Python. Oxford
637 Nanopore Technologies.

638 21. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.
639 BLAST+: architecture and applications. BMC Bioinformatics 10:421.

22. Seeman T. 2015. ABRicate-Github https://github.com/tseemann/abricate.

23. Ondov BD, Nicholas H. Bergman, Bergman NH, Adam M. Phillippy, Phillippy AM. 2011. Interactive metagenomic visualization in a Web browser. BMC Bioinformatics 12:385–385.

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

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

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

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

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

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

30. The pandas development team. 2024. pandas-dev/pandas: Pandas (v2.2.3). Zenodo.

31. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. 2011. Scikit-learn: Machine Learning in Python. J Mach Learn Res 12:2825–2830.

32. Matplotlib: A 2D Graphics Environment | IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/document/4160265. Retrieved 27 October 2024.

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

34. Hazlett M, Rebeccah McDowall, McDowall RM, McDowall R, J Delay, DeLay J, Margaret Stalker, Stalker M, Beverly McEwen, McEwen B, Tony van Dreumel, van Dreumel T, Maria Spinato, Spinato M, Brian Binnington, Brian Binnington, Binnington B, Đurđa Slavić, Slavic D, Slavic D, Susy Carman, Carman S, Hugh Y. Cai, Cai HY. 2013. A

676 prospective study of sheep and goat abortion using real-time polymerase chain reaction and
677 cut point estimation shows Coxiella burnetii and Chlamydophila abortus infection
678 concurrently with other major pathogens. J Vet Diagn Invest 25:359–368.

679 35. Dabisch-Ruthe M, Vollmer T, Adams O, Knabbe C, Dreier J. 2012. Comparison of three
680 multiplex PCR assays for the detection of respiratory viral infections: evaluation of xTAG
681 respiratory virus panel fast assay, RespiFinder 19 assay and RespiFinder SMART 22 assay.
682 BMC Infect Dis 12:163.

683 36. Di Tommaso P, Maria Chatzou, Chatzou M, Evan Floden, Floden EW, Pablo Prieto Barja,
684 Barja PP, Emilio Palumbo, Palumbo E, Cédric Notredame, Notredame C. 2017. Nextflow
685 enables reproducible computational workflows. Nat Biotechnol 35:316–319.

686 37. Mantas Sereika, Rasmus Hansen Kirkegaard, Kirkegaard RH, Søren Michael Karst, Karst
687 SM, Thomas Yssing Michaelsen, Thomas Yssing Michaelsen, Emil A. Sørensen, Sørensen
688 EA, Rasmus Dam Wollenberg, Rasmus Dam Wollenberg, Mads Albertsen, Albertsen M.
689 2022. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished
690 bacterial genomes from pure cultures and metagenomes without short-read or reference
691 polishing. Nat Methods https://doi.org/10.1038/s41592-022-01539-7.

692 38. Tianyuan Zhang, Hanzhou Li, Silin Ma, Jian Cao, Hao Liao, Qiaoyun Huang, Wenli Chen.
693 2023. The newest Oxford Nanopore R10.4.1 full-length 16S rRNA sequencing enables the
694 accurate resolution of species-level microbial community profiling. Appl Environ
695 Microbiol 89.

