Journal of Microbiological Methods 157 (2019) 81–87

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Targeting discriminatory SNPs in *Salmonella enterica* serovar Heidelberg genomes using RNase H2-dependent PCR

Geneviève Labbé[a,1], Marisa A. Rankin[a,1], James Robertson[a], Jonathan Moffat[a], Elissa Giang[a], Lok Kan Lee[a], Kim Ziebell[a], Joanne MacKinnon[a], Chad R. Laing[b], E. Jane Parmley[c], Agnes Agunos[c], Danielle Daignault[d], Sadjia Bekal[e], Linda Chui[f,g], Kimberley A. MacDonald[h,i], Linda Hoang[i], Durda Slavic[j], Danielle Ramsay[k], Frank Pollari[c], John H.E. Nash[l], Roger P. Johnson[a,*]

[a] National Microbiology Laboratory, Public Health Agency of Canada, Guelph, Ontario, Canada
[b] National Centres for Animal Disease Lethbridge Laboratory, Canadian Food Inspection Agency, Lethbridge, AB, Canada
[c] Centre for Foodborne, Environmental and Zoonotic Infectious Diseases, Public Health Agency of Canada, Guelph, Ontario, Canada
[d] National Microbiology Laboratory, Public Health Agency of Canada, St-Hyacinthe, Québec, Canada
[e] Laboratoire de Santé Publique du Québec, Sainte-Anne-de-Bellevue, Québec, Canada
[f] Provincial Laboratory for Public Health-Alberta Public Laboratories, Edmonton, Alberta, Canada
[g] Department of Laboratory Medicine and Pathology, University of Alberta, Edmonton, Alberta, Canada
[h] National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Manitoba, Canada
[i] British Columbia Centre for Disease Control, Public Health Microbiology and Reference Laboratory, Vancouver, British Columbia, Canada
[j] Animal Health Laboratory, Laboratory Services Division, University of Guelph, Guelph, Ontario, Canada
[k] Ministère de l'Agriculture, des Pêcheries, et de l'Alimentation du Québec, Québec, Canada
[l] National Microbiology Laboratory, Public Health Agency of Canada, Toronto, Ontario, Canada

ARTICLE INFO

Keywords:
rhPCR
Salmonella Heidelberg
SNP genotyping
RNase-H dependent PCR

ABSTRACT

We report a novel RNase H2-dependent PCR (rhPCR) genotyping assay for a small number of discriminatory single-nucleotide polymorphisms (SNPs) that identify lineages and sub-lineages of the highly clonal pathogen *Salmonella* Heidelberg (SH). Standard PCR primers targeting numerous SNP locations were initially designed *in silico*, modified to be RNase H2-compatible, and then optimized by laboratory testing. Optimization often required repeated cycling through variations in primer design, assay conditions, reagent concentrations and selection of alternative SNP targets. The final rhPCR assay uses 28 independent rhPCR reactions to target 14 DNA bases that can distinguish 15 possible lineages and sub-lineages of SH. On evaluation, the assay correctly identified the 12 lineages and sub-lineages represented in a panel of 75 diverse SH strains. Non-specific amplicons were observed in 160 (15.2%) of the 1050 reactions, but due to their low intensity did not compromise assay performance. Furthermore, *in silico* analysis of 500 closed genomes from 103 *Salmonella* serovars and laboratory rhPCR testing of five prevalent *Salmonella* serovars including SH indicated the assay can identify *Salmonella* isolates as SH, since only SH isolates generated amplicons from all 14 target SNPs. The genotyping results can be fully correlated with whole genome sequencing (WGS) data *in silico*. This fast and economical assay, which can identify SH isolates and classify them into related or unrelated lineages and sub-lineages, has potential applications in outbreak identification, source attribution and microbial source tracking.

Abbreviations: BHI, brain heart infusion; dNTP, Deoxyribonucleotide triphosphate; LB, Luria-Bertani; PCR, polymerase chain reaction; PFGE, pulsed field gel electrophoresis; rhPCR, RNase-H2 dependent PCR; SH, Salmonella Heidelberg; SNP, single nucleotide polymorphism; Tm, melting temperature; WGS, whole genome sequencing

* Corresponding author at: 110 Stone Road West, Guelph, ON N1G 3W4, Canada.
E-mail address: Roger.Johnson@canada.ca (R.P. Johnson).
[1] These authors contributed equally to the current work.

https://doi.org/10.1016/j.mimet.2018.12.021
Received 8 November 2018; Received in revised form 21 December 2018; Accepted 24 December 2018
Available online 25 December 2018

