

Original Article

Selective whole genome amplification and sequencing of *Coxiella burnetii* directly from environmental samples



Jill Hager Cocking^{a,b}, Michael Deberg^b, Jim Schupp^c, Jason Sahl^a, Kristin Wiggins^c, Ariel Porty^d, Heidie M. Hornstra^a, Crystal Hepp^{a,b}, Claire Jardine^e, Tara N. Furstenau^b, Albrecht Schulte-Hostedde^d, Viacheslav Y. Fofanov^{a,b,*}, Talima Pearson^{a,*}

^a Pathogen and Microbiome Institute, Northern Arizona University, Flagstaff, AZ, United States of America

^b School of Informatics, Computing and Cyber Systems, Northern Arizona University, Flagstaff, AZ, United States of America

^c Pathogen and Microbiome Division, TGen North, Flagstaff, AZ, United States of America

^d Department of Biology, Laurentian University, Sudbury, Ontario, Canada

^e Department of Pathobiology, University of Guelph, Guelph, Ontario, Canada

ARTICLE INFO

Keywords:

Whole genome sequencing

Coxiella burnetii

Selective whole genome amplification

ABSTRACT

Whole genome sequencing (WGS) is a widely available, inexpensive means of providing a wealth of information about an organism's diversity and evolution. However, WGS for many pathogenic bacteria remain limited because they are difficult, slow and/or dangerous to culture. To avoid culturing, metagenomic sequencing can be performed directly on samples, but the sequencing effort required to characterize low frequency organisms can be expensive. Recently developed methods for selective whole genome amplification (SWGA) can enrich target DNA to provide efficient sequencing. We amplified *Coxiella burnetii* (a bacterial select agent and human/livestock pathogen) from 3 three environmental samples that were overwhelmed with host DNA. The 68- to 147-fold enrichment of the bacterial sequences provided enough genome coverage for SNP analyses and phylogenetic placement. SWGA is a valuable tool for the study of difficult-to-culture organisms and has the potential to facilitate high-throughput population characterizations as well as targeted epidemiological or forensic investigations.

1. Background

Coxiella burnetii causes Q fever in humans and is typically characterized by ~10 days of self-resolving fever and general malaise, but can also lead to chronic infections, pneumonia, hepatitis, endocarditis, and in some cases, death [1]. *C. burnetii* is globally distributed and shed into the environment via urine, feces, and parturition tissues/fluids of mammals. It is listed as a Category B Select Agent by the CDC due to its high infectivity, ease of aerosolization, and environmental stability. Recent genotyping surveys of cow and goat milk in the USA suggests that it is also ubiquitous in dairy livestock with few barriers to rapid, frequent, and large-scale dissemination [2–4]. Despite these

characteristics, human disease and outbreaks are rare [5], possibly due to host-adapted genotypes circulating in the ruminant populations [2]. However, a recent outbreak of Q fever in the Netherlands with over 2000 human cases was probably due to spill-over from the large number of industrial-scale dairy goat farms [6]. This outbreak resulted in culling > 50,000 goats [7], although it is unknown whether these drastic measures were necessary for the abatement because we know very little about the maintenance and spread of *C. burnetii*. Our previous work has focused on using genomic and sub-genomic genotyping in a phylogenetic context to better understand persistence, transmission, and dissemination at the herd, regional, and global level [2–4,8–13]. To this end, we need a much larger collection of *C. burnetii* whole genomes.

Abbreviations: WGS, Whole genome sequencing; DNA, deoxyribonucleic acid; SNP, single nucleotide polymorphism; CDC, Centers for Disease Control and Prevention; NCBI, National Center for Biotechnology Information; *C. burnetii*, *Coxiella burnetii*; SWGA, selective whole genome amplification.; qPCR, quantitative polymerase chain reaction; dNTP, deoxynucleotides

* Corresponding authors.

E-mail addresses: Jill.Cocking@nau.edu (J.H. Cocking), jjschupp@tgen.org (J. Schupp), Jason.Sahl@nau.edu (J. Sahl), kwiggins@tgen.org (K. Wiggins), APorty@laurentian.ca (A. Porty), Heidie.Hornstra-ONeill@nau.edu (H.M. Hornstra), Crystal.Hepp@nau.edu (C. Hepp), cjardi01@uoguelph.ca (C. Jardine), Tara.Furstenau@nau.edu (T.N. Furstenau), aschultehostedde@laurentian.ca (A. Schulte-Hostedde), Viacheslav.Fofanov@nau.edu (V.Y. Fofanov), Talima.Pearson@nau.edu (T. Pearson).

<https://doi.org/10.1016/j.ygeno.2019.10.022>

Received 29 April 2019; Received in revised form 5 October 2019; Accepted 9 October 2019

Available online 31 October 2019

0888-7543/ © 2019 Published by Elsevier Inc.

Although discovered in the early 1930s, we know very little about the population dynamics and genotype-specific host-pathogen interactions of *C. burnetii*. There are currently 62 genomes of *C. burnetii* publicly available on NCBI (only 11 of which are complete) – not enough to perform high-resolution phylogenetic analyses needed to understand fine-scale patterns of spread, maintenance, and evolution of this pathogen. The lack of *C. burnetii* whole genome sequences is largely due to several major challenges: it is highly infectious, requiring only one to ten cells to produce an infection in humans [14] and therefore requires highly trained individuals and specialized facilities for its study. Secondly, *C. burnetii* is an obligate intracellular pathogen and is thus very difficult and slow to culture. Even in the age of high throughput sequencing, obtaining a complete genome of *C. burnetii* without a culturing step still poses significant problems. *C. burnetii*-positive samples often contain overwhelming amounts of the host's DNA. Thus, metagenomic sequencing, while technically possible, is very resource-inefficient and often not feasible for most research applications. A way to either enrich the *C. burnetii* DNA or reduce host DNA is needed for this and many other species.