696

697

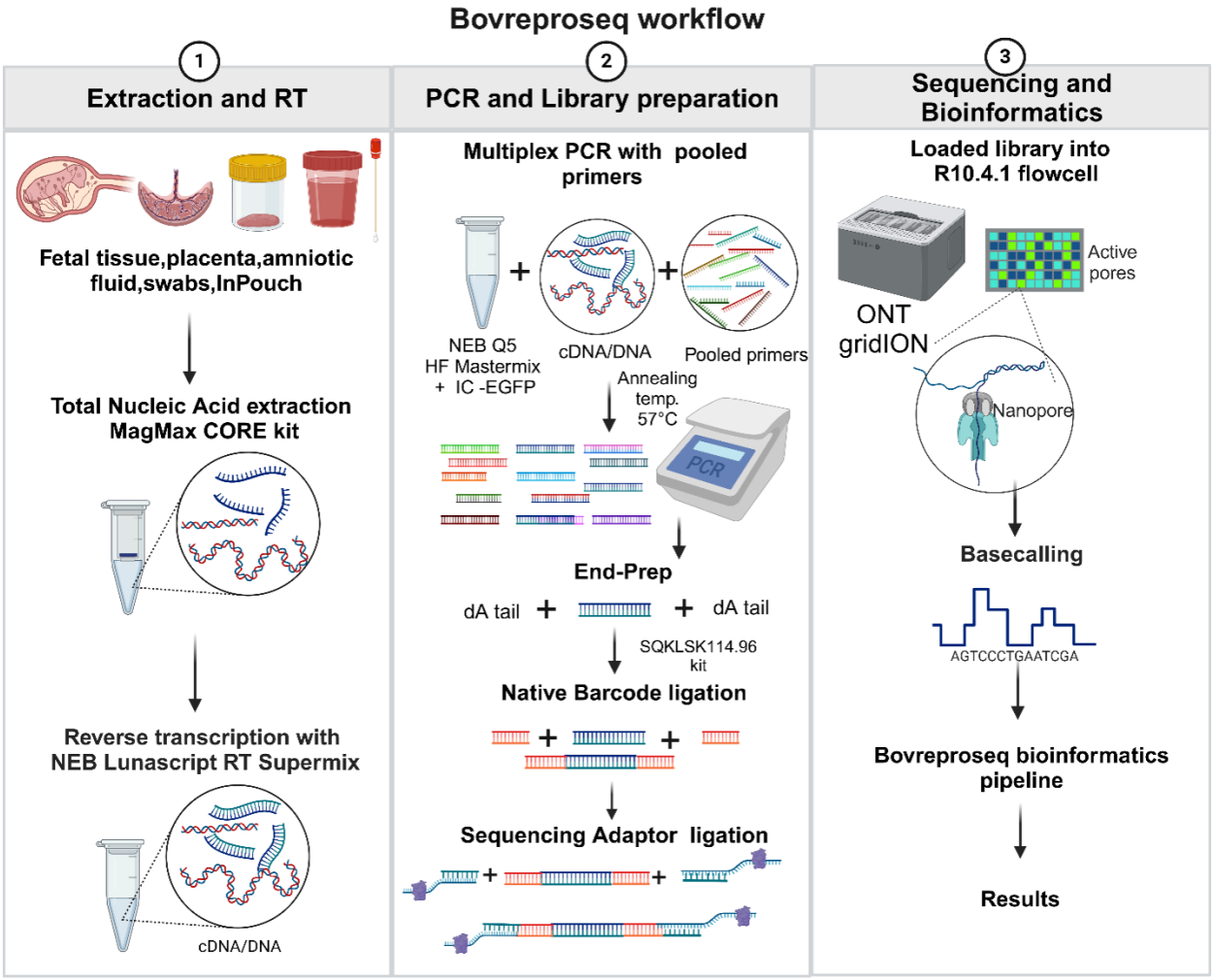


Figure 1: BovReproSeq workflow. 1) Total nucleic acid was extracted from samples using the MagMAX CORE kit, followed by a reverse transcription to enable detection *BVDv*. 2) cDNA/DNA was then used as a template for multiplex PCR with pooled primers and the generated amplicons were then prepared for sequencing using the Oxford Nanopore Ligation Sequencing Kit. 3) 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. (Created with BioRender.com)