1. Introduction

In clinical and public health settings, the need for fast, cost-effective and reliable determination of nucleic acid sequences has driven the development of a wide array of polymerase-chain-reaction (PCR) and isothermal amplification methods (Zanoli and Spoto, 2013). In infectious diseases, sequence variations such as single nucleotide polymorphisms (SNPs) or gene presence/absence within pathogens can have important impacts on the course of investigative and therapeutic actions; for example, in identification and tracking of outbreaks, and in the choice of antimicrobial drugs. Although the use of whole genome sequencing (WGS) methods to detect DNA sequence variation is increasing (Khodakov et al., 2016; Nadon et al., 2017), routine WGS adoption is limited mainly to large reference and diagnostic laboratories, due to the substantial capital and operational costs. Consequently, PCR remains the most widely used method for detection of DNA sequences (Niemz et al., 2011), and has the benefits of low material cost, high sensitivity and small capital costs. Numerous modifications to PCR amplification strategies to improve the resolution and reliability of detecting small sequence variations such as SNPs have been described, including primer modifications designed to block the extension of the wild-type sequence (Khodakov et al., 2016). Integrated DNA Technologies (IDT®, Skokie, IL) commercialized a blocking strategy based on RNase H cleavage of an RNA base within the 3′ end of a blocked primer sequence (Dobosy et al., 2011). The thermostable RNase H only recognizes DNA-RNA duplexes and only cleaves the blocked end of the rhPrimer when the complementary target DNA base is present, thereby allowing the primer extension to proceed. Since cleavage of the blocked rhPrimer can only occur when it is bound tightly to the complementary target sequence, the formation of non-specific amplification products and primer-dimers has been reported to be minimized (Dobosy et al., 2011). Assay development of an RNase H2-dependent PCR (rhPCR) from IDT has so far been limited to a few targets (Broccanello et al., 2018; McAllister et al., 2018), so the scalability of the method to more targets is not known.

Here we describe the development and evaluation of an rhPCR multi-target genotyping assay for discrimination of related and unrelated isolates of the highly clonal pathogen, *Salmonella enterica* subsp. *enterica* serovar Heidelberg (SH), one of the top three reported *Salmonella* serovars causing human illness in Canada. Due to the high genetic homogeneity SH isolates, standard subtyping by Pulsed Field Gel Electrophoresis (PFGE) and phage typing often lack the resolution for reliable isolate discrimination (Bekal et al., 2016). By genomic analyses of multiple related and unrelated SH isolates, we recently identified numerous discriminatory SNPs that define phylogenetic SH lineages and sub-lineages (Labbé et al., manuscript in preparation). Our objective was to select and validate a panel of rhPrimers for these SNPs in a minimal rhPCR genotyping assay that could be performed efficiently in any molecular biology laboratory. We found that numerous factors must be considered when designing an rhPCR assay to minimize non-specific amplification. Nevertheless, the final rhPCR assay targeting 14 SNPs effectively discriminated related from unrelated SH isolates. As well, the assay distinguished SH isolates from isolates of other serovars, since all 14 selected SNPs are fully conserved across the SH population, but one or more are absent from other serovars. With a turnaround time of less than five hours after culture, this method provides a quick, effective way to both identify and genotype SH isolates.

2. Materials and methods

2.1. Selection of isolates

A set of 150 SH isolates reflecting the diversity of this serovar in Canada based on PFGE, phage type, sample source and geographic origin was selected for WGS. These isolates were collected from human, poultry, bovine, and equine sources between 1998 and 2014 across nine Canadian provinces, and were provided by several partners, including the Canadian Integrated Program for Antimicrobial Resistance Surveillance, FoodNet Canada, the British Columbia Centre for Disease Control Public Health Laboratory, the Laboratoire de Santé Publique du Québec, the University of Guelph Animal Health Laboratory, and the Provincial Laboratory for Public Health – Alberta Public Laboratories. The sequencing and closing of genomes of 20 of these isolates, including four pairs of epidemiologically related strains, have been reported previously (Labbé et al., 2016a, 2016b; Bekal et al., 2016). The other 130 isolates were sequenced on the Illumina MiSeq platform (Illumina, Inc., San Diego, CA) after DNA and library preparation with the Illumina Nextera XT DNA Library Preparation Kit, as described earlier (Labbé et al., 2016b). As the project progressed, 75 of these SH isolates (Table S1), representing 12 SNP genotypes, were selected for evaluation of the rhPCR assay. We also selected one previously sequenced isolate of four other common *Salmonella* serovars (Enteritidis, Kentucky, Javiana and Typhimurium) for inclusion in our investigation of the presence of the chosen SH SNPs in other serovars. The available GenBank Accession numbers for the WGS data of these and the 75 selected SH isolates are listed in Table S1.