Selective whole genome amplification (SWGA) was originally developed and tested on *Borrelia burgdorferi* and the parasite *Wolbachia pipientis* in simple synthetic mixtures containing *Escherichia coli* and *Drosophila melanogaster*, respectively [15]. In simple, natural mixtures, SWGA has also been used to successfully amplify *Plasmodium* spp. and *Mycobacterium tuberculosis* from human blood [16–19]. This method presents an intriguing opportunity to bypass culturing steps and obtain quality whole genome sequence data directly from a clinical or environmental sample. SWGA was specifically designed to amplify a target in a background of another organism(s) – for example the host [15]. This process is dependent on identifying certain sequences, or motifs, that are more common in the target species than other genomes. These motifs serve as primer binding sites and allow the target genome to be preferentially amplified.

Here we apply this method to *C. burnetii* to demonstrate both the utility and challenges of SWGA for population genetics analyses of a bacterial pathogen derived from typical samples that are likely to be dominated by mammalian host DNA as well as a complex community of microbes. For our study, samples of unpasteurized milk and goat anterior vaginal swabs were analyzed for the presence of *C. burnetii*. The SWGA software [15] was used to design primers that would selectively amplify *C. burnetii* in a background of cow or goat. These selective primers were then used in an isothermal multiple displacement amplification. Under some conditions, our results show a 68- to 147-fold enrichment of *C. burnetii* in a goat background, and robust sequence reads, resulting in confident phylogenetic placement of these strains.

2. Methods

2.1. DNA sources and extraction protocols

Known *C. burnetii* positive samples from three goat anterior vaginal swabs and five commercial cow milk samples were used to test the utility of SWGA for *C. burnetii* amplification. The anterior vaginal swabs were obtained as part of another study [20] from goats in Canada that had recently given birth (Laurentian University IACUC approval: certificate number 2014-01-02). Raw cow milk was obtained from small local farms throughout Arizona. Here, sourcing was particularly important as most commercially bought milk contains a mixture of milk from many individual cows, potentially resulting in multiple *C. burnetii* strains and generally diluting a positive signal. DNA from anterior vaginal swabs was extracted in Canada using the DNA purification from buccal swabs (spin protocol) from the QIAamp DNA mini and Blood mini handbook. These DNAs were shipped to Arizona and were not autoclaved upon arrival in the lab. The raw unpasteurized milk samples were brought into the BSL2 lab and immediately autoclaved for 30 min at 121.1 °C to destroy any live bacteria. DNA was extracted from the

milk with the Qiagen DNeasy® Blood & Tissue Kit after pelleting 1 ml of the milk and physical removal of the creamy top layer. The following modifications were made to the kit protocol (July 2006): 1) extractions were incubated overnight with shaking at 56 °C and 2) after the addition of buffer AE but before the final spin, a 5-min incubation at 70 °C was added. All genomic DNA extractions were verified to contain *C. burnetii* DNA using a qPCR detection assay [11,21,22].

2.2. SWGA primer panel design

Selective whole genome amplification (SWGA) [15] uses Φ 29 polymerase for long range multiple displacement amplification. The amplification process is reliant on designing a set of primers that preferentially amplify the *target* genome in the presence of a typically much larger *background* genome (e.g. host), while ideally avoiding priming within the *exclusion* genomic sequences (e.g. circular plasmids that may overtake amplification).

For our study, we used the Dugway 5 J108–111 strain as our target reference genome (GenBank: CP000733.1). This strain has the largest known *Coxiella* genome and has had the fewest deletions compared to other genomes [12]. This genome represents 88% of the *C. burnetii* pan-genome, thus, regions in other genomes are likely to also be found in the Dugway genome. This characteristic may be important as it reduces the likelihood of failing to amplify parts of some genomes due to the presence of unknown genomic regions. The goat (*Capra hircus* scaffolds NW_005100758 - NW_005101023) and cow (*Bos taurus* scaffolds from the Bos_taurus_UMD_3.1.1 assembly) were used as the *background* genomes. Finally, the goat and cow mitochondria sequences were used as *exclusion* genomes to make sure the primers would not amplify any part of the mitochondrial genome as this would dominate the SWGA reaction. This is known to happen if there is low-level nonspecific priming because the small size and circular shape of the mitochondrial genome would result in a “rolling-circle” amplification by the Φ 29 polymerase [15].

These genomes were used with the SWGA pipeline [15] to identify the primer sequences that are present in high frequency (< 7Kbp apart) in the *target* genome, low frequency (> 80K bp apart) in the *background* genomes, and absent in goat and cow mitochondria genomes. SWGA uses multiple displacement amplification and a Φ 29 DNA polymerase. Φ 29 works best at 30 °C so the primers were designed to have a melting temperature at or < 30 °C. Our primers had a melting temperature range of 19.7 °C to 25.6 °C. Our background genome, the goat, is much larger than *E. coli* or *Drosophila melanogaster* genomes used in the original study. Our primer output thus consisted of 40 primers (Supplemental Table 1). We predicted that 37 of the 40 primers contribute additional unique sequence amplification (Supplemental Fig. 1). Primers were ordered from Integrated DNA Technologies and included phosphorothioate bonds between the third to last and second to last as well as between the penultimate and last nucleotides (indicated with a * in Supplemental Table 1). This modification prevents degradation by the Φ 29's exonuclease activity.