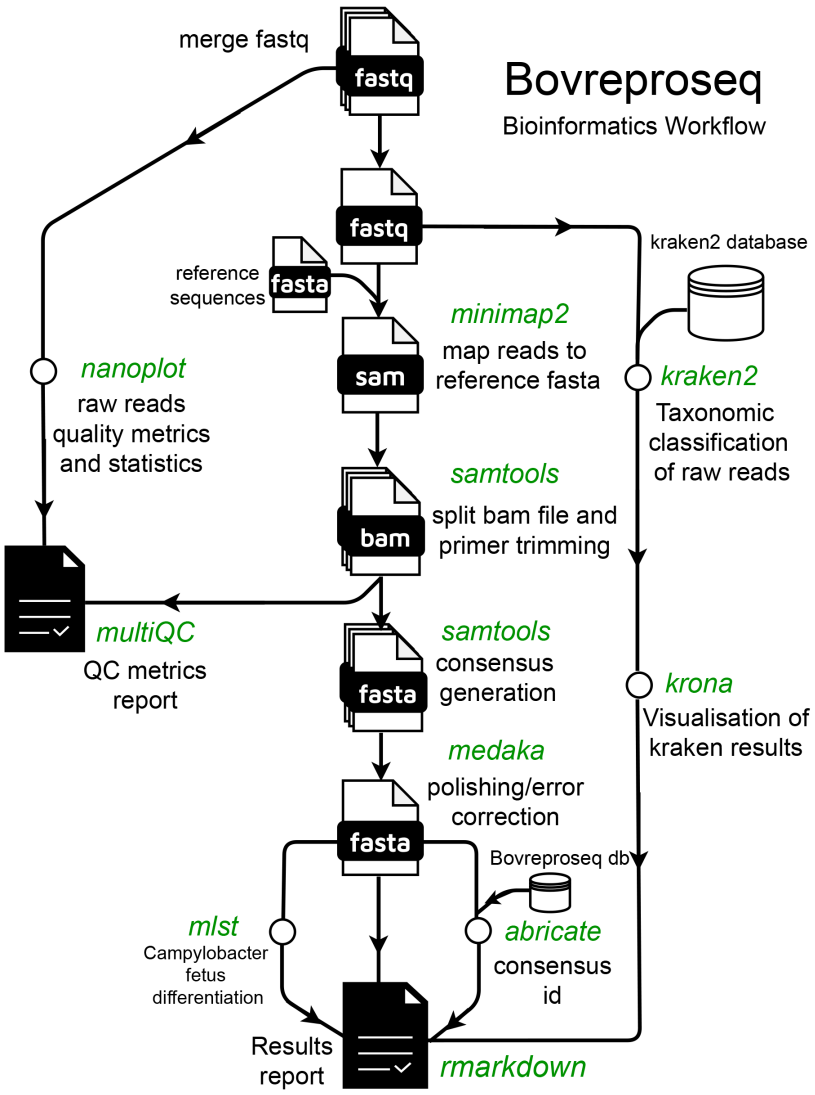


Figure 2: BovReproSeq Bioinformatics workflow. Multiple FASTQ files generated for each sample were merged into one and renamed. The merged file was then mapped against a reference using minimap2. The generated SAM file was subsequently used to generate consensus and read statistics using SAMtools. The 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.