2.2. Targeting discriminatory SNPs with rhPCR

2.2.1. Primer design

Our aim was to determine if rhPCR assays are suitable for parallel testing of multiple canonical SNPs in the SH genome that allow classification of SH isolates into specific phylogenetic groups. At each SNP site, one of two nucleotides is present. Accordingly, two forward and one reverse rhPrimers were designed for each SNP site, with the two forward primers targeting the nucleotide as either “positive” (+) for a given phylogenetic group, or “negative” (−), when not within the target phylogenetic group. After electrophoresis, the presence or absence of amplicons for each rhPrimer set can be used to generate a SNP profile of the strain, which determines its SNP genotype and relatedness to other strains.

Candidate SNPs for the rhPCR assay were identified by assembly and analysis of the WGS data of the selected 150 SH isolates against a closed reference SH genome (GenBank ID CP012921, Labbé et al., 2016a). This reference genome was also used to determine if the 20 base pair sequences flanking the SNP sites were conserved among SH isolates and for *in silico* primer design over the SNP regions. Phylogenetic analysis generated 23 distinct genotypes defined by 65 SNPs, corresponding to main lineages and sub-lineages in the SH population structure (results not shown). Gen2 rhPrimers for each SNP were first designed as standard PCR primers with a query sequence of 1000 base pairs on either side of the target SNP using the PrimerQuest® online tool (IDT Inc., Skokie, IL). The “Force Forward Stop Position (3’ End)” field was set to the SNP site and the acceptable melting temperature (Tm) range was set between 58 and 63 °C, with 60 °C as optimal. Amplicon sizes were restricted to 200–1000 base pairs. Reagents for the rhPCR were used at the default concentrations provided by IDT: 50 nM sodium, 200 nM primers, 3 mM magnesium, and 0.8 mM dNTPs. Candidate SNPs were eliminated if the genome surrounding the target SNP base had a high local GC content or if the primer Tms fell outside 58–63 °C. Further examination with the OligoAnalyzer 3.1 online tool (IDT Inc., Skokie, IL) identified and rejected primers with predicted secondary structures that might interfere with amplification, including those with hairpin loops having predicted Tms > 50 °C, or homodimers or heterodimers with ΔG values < −7.0 kcal/mol. Primers were also rejected if they had high homology to non-target sites in the genome of strain CP012921. The remaining candidate primers were modified from standard PCR primers to RNase H2 compatible primers by appending terminal bases in the following pattern “rDxxDM” where “r” is the RNA base of the target SNP, “D” is a DNA base matched to the target sequence, “xx” is two C3 spacers, and “M” is a DNA base which mismatches the target sequence. These *in silico* designed and modified

rhPrimer sets were subsequently optimized and evaluated by laboratory testing with the rhPCR assays described below.

2.2.2. DNA template preparation

Isolates were streaked onto LB agar from frozen glycerol stocks and incubated overnight at 37 °C. A single colony from each plate was then inoculated into tubes containing 1 mL of Brain Heart Infusion (BHI) broth and incubated overnight at 37 °C. A 900 µL volume of each culture was centrifuged at 5000 x *g* for 5 min and the DNA was extracted from the cell pellets with two kits: the Qiagen DNA Tissue kit with the Qiagen EZ1 robot (Qiagen, Germany), or the EZ10 Spin Column Genomic DNA Extraction Mini Preps Kit (BioBasic, Markham, ON). The two kits differ in hands-on time (30 mins *vs.* 1 h and overall turnaround time (~3 h *vs.* ~1 h, respectively), but the quality of the DNA was equivalent for the rhPCR assay. The extracted DNA concentrations were determined with the Nanodrop 8000 (Thermo Fisher Scientific, Waltham, MA), using a 2 µL sample volume. To estimate the optimal amount of extracted DNA per rhPCR reaction, 10-fold dilutions of the extracted DNA (100–0.00001 ng/reaction) from SH Strain CP012921 were tested with the rhPrimers for SNP location 1,047,714.

2.2.3. rhPCR conditions

All rhPrimers, RNase H2 and rhPCR mastermix (Primetime® Mastermix) were purchased from IDT (Skokie, IL). To ensure robust evaluation, rhPCR and primer testing included the following varying reaction conditions: 5–20 mU/uL RNase H2, 4–1000 nM rhPrimer, 0.1–0.2 ng/uL DNA template, and annealing/extension temperatures of 62–66 °C.

For the final optimized rhPCR assay, reactions were performed in 10 µL volumes that contained 1 ng of extracted DNA (0.1 ng/µL), 5 µL of Primetime Mastermix and varying concentrations of rhPrimers and RNase H2 (Table S2). The rhPCR was performed using TGradient™ thermal cyclers (Biometra, Germany), with the following thermal cycling parameters: 95 °C for 3 min, 40 cycles of 95 °C for 10 s, and annealing/extension at 62 °C or 64 °C for 50 s. The annealing/extension temperatures for each rhPrimer set were optimized for either 62 °C or 64 °C (Table S2).