2.3. SWGA wet-lab protocol

Forty nanograms in a volume of 2 μ l of each extracted DNA sample was first brought to 35 °C for 5 min. Added to this DNA was 30 units of Φ 29 DNA polymerase (New England Biolabs), reaction buffer and BSA to 1 \times , and 1 mM of dNTPs. 0.25 μ l of each primer from a 100 μ M stock was added to each reaction for a final primer concentration of 0.5 μ M. The total volume for the reaction was 50 μ l. Amplification was carried out in a stepwise fashion starting at 35 °C for 5 min, 34 °C for 10 min, 33 °C for 15 min, 32 °C for 20 min, 31 °C for 30 min, 30 °C for 16 h, and 65 °C for 15 min.

Amplified samples were purified using the QIAquick PCR purification kit using the manufacturer's recommendations. To prepare for whole genome sequencing, both pre and post SWGA samples were

prepped into DNA libraries following standard methods [23]. The sequences were obtained using an Illumina MiSeq instrument. To determine sequencing coverage, resulting reads were aligned to the Dugway 5 J108–111 strain using Bowtie 2 sequence aligner [24].

2.4. Sequence quality assessment and placement on tree

SNPs were discovered by running the NASP pipeline [25] on 62 publicly available *C. burnetii* sequences. Genomes were aligned against *C. burnetii* RSA493 (GCA_000007765) with NUCmer [26] and SNPs were identified with NASP. Any SNP that fell within a duplicated region, based on a NUCmer reference self-alignment, was filtered from downstream analyses. Using WG-FAST [27], the sequencing reads from our samples were aligned to the reference genome (*C. burnetii* RSA493) that was annotated with the SNPs discovered among the 62 public genomes. This allowed us to determine the SNP alleles in our samples and place them onto a phylogenetic tree using WG-FAST [27] and the evolutionary placement algorithm incorporated into RAxML [28]. The samples were also genotyped using a previously published Taqman SNP assay (Cox51bp67) which determines if a sample is in the ST8 clade [29], thus providing an approximate location on the phylogenetic tree.

3. Results

3.1. Selective whole genome amplification

Using the SWGA process, we were able to selectively amplify *C. burnetii* from three goat anterior vaginal swabs (Table 1). 32 of the 40 primers showed an increase in coverage in over 50% of predicted touchdown locations in the *C. burnetii* genome, suggesting lower performance of only 8 primers (Supplemental Fig. 2). For the most successful sample, Cb0628, the results showed a 127-fold increase in *C. burnetii* DNA (Fig. 1). Prior to amplification, 81% of the reads were from the goat host while only 0.6% were from *C. burnetii*. After selective amplification, 74% of reads were aligned to the *C. burnetii* Dugway 5 J108–111 genome and only 7.2% were from the goat (Table 1). With SWGA, approximately 92% coverage of the targeted genome was achieved. Two other cervical samples, Cb0635 and Cb0648, showed lower final quantities compared to sample Cb0628, likely due to the very low starting amounts of *C. burnetii* in the original samples. For sample Cb0635, there was a 147-fold increase in *C. burnetii* DNA after selective amplification. Goat reads went from 54% to 19% of the resulting sequencing reads. Sample Cb0648 had a 68-fold increase in *C. burnetii* reads and the percent of goat reads dropped from 73% to 27%. For Cb0635, there was 20% coverage of the genome after SWGA and 24% for Cb0648 (Table 1). Sequences are available at NCBI under BioProject accession PRJNA498329.

None of the five DNA extractions from raw milk successfully amplified using the SWGA method. We suspected that autoclaving the raw milk compromised the DNA quality. To test this, equivalent amounts of *C. burnetii* DNA were spiked into two separate aliquots of autoclaved raw milk that did not previously contain *C. burnetii* DNA (real-time PCR

Table 1

Sequencing coverage results of *C. burnetii* goat anterior vaginal swab samples. The three sequenced samples include both pre and post SWGA preparations. Percent coverage is based on alignment to the *C. burnetii* MSU Goat genome.

Sample	Total MiSeq reads	Reads aligned to goat genome	Percent goat reads	Reads aligned to <i>C. burnetii</i> genome	Percent <i>C. burnetii</i> reads	Fold bacteria enrichment	Coverage (%) of <i>C. burnetii</i> genome
Cb0628pre	1,347,834	1,095,984	81	7919	0.6	N/A	13
Cb0628post	1,623,630	117,375	7.2	1,196,048	74	127	99
Cb0635pre	1,522,938	8,242,150	54	61	0	N/A	N/A
Cb0635post	884,630	64,367	19	5307	0.6	147	20
Cb0648pre	1,638,174	1,200,185	73	189	0	N/A	N/A
Cb0648post	824,530	226,014	27	6425	0.8	68	24

Reference Dugway 5J108-111 size = 2,158,758 bp.



Fig. 1. Coverage plot of sample Cb0628. The outermost track is the pre-SWGA coverage (mean = 0.69; max = 4; min = 0) followed (inward) by the post-SWGA coverage (mean = 152.93; max = 5696; min = 21) — the average read depth in 1 kb windows is shown with values ranging from 0 (yellow) to 5696 (dark blue). The read depth is based on reads aligned using BWA-MEM v. 0.7.8 with the *Coxiella burnetii* MSU Goat chromosome sequence (NZ_CP018150.1) as a reference. The next track indicates the touchdown locations for all of the primers (forward primers on outer track and reverse primers on inner track); the primer locations were predicted based on the same sequence used as the alignment reference. The remaining tracks indicate the predicted touchdown locations of each individual primer starting with primer #1 in the center moving out to primer #40. The arrows indicate the amplification direction and are extended 10 kb to indicate elongation potential. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

negative). One aliquot was autoclaved again after the addition of *C. burnetii* DNA and the other was not. DNA was extracted from both milk treatments and subjected to SWGA and sequencing. Autoclaving the milk reduced the number of reads (post SWGA) attributed to *C. burnetii* by ~75%.