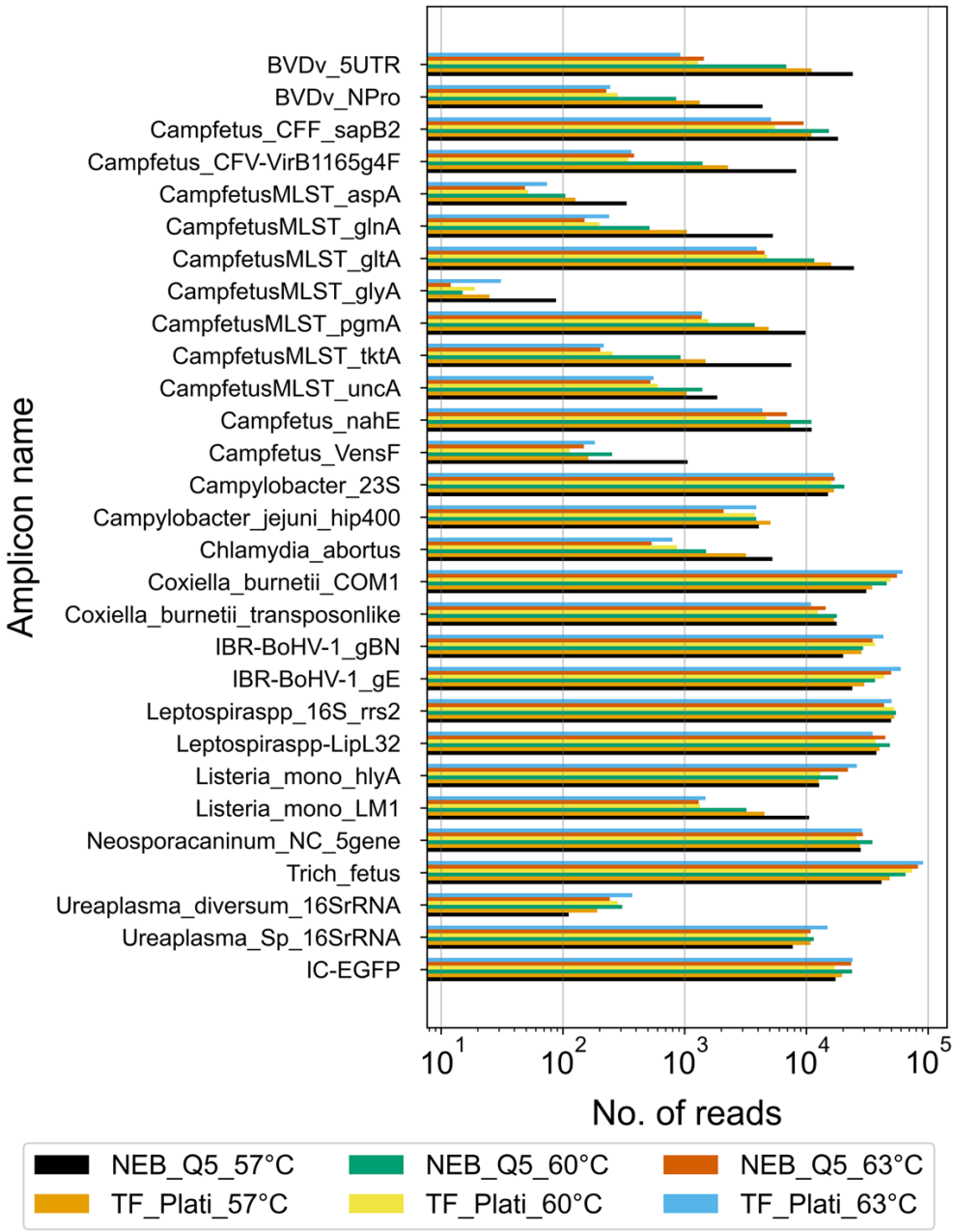


Figure 3: Comparison of sequencing depth of amplicons produced with NEB Q5 High-Fidelity 2x Master Mix and TF Invitrogen Platinum superfi ii polymerase, annealing temperatures of 57°C, 60°C and 63°C, and primer concentration of 0.015 µM. NEB Q5 High-Fidelity 2x Master Mix at 57°C consistently produced higher read depths across most amplicons. This experiment was performed without the reverse-transcription step and includes results only from synthetic controls for 12 pathogens (29 synthetic targets).