2.2.4. Gel electrophoresis

The rhPCR amplicons were visualized using double-tiered 1.2% agarose gel cassettes (FlashGel™, Lonza Group, Switzerland), according to the manufacturer's instructions. The DNA Marker was 100 bp - 4 Kb for most gels, or 50 bp - 1.5 Kb. Electrophoresis was at 200 V for approximately 6 min using a Bio-Rad PowerPac™ unit (Bio-Rad, Hercules, CA), and results were visualized and captured using a Gel Doc™ XR + UV imaging device (Bio-Rad).

2.3. Evaluation of the rhPCR assay

The optimized rhPCR assay for 14 SNPs incorporating the above assay conditions was evaluated for its ability to classify SH isolates correctly into SNP genotypes by testing the 75 diverse SH isolates, including four epidemiologically related isolate pairs (Table S3). These 75 isolates are of 12 different genotypes predicted from WGS data assembly to the closed genome of the reference isolate CP012921.

2.4. Presence of the 14 selected SH SNP sites in other Salmonella serovars

Once the final panel of SNPs was identified and the rhPCR was optimized, we applied both *in silico* and laboratory testing to explore if this rhPCR assay is specific for SH, or if it would erroneously yield a genotyping classification for other *Salmonella* serovars, as in the rare event of a sample mix-up or serotyping error. For *in silico* testing, the rhPrimer sequences for the 14 SNP targets were trimmed to exclude the variable SNP site, then the trimmed sequences were used to locate and extract the amplicon sequences of each primer pair for each SNP from a closed SH genome (Strain SL476, Fricke et al., 2011). The primers and full predicted sequences and locations of the amplicons of the 14 SNPs in Strain SL476 (GenBank NC_011083.1) are shown in Table S4. These amplicon sequences were used in a BLAST® search (Camacho et al., 2009) of 500 publicly available closed *Salmonella* genomes that represent 103 different *Salmonella* serovars and include the genomes of 27 SH strains (Robertson et al., 2018). The best hits for the amplicons in each of the genomes were obtained by creating fasta files for each of the primer groups and aligning them using mafft (mafft –auto –reorder –thread 2) (Nakamura et al., 2018). The matched sequences that were missing the primer binding sites were excluded. The genomes of the selected representative isolates of four prevalent *Salmonella* serovars, *S*. Enteritidis, *S*. Typhimurium, *S*. Kentucky, and *S*. Javiana (Table S1) were examined similarly. These same four isolates were also tested by the optimized rhPCR assay to verify the accuracy of the resulting *in silico* predictions, and to further evaluate the rhPCR assay.

3. Results

3.1. rhPrimer design

Initial *in silico* analysis of the WGS data from 150 select SH isolates identified 65 candidate SNP sites that were present in all isolates, had conserved 20 bp flanking sequences, and based on phylogenetic analysis distinguished 23 SNP genotypes. The primers were designed and the final targets selected as described above in the Materials and Methods. Based on these *in silico* analyses, 139 rhPrimers targeting 33 of the 65 SNP sites (Tables S5 and S6) were designed and taken forward for subsequent evaluation and optimization in the rhPCR assay.

3.2. rhPrimer optimization

The 33 SNPs selected by *in silico* analysis were tested in rhPCR assays using a total of 139 candidate rhPrimers (Table S6). One rhPrimer set per genotype was tested initially, starting with the primers that were least likely to form problematic secondary structures. If the rhPrimer sets failed to generate an amplicon under the optimized PCR conditions, the rhPrimer set was re-designed, either by increasing rhPrimer length, or targeting the reverse complementary DNA strand. In the end, 97 of 139 rhPrimers were rejected after this testing (Table S6), usually due to non-specific amplification in isolates that possessed the alternate DNA base at the target SNP site, and less frequently due to weak amplification of the targeted genomic region. Notably, many SNP targets that appeared ideal *in silico* failed to produce specific amplicons in the rhPCR assays. Due to the extensive time and effort required to optimize targeting of all the selected SNP sites, we did not continue designing alternative primers that could identify 8 of the 23 SH sub-lineages that were initially targeted. The final number of targets for the rhPCR assay was reduced to 14 SNP sites that define 15 genotypes (Table S2).

3.3. Optimization of the rhPCR assay

3.3.1. Preparation of the DNA template

Estimation of the optimal concentrations of DNA for the assay (Fig. 1) showed that amplicons were produced using 0.01–100 ng of extracted DNA per reaction. The intensity of the amplicon from extracted DNA decreased below 1 ng of purified DNA and at 0.1 ng, the amplicon was no longer visible in several reactions (data not shown). Also, non-specific amplification in the lower part of the gel was visibly greater above 1 ng of extracted DNA per reaction. Based on these findings, 1 ng of extracted DNA per reaction was used for the subsequent development and optimization of the rhPCR assay.