3.2. Assessment of sequence quality and phylogenetic placement

NASP identified 12,068 SNPs across the 62 publicly available whole genome sequences of *C. burnetii* (Supplemental Table 2) and these SNPs were used to build a maximum likelihood tree (Fig. 2). WG-FAST placed samples Cb0628pre, Cb0628post, Cb0635post, and Cb0648post onto the tree (Fig. 2, in red) with insertion likelihood values of 100%, 100%,

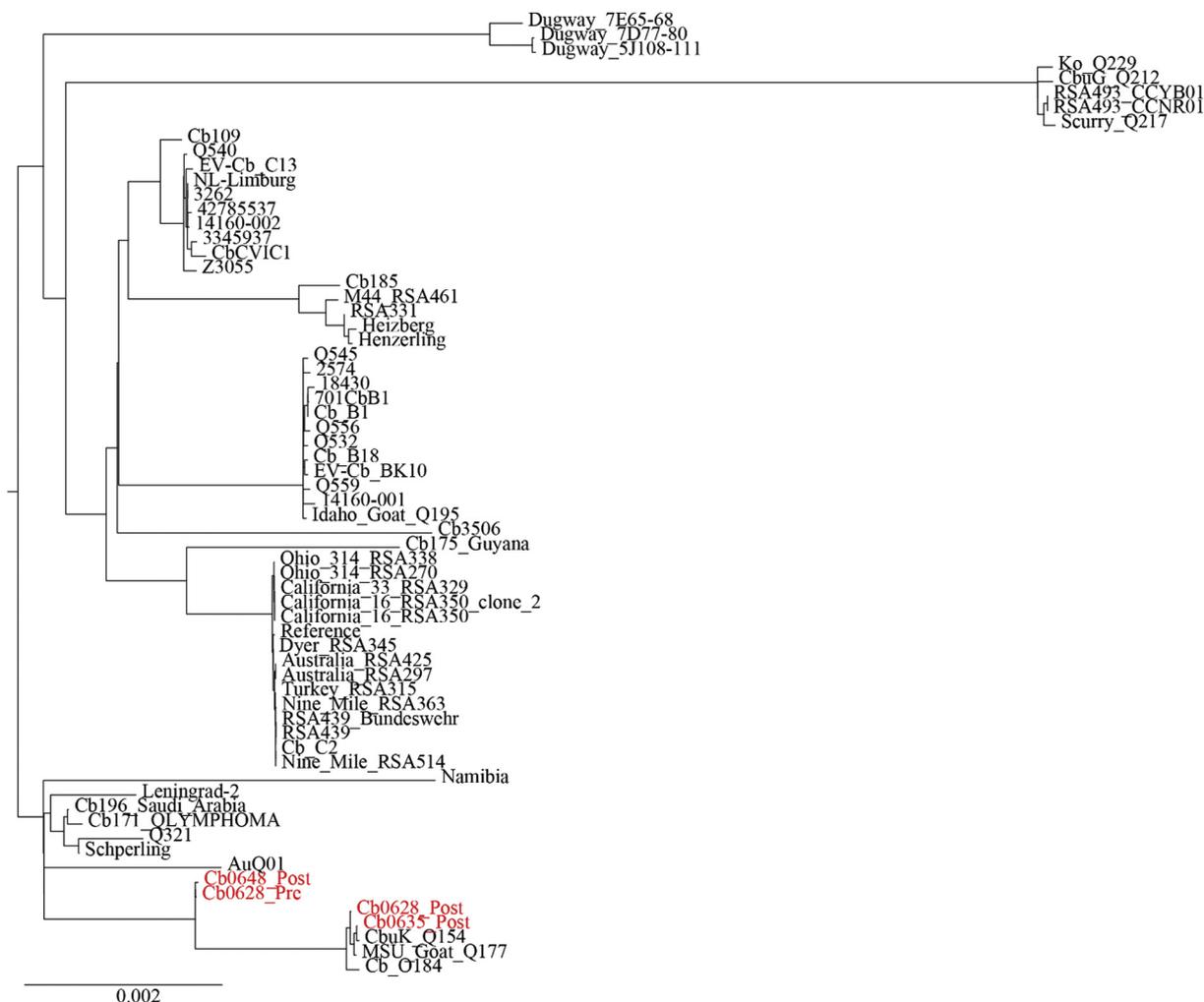


Fig. 2. Phylogeny of *C. burnetii* including SWGA samples sequenced in this study. Phylogenetic tree generated using 12,068 SNPs (consistency index = 0.9814, excluding parsimony uninformative SNPs). Tree is rooted according to Pearson et al. [12]. Samples Cb0628, Cb0635, and Cb0648 are located in the ST8 genetic group, consistent with the Cox51bp67 Taqman assay.

31%, and 99% respectively. The location on the tree of the SWGA samples are consistent with the genotyping assay results that placed Cb0628, Cb0635 and Cb0648 in the ST8 group. The phylogenetic grouping of Cb0648 post-SWGA and Cb0628 pre-SWGA samples away from the other two samples is likely due to the large number of missing SNP calls (see Discussion for additional details).