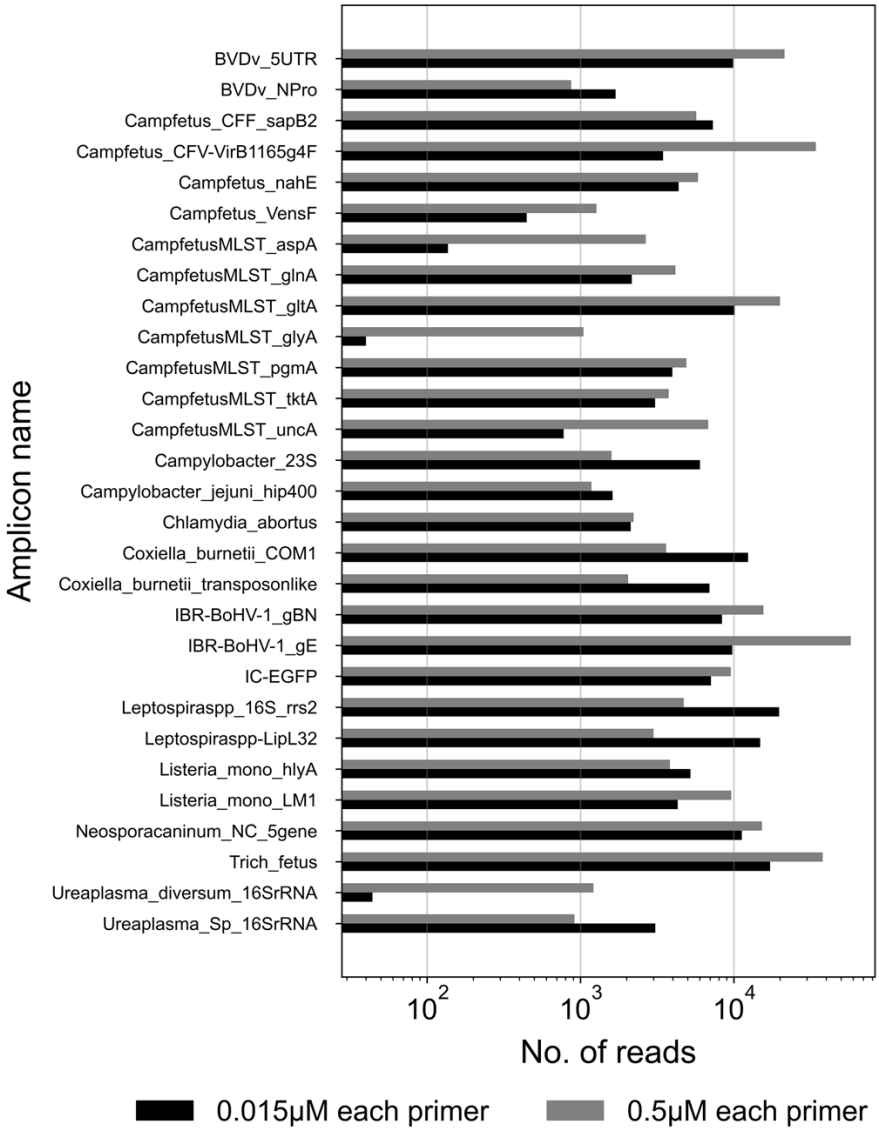


Figure 4: Comparison of sequencing depth for amplicons produced with two different primer concentrations (0.015 μM and 0.5 μM) using NEB Q5 High-Fidelity 2x Master Mix and an annealing temperature of 57°C. The higher primer concentration of 0.5 μM consistently yielded greater read depths across most amplicons compared to the lower concentration of 0.015 μM. This experiment was performed without the reverse-transcription step and includes results only from synthetic controls for 12 pathogens (29 synthetic targets).