Extracted DNA (100 to 0.00001 ng/reaction) from Strain CP012921 was tested with rhPrimers for target SNP 4 (at location 1,047,714), where the expected amplicon is 744 bp (Table S2). Strain CP012921 has a “T” DNA base at this genome position, so the amplicons are expected

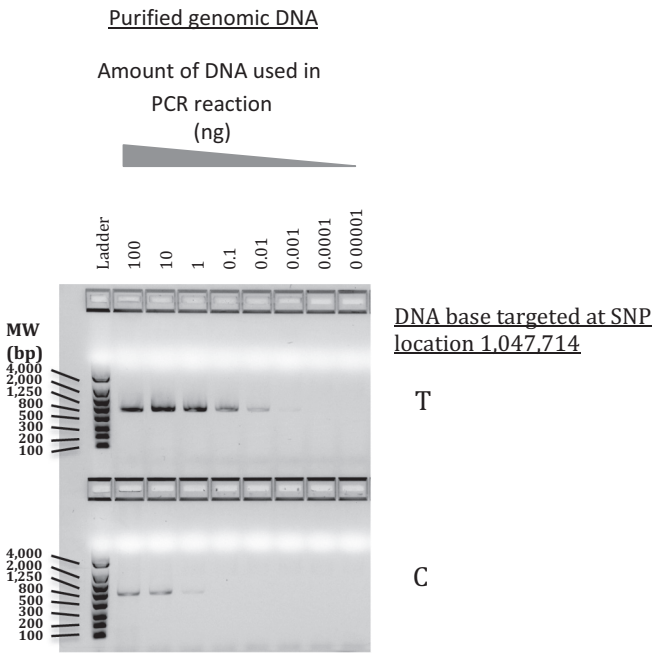

Fig. 1. Results of rhPCR assay with dilutions of extracted DNA to estimate the optimal amount per reaction.

to be present in the upper part of the gel. The bands visible in the lower part of the gel are the result of non-specific amplification.

3.3.2. *rhPCR assay conditions*

The success of an rhPCR assay is heavily influenced by the annealing/extension temperature and by primer and RNase H2 enzyme concentrations. During the primer optimization in the rhPCR assay, it proved very difficult to find a single annealing/extension temperature for specific amplification of all target SNPs. Consequently, the rhPrimers were designed to work at one of two annealing/extension temperatures (62 °C or 64 °C) so that only two thermal cyclers were required for testing all rhPrimer sets simultaneously. The final panel of 14 SNP sites for optimization and evaluation of the rhPCR assay comprises the one SNP site in each of the 15 SNP genotypes (Table S6) that had the highest primer specificity under the selected rhPCR conditions. The final rhPCR assay, described in the Materials and Methods, includes 42 rhPrimers targeting the 14 SNPs (Table S2) in 28 independent PCR reactions that allow classification of the SH isolates into the possible 15 different genotypes corresponding to main lineages and sub-lineages in the SH population (Fig. 2). The DNA base predicted to be present at each of the final 14 SNP target sites for each of the 15 genotypes is shown in Table S7. Note that only 15 profiles are possible for SH isolates, based on the SH population structure analysis: any results showing a profile not included in Table S7 would be deemed invalid, as other SNP combinations are not consistent with phylogeny (Labbé and coll., manuscript in preparation).

3.4. *Evaluation of the rhPCR assay*

Testing the 75 SH isolates by rhPCR comprised 1050 reaction pairs, with amplicons obtained from 1044 reactions on the first run (Table 1). Three of the missing amplicons were from the same isolate, while the other three were from separate isolates. On repeat testing of these six targets in the four isolates, all the expected amplicons were generated, confirming our suspicions that the source of the initial failures was technical.

On visual inspection of the electrophoretic gels, the distribution of bands between the upper and lower portions of the gels was unique for

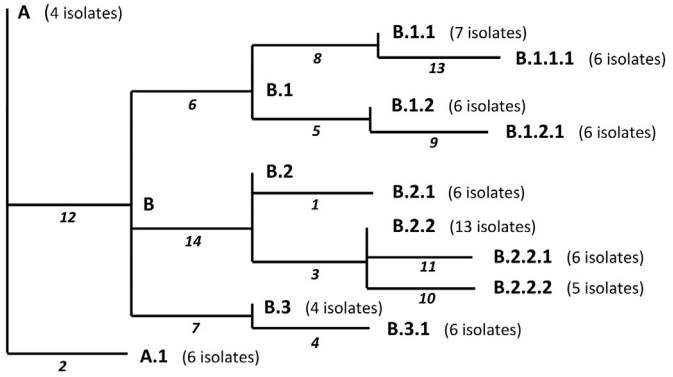

Fig. 2. Phylogenetic tree backbone for the 15 *S.* Heidelberg SNP genotypes (bold font) based on the 14 target SNPs (italic font).
Footnote: See Table S7 for the target SNP numbers and the DNA bases present at each SNP site for each genotype.