For Cb0628 (pre-SWGA), *WG-FAST* was able to make calls for 1588 of the 12,068 discovered SNPs and 188 of them had alleles that differed from the reference (RSA 493) genome. After amplification (post-SWGA), there were 12,049 positions called and 1749 were different from the reference. After SWGA, samples Cb0635 and Cb0648 had 1073 and 1438 callable positions with 135 and 216 positions that differed from the reference, respectively. The coverage and proportion filters for the *WG-FAST* analysis were set to default values of 3× and 0.9, respectively.

The SWGA process showed high fidelity with close to all the SNP alleles matching those called in a phylogenetically similar strain, MSU Goat. For sample Cb0628, which had enough sequence data to be analyzed pre and post amplification, almost all of the positions that were called in both genomes had identical alleles (1580 out of 1587). The seven mismatched alleles (positions 142,993, 283,472, 325,792, 743,560, 1,255,635, 1,490,484, 1,793,693) appear to be bad allele calls due to lack of coverage in the pre-amplification sample (average coverage of 2.85× vs 155× for pre- and post- amplification, respectively).

Discovery of novel SNPs can also be important for phylogenetic

analyses but may also indicate sequencing or SWGA errors. We added the SWGA genomes to the previously described NASP run, using coverage and proportion filters of 3× and 0.9, respectively. The low coverage threshold was designed to increase the likelihood of finding false positive SNPs. Only 3 novel SNPs were discovered among SWGA genomes (all in Cb0635). These SNPs were in regions with 6× coverage with all reads supporting each call. While SNP calls that match alleles in other genomes can be assumed to represent true alleles, novel SNPs are more difficult to validate with sequencing data alone. We did not design and run assays to independently confirm the validity of these SNPs. For *C. burnetii*, most SNPs are found between, rather than within clades. Therefore it is not surprising that few SNPs were discovered among the SWGA genomes. The few novel SNPs, the pre/post amplification comparison of SNP alleles, and the adherence to the expected phylogenetic placement of all SWGA samples, provide strong evidence that that errors are not being made during the multiple displacement amplification.

4. Discussion

Whole genome sequencing produces unparalleled data for population, epidemiological, and forensic investigations of bacterial pathogens. However, the need for minimum sequenceable quantities of DNA and difficulties in dealing with pathogen-host mixtures present a massive impediment to these types of studies and has thus prompted

multiple proposed solutions [30,31]. We have successfully used selective whole genome amplification to demonstrate the potential for obtaining whole genome sequences from a microbial pathogen in live-stock samples containing a mixture of DNA from eukaryotic and multiple bacterial sources. The size and complexity of the SWGA primer panels is highly dependent on the size and relative k-mer composition of the foreground and background genomes. While larger than previously reported panels, our 40-primer SWGA panel appeared to successfully and preferentially amplify the foreground genome of *C. burnetii*, with only minimal overlap between primers. The Pearson correlation coefficient between the number of touchdown locations and coverage level was only ~ 0.2 (Supplemental Fig. 3), suggesting that primer saturation has not yet been reached and that additional primers may have further improved amplification performance. Resulting whole genome sequences were of sufficient quality for fine-scale phylogenetic analyses. The SWGA process is relatively simple, inexpensive, and can be run in a comparatively high-throughput manner. Sequencing a single SWGA sample requires only a fraction of a sequencing run. It would be possible to obtain the level of information obtained in this study with sample Cb0628 by pooling 15 SWGA samples on a single MiSeq run.

4.1. SWGA preserves and enables fine-scale phylogenetic analyses

All three successfully amplified goat cervical samples were placed onto a phylogenetic tree (Fig. 2) in clades that were consistent with our genotyping assay. The pre-SWGA sample for Cb0628 had enough sequencing reads to be placed onto the tree as well and provided a reference for the accuracy of sequencing after SWGA. The sequencing of Cb0635 and Cb0648 samples without SWGA did not yield sufficient *C. burnetii* reads and thus could not be placed on the tree. All three samples are located close together on the *C. burnetii* tree and near to other goat-derived samples in the ST8 clade. Most of the phylogenetic separation of the pre-SWGA Cb0628 and Cb0648 samples from the other SWGA samples is an algorithmic artefact of the large number of missing data. The closely related taxa, MSU Goat and Q154, are from Montana and Oregon, respectively. GP-WA1 and HHV-WA2 are both from Washington State. The samples used in this study were collected from the southern portion of Ontario, Canada. The proximity of these samples on the phylogenetic tree suggest little genetic variability and is consistent with a model of rapid and continuous dispersal of genotypes [4]. Host type seems to be correlated with genotype [2]. For example, goats usually harbor the ST8 strains. The availability of more genomes from the SWGA process may reveal additional differences and finer scale patterns of dispersal among these ST8 genotypes.

The SNPs identified from the SWGA samples (Supplemental Table 2) were used for phylogenetic placement and for further assessment of sequencing accuracy and utility. A total of 12,068 SNPs were identified within our dataset of 62 *Coxiella burnetii* genomes. Prior to selective amplification, only 1588 were called in sample Cb0628. After SWGA, 12,049 were called. This is over 7 times more SNP information than would be available without the SWGA process. Importantly, the SNPs called for Cb0628 (pre-SWGA) closely matched the SNPs in Cb0628 post SWGA. This suggests that few, if any, mistakes are being introduced during the selective whole genome amplification. Samples Cb0635 and Cb0648 had fewer called SNPs as they did not amplify to the level of Cb0628, however there was still enough information to place these isolates in the expected position on a phylogenetic tree. As more closely related genomes are sequenced, we will be able to ascertain the accuracy of the SWGA sample-specific SNPs. Importantly, only 35 SNPs from Cb0628 showed an allele that was different from MSU Goat, a closely related genome, and probably represents divergent evolution rather than sequencing or SWGA errors. SNPs are not expected to cause phylogenetic conflict (homoplasy) for clonal organisms such as *C. burnetii* [8,12] and the near total sharing of allele states suggests a high degree of allele calling accuracy and the potential for

fine-scale phylogenetic accuracy.