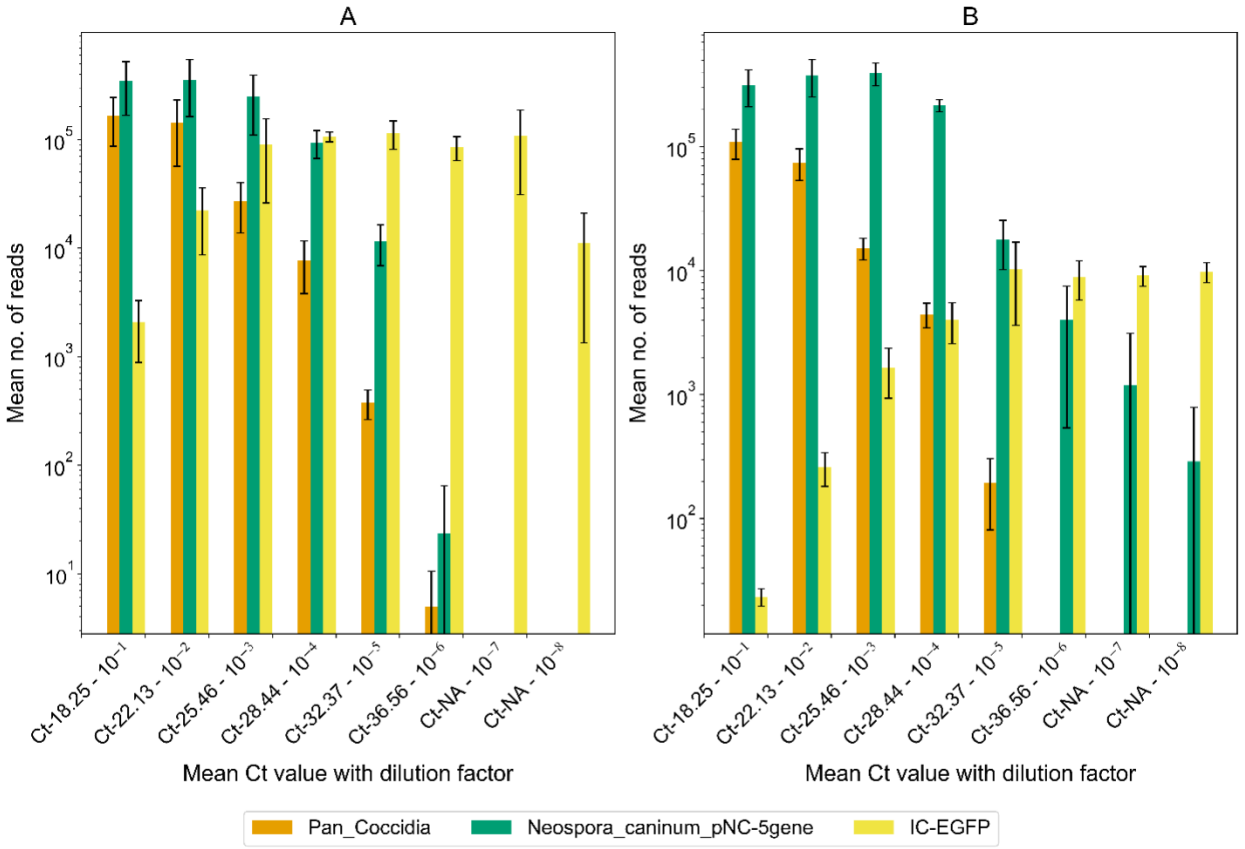
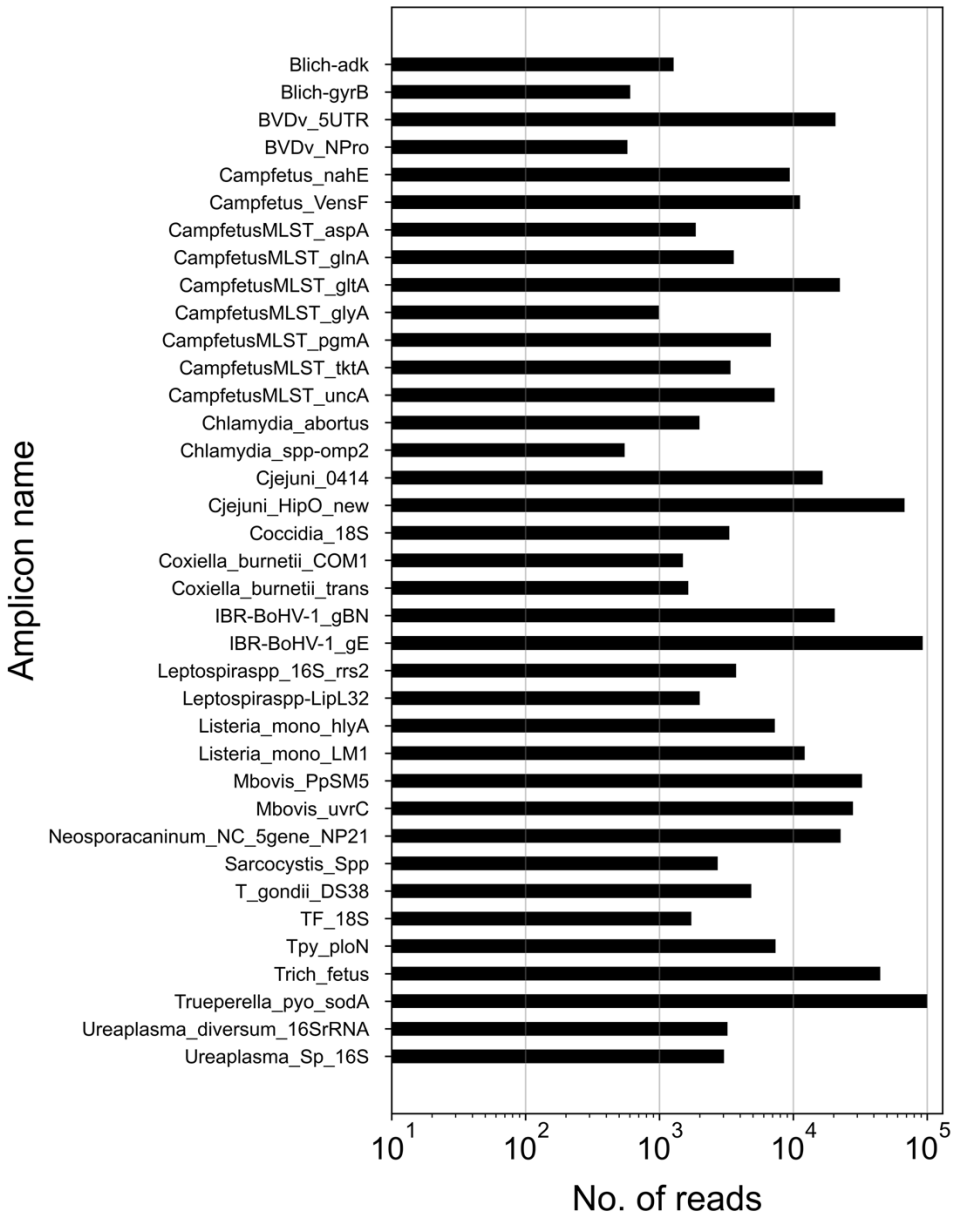