each genotype, consistent with the genotype-specific target SNPs in Table S7. Overall, the 75 tested SH isolates, including the four pairs of known epidemiologically related strains (Table S3), were clustered appropriately by the rhPCR assay into 12 of the possible 15 genotypes that were derived by WGS analysis (Table 1, Fig. 2). The three other possible SNP genotypes that can be detected by this assay correspond to ancestral nodes defining main lineages in the SH population structure, and upon analysis of a broader range of publicly available international SH isolates, few were found to be classified into these genotypes (results not shown). The lack of isolates with these three genotypes in the smaller set of 75 Canadian isolates tested in this rhPCR evaluation reflects their rarity among Canadian SH isolates, which are instead frequently classified into sub-lineages stemming from these three ancestral genotypes.

Sample agarose gels displaying the rhPCR genotyping assay results for the SH reference strain and two epidemiologically related isolates are shown in Fig. 3. The epidemiologically related strains CP16581 (Fig. 3B) and CP16510 (Fig. 3C) have identical amplicon patterns, indicating they are of the same genotype, and distinct from the genotype of the reference strain CP012921, which has a different pattern of amplicons (Fig. 3A). Additional gels comparing rhPCR assay results for each of the target SNPs in 16 isolates are shown in Fig. S2.

Over the many rhPCR reactions conducted during this evaluation, non-specific amplification of other regions of the genomes of SH or other *Salmonella* serovars was never observed. However, evidence of non-specific amplification of the alternate SNP bases was observed within the targeted SNP sites. Typically this appeared as faint bands in one half of the gel, but with strong corresponding bands from the expected positive base at the same site in the other half of the gel (Fig. 3B and C). During the testing of the 75 isolates in the evaluation, these non-specific bands were seen in 160 (15.2%) of the 1050 reactions, occurring with 13 of the 14 SNP sites, at within-site frequencies ranging from 0 to 100% for individual primers.

Despite their frequent occurrence, the non-specific amplicon bands were visibly of much lower intensity than those from the corresponding specific amplicons at the same SNP site. By quantification of the band intensities (Image Lab™ Software; BioRad, Berkeley, CA), these non-specific bands had ≤17.5% of the intensity of the corresponding correct SNP base amplicon bands, and a summary of the specificity of the assays based on the relative intensity of the bands is presented in Table 1. Due to the lower intensity of the non-specific amplification bands in the gels, we found that the layout of the electrophoresis gels used in Figs. 3 and S1, allowing direct comparison of the intensities of specific and non-specific bands is critical for the correct interpretation of the rhPCR results.

The agarose gels show amplicons from rhPCR assays of the 14 SNP targets in strain CP012921, Genotype B.2.2.2 (A); and strains

Table 1
Summary of results of evaluation of the rhPCR targeting 14 SNPs for genotyping 75 selected strains of *S.* Heidelberg.

SNP Locations in Reference GenomeCP012921	SNP base on forward strand	Primer ID	No. of amplicons expected[a]	No. of amplicons detected[a]	No. of non-specific amplicons observed[b]	Average relative intensity of non-specific amplicons compared to the targeted amplicons[c]	Figure no. for reference
250,467	T	−	69	69	3	0.16 ± 0.02	Fig. S2a
	A	+	6	6	12	0.12 ± 0.05	
636,931	G	−	69	69	0	N/A	Fig. S2b
	A	+	6	6	0	N/A	
848,800	A	+	24	24	0	N/A	Fig. S2c
	G	−	51	51	10	0.04 ± 0.02	
1,047,714	T	−	**66**	**63**	0	N/A	Fig. S2d
	C	+	9	9	31	0.08 ± 0.03	
1,061,770	G	−	63	63	11	0.05 ± 0.01	Fig. S2e
	T	+	12	12	0	N/A	
2,069,216	G	−	50	50	6	0.11 ± 0.05	Fig. S2f
	T	+	25	25	0	N/A	
2,283,877	G	−	65	65	0	N/A	Fig. S2 g
	T	+	10	10	34	0.03 ± 0.01	
2,432,587	C	−	14	14	0	N/A	Fig. S2 h
	A	+	**61**	**60**	5	0.04 ± 0.03	
2,823,451	T	−	69	69	1	0.09	Fig. S2i
	G	+	6	6	0	N/A	
3,114,352	C	+	5	5	4	0.04 ± 0.02	Fig. S2j
	G	−	70	70	0	N/A	
3,468,405	G	−	**69**	**68**	1	0.06	Fig. S2k
	A	+	6	6	0	N/A	
3,766,668	T	−	65	65	10	0.07 ± 0.04	Fig. S2 l
	G	+	10	10	14	0.01 ± 0.01	
3,827,083	C	−	**69**	**68**	2	0.0417 ± 0.0004	Fig. S2 m
	A	+	6	6	11	0.06 ± 0.02	
4,249,810	T	+	45	45	5	0.08 ± 0.05	Fig. S2n
	G	−	30	30	0	N/A	