As the aim of this work is to determine the effectiveness of SWGA for SNP allele calling accuracy, we focused mainly on SNPs that could be verified (discovered in one or more publicly available genomes). We therefore excluded novel SNPs among the SWGA samples for phylogenetic analyses, however, our data suggest that the SWGA process is not introducing SNP errors. As such, novel, autapomorphic, SNPs among SWGA genomes can be included in phylogenetic analyses if they meet certain criteria. In our experience using NASP for SNP discovery, our typical requirement that a SNP allele is found in $> 90\%$ of reads with a minimum of $10\times$ coverage results in few false positives. SWGA sequences with sufficient coverage should therefore be able to be utilized for novel SNP discovery like any other genome.

4.2. Challenges and limitations

The SWGA process holds great potential for fine-scale analysis of a target from environmental samples, however challenges persist and more work is needed to identify and overcome limitations. We were not able to selectively amplify *C. burnetii* DNA from milk, despite identifying and optimizing extraction methods designed to maximize DNA quantity. This is due, at least in part, to exposure to very high temperatures during autoclaving and we suspect that pasteurization may cause similar problems. The SWGA process is reliant on very long fragments of DNA that may be highly fractured upon exposure to high temperatures. Sundararaman et al. (2016) were not able to successfully amplify *Plasmodium* spp. from whole blood samples that had been frozen without preservation, suggesting that exposure to a range of harsh environmental conditions may impact the success of SWGA. The process of multiple displacement amplification requires extension of the amplicon beyond the point at which an adjacent primer is encountered. If the DNA is degraded, and only short pieces are available, the primers will extend but not displace another growing strand. Without displacement, the amplification will not have the “branching” effect that allows large genomes to be amplified in an exponential fashion. In order to successfully work with degraded samples, perhaps designing more primers that are spaced closer together may provide a potential solution, although additional primers increase the cost of these reactions.

It is also possible that the purification kit used for sequence preparation may be losing larger PCR products. The QIAquick PCR purification kit used in our study is recommended for PCR products between 100 and 10,000 bp, however the multiple displacement amplification is known to produce fragments as large as 70,000 bp [18]. Loss of some larger products with this purification method may be contributing to the failure of some samples. The QIAquick kit was not used on samples Cb0635 and Cb0648 which were successfully amplified and sequenced. Further investigations into appropriate purification kits are warranted.

Increasing the DNA concentration for the reaction, performing a second selective amplification with a different set of primers and combining independent amplifications appear to improve results [18]. Further work into understanding why some samples can be amplified successfully while others cannot will provide the foundation for increasing the robustness of this procedure and will allow predictions of which sample types will be most successful.

5. Conclusions

We were able to selectively amplify an important pathogen that is typically found as a minor component in highly complex communities. Our results also enable us to better understand the limitations and possibilities for this method. The promise of obtaining whole genomes directly from samples while bypassing culturing methods opens important avenues of research across taxonomic groups.

Ethics approval and consent to participate

IACUC approval from Laurentian University for collection of goat samples.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analyzed during the current study are available in NCBI under BioProject accession PRJNA498329.

Funding

This work was funded under the State of Arizona Technology and Research Initiative Fund (TRIF), administered by the Arizona Board of Regents, through Northern Arizona University, the National Institute On Minority Health And Health Disparities under the National Institutes of Health award numbers U54MD012388, The Centre of Excellence in Goat Research and Innovation, and by the Canada Research Chair in Applied Evolutionary Ecology. The University of Guelph - OMAFRA Partnership (Emergency Management program) provided funding to support sample collection in Canada. The funders had no role in determining the content of this work.

Authors' contributions

JHC, JMS, VYF, and TP conceived and planned the project. JHC, VYF, and TP wrote the manuscript. JHC performed the SWGA lab work. KW assisted with lab work and sequencing. AP, CJ, and ASH collected and extracted DNA from the goat samples. MD, JS, HMH, CH, TNF, VYF, and TP performed bioinformatic analysis. All authors have read and approved this manuscript.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgements

We greatly appreciate the technical guidance and consultations with Drs. Dustin Brisson, Erik Clarke and Sesh Sundararaman, the initial developers of the SWGA process. The authors thank Dr. Paula Menzies for assisting with project management in Canada.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ygeno.2019.10.022>.