Figure 5: Comparison of mapped reads in serially diluted mock *Neospora caninum* positive samples spiked with 150 copies (A) and 15 copies (B) of IC-EGFP per reaction. Each dilution was tested in triplicate. The x-axis shows mean Ct values from real-time PCR of serial dilutions of the mock sample while y-axis represent the mean number of reads generated for each amplicon from BovReproSeq. "NA" on X-axis indicates that *Neospora caninum* was not detected using real-time PCR. Error bars represent the standard deviation of read counts across triplicates. *Neospora caninum* was consistently detected in all replicates up to Ct of 32, with inconsistent detection in subsequent dilutions. 15 copies of IC-EGFP per reaction yielded higher read depth for *Neospora caninum* in higher dilutions than 150 copies per reaction.

745



746

747 **Figure 6:** Comparison of read counts for each synthetic control from a single PCR reaction (including
748 reverse transcription step), where the final version of the pooled control was used as a template. All 37
749 synthetic targets were detected from a single PCR reaction.

750

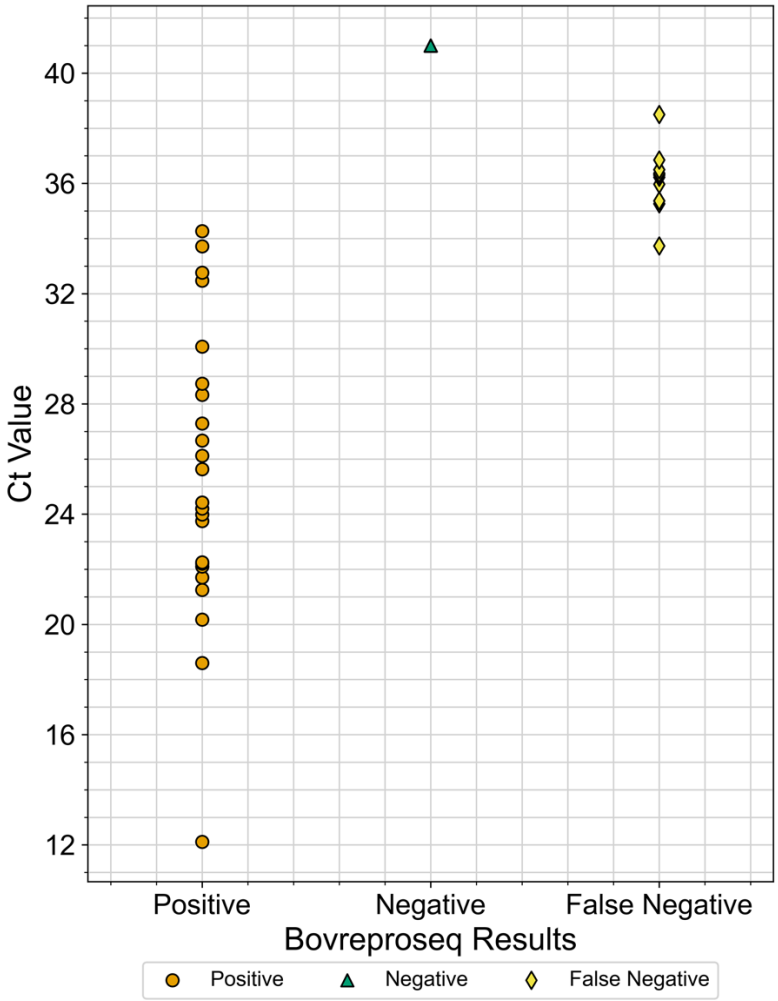


Figure 7: Real-time PCR Ct values of the samples and the corresponding results from the BovReproSeq method. Ct values for the positive samples ranged from 12 to 36, while all the false negatives had a Ct value close to 35 or above indicating that these samples had a very low pathogen load. Additionally,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