[a] Total numbers of expected (1050) and detected (1044) targeted amplicons obtained on initial testing. Data in bold typeface indicate reactions in which the six expected amplicons were not obtained in the first run but were generated on repeating the rhPCR assays once.

[b] Non-specific amplicons appeared as faint bands in the electrophoretic gel due to amplification of sequences containing the non-targeted, alternate base, when the corresponding sequences containing the targeted base generated a strong amplicon band.

[c] As determined by quantification of the targeted and non-specific band intensities using Image Lab™ Software (BioRad, Berkeley, CA). NA = not applicable, since non-specific amplicons were not observed for these reactions. Overall, the intensities of the non-specific amplicons relative to the targeted amplicons ranged from 0.01 to 0.175.

CP016581 (B) and CP016510 (C), which are two epidemiologically related isolates with Genotype B.1.2.1 (see Table S3 for the strain list, and Table S7 for SNP profiles for each genotype). The SNP positions in the reference strain CP012921 are indicated above the lanes of Fig. A. Molecular weight markers are shown on the left. Reactions with positive (+) primers (Table S2) were loaded into the upper tier of the gels, and those with negative primers (−) were loaded into corresponding lanes in the lower tier of the gels, as indicated on the right. The DNA base present at the SNP sites for each strain is shown below the agarose gels (for clarity, the DNA bases shown correspond to the lead (+) strand of reference genome CP012921, regardless of which DNA strand was targeted by the PCR assay). Non-specific amplification can be seen as faint bands in Figs. B and C in the upper half of the gel in the lanes for SNP sites at 1,047,714, 2,283,877 and 3,766,668.

Visual comparison of the gels readily reveals that the amplicon pattern in A differs from those of the epidemiologically related strains in B and C, which are identical. The patterns for all 14 SNP targets together reflect that strain CP012921 has a SNP profile different from the identical profiles of the related strains CP016581 and CP016510, and the results match the genotypes derived from DNA bases found at the SNP locations on the closed genome of these three strains.

3.5. Presence of the 14 selected SH SNP sites in other Salmonella serovars

On investigating if the 14 selected SH SNP sites were specific to SH, *in silico* analysis of 500 closed genomes of 103 *Salmonella* serovars, predicted that only the 27 SH genomes had all 14 rhPCR target sites, whereas one or more target regions were absent in the genomes of the 102 other serovars tested (Tables S8 and S9). Similarly, alignment of the core target sequences of the rhPCR primers with the corresponding genome sequences of the selected isolates of *S.* Enteritidis, *S.* Typhimurium, *S.* Kentucky, and *S.* Javiana (Table S10) predicted that all four isolates lacked at least one of the 14 target sites in our assay. These *in silico* predictions were confirmed in the laboratory, as no amplicon was obtained from these missing SNP sites when these four isolates were tested in the rhPCR (Table S10, Fig. S1). Such concordant results strongly suggest that the assay is specific for the *Salmonella* serovar Heidelberg, as it is the only serovar found to produce amplicons for all 14 target SNP sites.

4. Discussion

This research developed and evaluated an rhPCR method based on multiple canonical SNPs that is able to define genotypes of the highly clonal pathogen, SH, with a turn-around time of less than five hours after culture. The approach of testing multiple target SNPs simultaneously proved challenging in both primer design and optimization of rhPCR assay conditions. When designing rhPrimers *in silico,* the DNA sequences surrounding the limited number of possible SNP targets often proved unsuitable, due to their propensity to form strong secondary structures. As well, primers that appeared very promising *in silico* often did not perform well in the rhPCR assay because of the need for varied PCR conditions and individualized reagent concentrations. Moreover, optimizing numerous reactions for specificity within narrow PCR temperature conditions, so that the samples could be divided between two thermocyclers set at different annealing/extension temperatures, was difficult. For lineages or sub-lineages that are defined by multiple SNP targets, in theory we believe it should be possible to find suitable