References

- [1] P.E. Fournier, T.J. Marrie, D. Raoult, Diagnosis of Q fever, *J. Clin. Microbiol.* 36 (1998) 1823–1834 <https://www.ncbi.nlm.nih.gov/pubmed/9650920>.
- [2] T. Pearson, H.M. Hornstra, R. Hilsabeck, L.T. Gates, S.M. Olivas, D.M. Birdsell, C.M. Hall, S. German, J.M. Cook, M.L. Seymour, R.A. Priestley, A.V. Kondas, C.L. Clark Friedman, E.P. Price, J.M. Schupp, C.M. Liu, L.B. Price, R.F. Massung, G.J. Kersh, P. Keim, High prevalence and two dominant host-specific genotypes of *Coxiella burnetii* in U.S. milk, *BMC Microbiol.* 14 (2014), <https://doi.org/10.1186/1471-2180-14-41>.
- [3] A.E. Bauer, S. Olivas, M. Cooper, H. Hornstra, P. Keim, T. Pearson, A.J. Johnson, Estimated herd prevalence and sequence types of *Coxiella burnetii* in bulk tank milk samples from commercial dairies in Indiana, *BMC Vet. Res.* 11 (2015), <https://doi.org/10.1186/s12917-015-0517-3>.
- [4] S. Olivas, H. Hornstra, R.A. Priestley, E. Kaufman, C. Hepp, D.L. Sonderegger, K. Handady, R.F. Massung, P. Keim, G.J. Kersh, T. Pearson, Massive dispersal of *Coxiella burnetii* among cattle across the United States, *Microb. Genomics.* 2 (2016), <https://doi.org/10.1099/mgen.0.000068>.
- [5] G.J. Kersh, K.A. Fitzpatrick, J.S. Self, R.A. Priestley, A.J. Kelly, R.R. Lash, N. Marsden-Haug, R.J. Nett, A. Bjork, R.F. Massung, A.D. Anderson, Presence and persistence of *Coxiella burnetii* in the environments of goat farms associated with a Q fever outbreak, *Appl. Env. Microbiol.* 79 (2013) 1697–1703, <https://doi.org/10.1128/AEM.03472-12>.
- [6] W. van der Hoek, F. Dijkstra, B. Schimmer, P.M. Schneeberger, P. Vellema, C. Wijkmans, R. ter Schegget, V. Hackert, Y. van Duynhoven, Q fever in the Netherlands: an update on the epidemiology and control measures, *Euro. Surveill.* 15 (2010), <https://www.ncbi.nlm.nih.gov/pubmed/20350500>.
- [7] B. Schimmer, G. Morroy, F. Dijkstra, P.M. Schneeberger, G. Weers-Pothoff, A. Timen, C. Wijkmans, W. van der Hoek, Large ongoing Q fever outbreak in the south of the Netherlands 2008, *Euro. Surveill.* 13 (2008), http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&opt=Citation&list_uids=18761906.
- [8] T. Pearson, J.D. Busch, J. Ravel, T.D. Read, S.D. Rhoton, J.M. U'Ren, T.S. Simonson, S.M. Kachur, R.R. Leadem, M.L. Cardon, M.N. Van Ert, L.Y. Huynh, C.M. Fraser, P. Keim, Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing, *Proc. Natl. Acad. Sci. U. S. A.* 101 (2004), <https://doi.org/10.1073/pnas.0403844101>.
- [9] H.M. Hornstra, R.A. Priestley, S.M. Georgia, S. Kachur, D.N. Birdsell, R. Hilsabeck, L.T. Gates, J.E. Samuel, R.A. Heinzen, G.J. Kersh, P. Keim, R.F. Massung, T. Pearson, Rapid typing of *Coxiella burnetii*, *PLoS One* 6 (2011), <https://doi.org/10.1371/journal.pone.0026201>.
- [10] G.J. Kersh, R.A. Priestley, H.M. Hornstra, J.S. Self, K.A. Fitzpatrick, B.J. Biggerstaff, P. Keim, T. Pearson, R.F. Massung, Genotyping and axenic growth of *Coxiella burnetii* isolates found in the United States environment, *Vector-Borne Zoonotic Dis.* 16 (2016), <https://doi.org/10.1089/vbz.2016.1972>.
- [11] T. Pearson, J.H. Cocking, H.M. Hornstra, P. Keim, False detection of *Coxiella burnetii*-what is the risk? *FEMS Microbiol. Lett.* 363 (2016), <https://doi.org/10.1093/femsle/fnw088>.
- [12] T. Pearson, H.M. Hornstra, J.W. Sahl, S. Schaack, J.M. Schupp, S.M. Beckstrom-Sternberg, M.W. O'Neill, R.A. Priestley, M.D. Champion, J.S. Beckstrom-Sternberg, G.J. Kersh, J.E. Samuel, R.F. Massung, P. Keim, When outgroups fail; Phylogenomics of rooting the emerging pathogen, *Coxiella burnetii*, *Syst. Biol.* 62 (2013), <https://doi.org/10.1093/sysbio/syt038>.
- [13] K.M. Sulyok, Z. Kreizinger, H.M. Hornstra, T. Pearson, A. Szegedi, T. Dán, E. Balla, P.S. Keim, M. Gyuranecz, Genotyping of *Coxiella burnetii* from domestic ruminants and human in Hungary: indication of various genotypes, *BMC Vet. Res.* 10 (2014), <https://doi.org/10.1186/1746-6148-10-107>.
- [14] A.S. Benenson, W.D. Tigert, Studies on Q fever in man, *Trans. Assoc. Am. Phys.* 69 (1956) 98–104 <https://www.ncbi.nlm.nih.gov/pubmed/13380951>.
- [15] A.R. Leitch, D. Brisson, Selective whole genome amplification for resequencing target microbial species from complex natural samples, *Genetics.* 198 (2014) 473–481, <https://doi.org/10.1534/genetics.114.165498>.
- [16] S.O. Oyola, C.V. Ariani, W.L. Hamilton, M. Kekre, L.N. Amenga-Etego, A. Ghansah, G.G. Rutledge, S. Redmond, M. Manske, D. Jyothi, C.G. Jacob, T.D. Otto, K. Rockett, C.I. Newbold, M. Berriman, D.P. Kwiatkowski, Whole genome sequencing of *Plasmodium falciparum* from dried blood spots using selective whole genome amplification, *Malar. J.* 15 (2016) 597, <https://doi.org/10.1186/s12936-016-1641-7>.
- [17] A.N. Cowell, D.E. Loy, S.A. Sundararaman, H. Valdivia, K. Fisch, A.G. Lescano, G.C. Baldeviano, S. Durand, V. Gerbasi, C.J. Sutherland, D. Nolder, J.M. Vinetz, B.H. Hahn, E.A. Winzler, Selective whole-genome amplification is a robust method that enables scalable whole-genome sequencing of *Plasmodium vivax* from unprocessed clinical samples, *MBio.* 8 (2017), <https://doi.org/10.1128/mBio.02257-16>.
- [18] S.A. Sundararaman, L.J. Plenderleith, W. Liu, D.E. Loy, G.H. Learn, Y. Li, K.S. Shaw, A. Ayoub, M. Peeters, S. Speede, G.M. Shaw, F.D. Bushman, D. Brisson, J.C. Rayner, P.M. Sharp, B.H. Hahn, Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria, *Nat. Commun.* 7 (2016) 11078, <https://doi.org/10.1038/ncomms11078>.
- [19] E.L. Clarke, S.A. Sundararaman, S.N. Seifert, F.D. Bushman, B.H. Hahn, D. Brisson, Swga: a primer design toolkit for selective whole genome amplification, *Bioinformatics.* 33 (2017) 2071–2077, <https://doi.org/10.1093/bioinformatics/btx118>.
- [20] A. Porty, Investigating *Coxiella burnetii* at the livestock - wildlife, *Interface.* (2016).
- [21] A.D. Loftis, W.K. Reeves, D.E. Szumlas, M.M. Abbassy, I.M. Helmy, J.R. Moriarity, G.A. Dasch, Rickettsial agents in Egyptian ticks collected from domestic animals, *Exp. Appl. Acarol.* 40 (2006) 67–81, <https://doi.org/10.1007/s10493-006-9025-2>.
- [22] O. Duron, The IS1111 insertion sequence used for detection of *Coxiella burnetii* is widespread in *Coxiella*-like endosymbionts of ticks, *FEMS Microbiol. Lett.* 362 (2015), <https://doi.org/10.1093/femsle/fnv132>.
- [23] P. Keim, R. Grunow, R. Vipond, G. Grass, A. Hoffmaster, D.N. Birdsell, S.R. Klee, S. Pullan, M. Antwerpen, B.N. Bayer, J. Latham, K. Wiggins, C. Hepp, T. Pearson, T. Brooks, J. Sahl, D.M. Wagner, Whole genome analysis of injectional anthrax identifies two disease clusters spanning more than 13 years, *EBioMedicine.* 2 (2015), <https://doi.org/10.1016/j.ebiom.2015.10.004>.
- [24] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with bowtie 2, *Nat. Methods* 9 (2012) 357–359, <https://doi.org/10.1038/nmeth.1923>.
- [25] J.W. Sahl, D. Lemmer, J. Travis, J.M. Schupp, J.D. Gillette, M. Aziz, E.M. Driebe, K.P. Drees, N.D. Hicks, C.H.D. Williamson, C.M. Hepp, D.E. Smith, C. Roe, D.M. Engelthaler, D.M. Wagner, P. Keim, NASP: an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats, *Microb. Genomics.* 2 (2016), <https://doi.org/10.1099/mgen.0.000074> doi: UNSP 000074.
- [26] A.L. Delcher, A. Phillippy, J. Carlton, S.L. Salzberg, Fast algorithms for large-scale