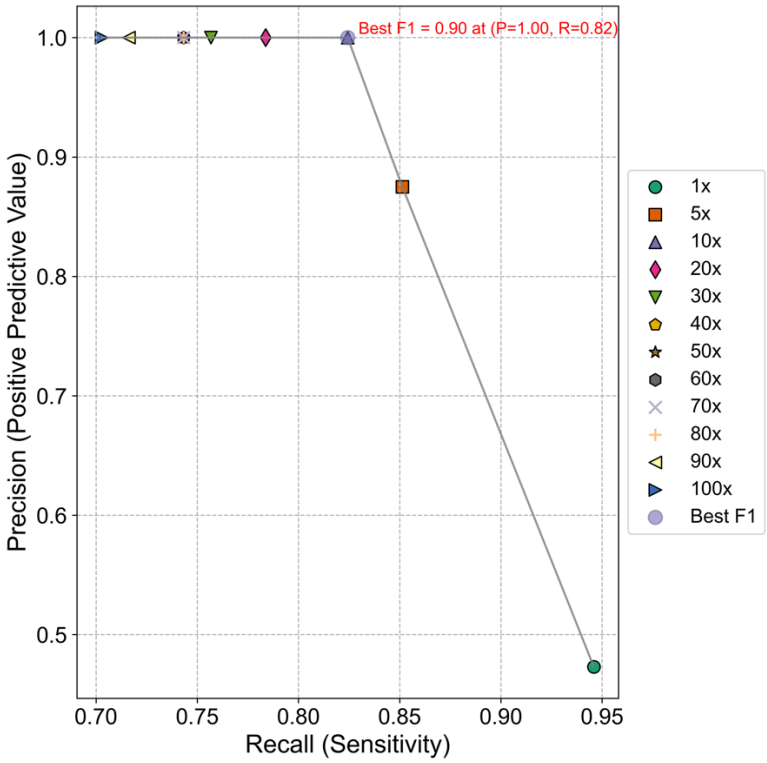


Figure 8: Precision-Recall curve. The graph shows the trade-off between precision (positive predictive value) and recall (sensitivity) at various read-count thresholds. The grey line represents the precision-recall curve, with data points corresponding to different read-count threshold (1x to 100x).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

767

768 **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

769

770

771

772 **Table 2**: Comparison of results between current laboratory methods and BovReproSeq. It includes
773 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

774

775

776 **Table 3**: Comparison of results from initial test request (current methods) and BovReproSeq method.

Sample ID	Tests requested by veterinarians (Initial test request)	Results from current methods	Results BovReproSeq
6135-2	*Chlamydia abortus, Coxiella burnetii*	Negative	*Campylobacter jejuni*
6156-2	*Chlamydia abortus, Coxiella burnetii*	Negative	*Campylobacter jejuni*
6157-2	*Chlamydia abortus*	Negative	*Campylobacter jejuni*
6783-3	*BVDv, Campylobacter fetus* subsp. *venerealis, Coxiella burnetii, BoHV-1, Tritrichomonas fetus*	Negative	*Trueperella pyogenes*
1581-2	*Coxiella burnetii*	Negative	*Chlamydia abortus, Trueperella pyogenes*
2760-2	*Coxiella burnetii*	Negative	*Campylobacter fetus* subsp. *fetus*
8292-1	*Ureaplasma diversum*	Negative	*Coxiella burnetii*
4122-2	*Bovine alphaherpesvirus 1 (BoHV-1)*	Positive	*BoHV-1, Mycoplasma bovis, Coxiella burnetii, Trueperella pyogenes*
4123-2	*Bovine alphaherpesvirus 1 (BoHV-1)*	Positive	*BoHV-1, Ureaplasma diversum, Mycoplasma bovis, Trueperella pyogenes, Pestivirus tauri (BVDv type-2)*

777

778

779 **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

780

781