- genome alignment and comparison, *Nucleic Acids Res.* 30 (2002) 2478–2483, <https://doi.org/10.1093/nar/30.11.2478>.
- [27] J.W. Sahl, J.M. Schupp, D.A. Rasko, R.E. Colman, J.T. Foster, P. Keim, Phylogenetically typing bacterial strains from partial SNP genotypes observed from direct sequencing of clinical specimen metagenomic data, *Genome Med.* 7 (2015) 52, <https://doi.org/10.1186/s13073-015-0176-9>.
- [28] S.A. Berger, D. Krompass, A. Stamatakis, Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood, *Syst. Biol.* 60 (2011) 291–302, <https://doi.org/10.1093/sysbio/syr010>.
- [29] A. Baker, T. Pearson, E.P. Price, J. Dale, P. Keim, H. Hornstra, A. Greenhill, G. Padilla, J. Warner, Molecular phylogeny of *Burkholderia pseudomallei* from a remote region of Papua New Guinea, *PLoS One* 6 (2011) e18343, <https://doi.org/10.1371/journal.pone.0018343>.
- [30] S.N. Gardner, K.G. Frey, C.L. Redden, J.B. Thissen, J.E. Allen, A.F. Allred, M.D. Dyer, V.P. Mokashi, T.R. Slezak, Targeted amplification for enhanced detection of biothreat agents by next-generation sequencing, *BMC Res. Notes.* 8 (2015) 682, <https://doi.org/10.1186/s13104-015-1530-0>.
- [31] G. Carpi, K.S. Walter, S.J. Bent, A.G. Hoen, M. Diuk-Wasser, A. Caccone, Whole genome capture of vector-borne pathogens from mixed DNA samples: a case study of *Borrelia burgdorferi*, *BMC Genomics* 16 (2015) 434, <https://doi.org/10.1186/s12864-015-1634-x>.