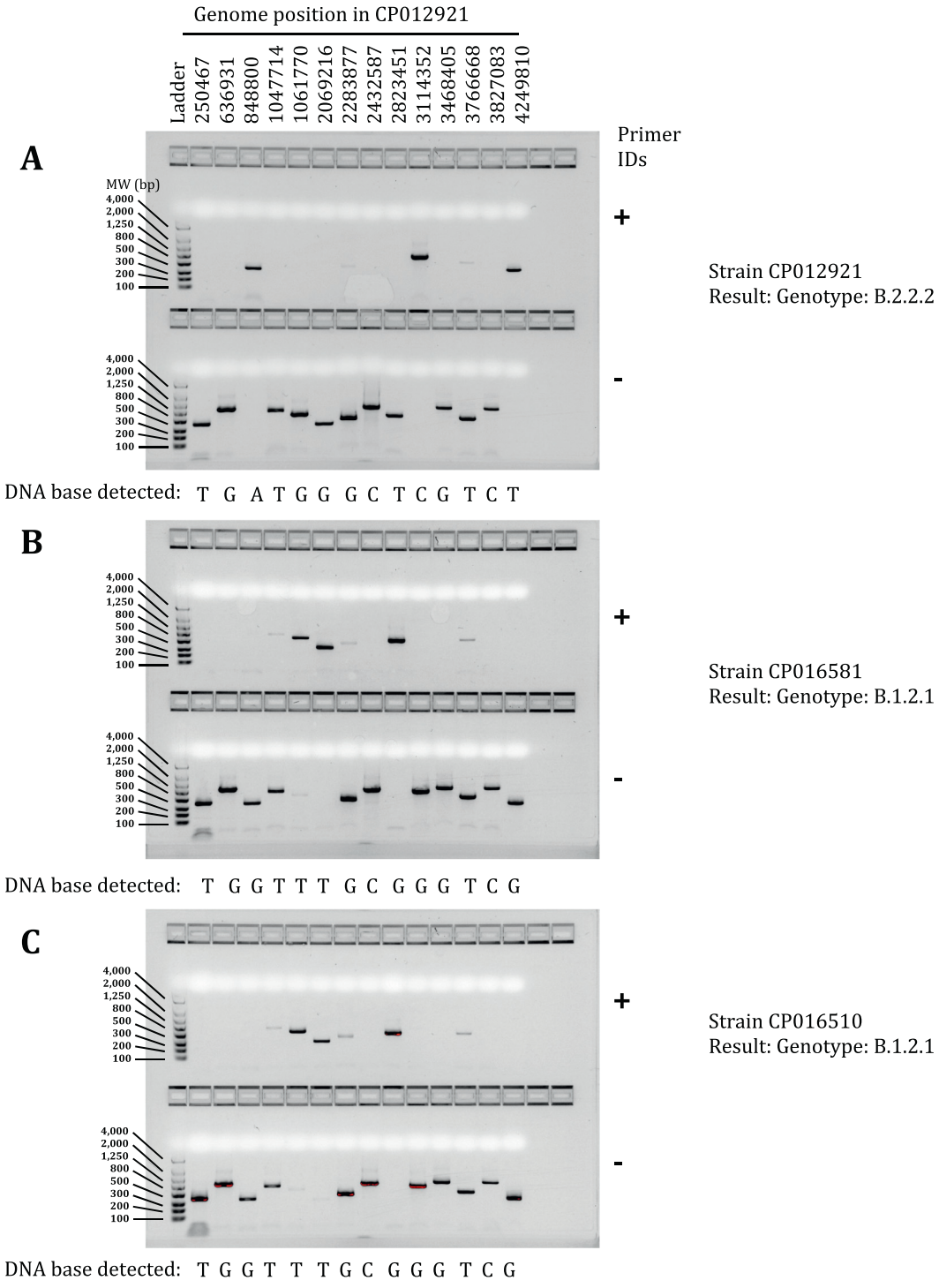

Fig. 3. Example rhPCR assay results for three *S.* Heidelberg strains showing discrimination of related and unrelated strains.

primer designs and assay conditions to specifically detect at least one of those targets. In the present study however, numerous primers and target SNPs were abandoned despite extensive optimization attempts, leading to the application of 42 primers targeting 14 SNPs that together define 15 possible SNP genotypes of SH. After re-testing of the six reactions that failed to generate amplicons in the first run, all 1050 reaction assays yielded results in full agreement with those anticipated from SNP-based genotyping of SH by WGS analysis. The double-tiered layout of the gels enables direct comparison of the targeted and alternate base reactions, allowing the user to verify the presence of all SNP targets in a given sample. This allows the user to quickly establish the SNP profiles of the tested isolates. The 14 targeted SNPs were able to identify an isolate as SH by the presence of all 14 amplicons. None of the other 102 *Salmonella* serovars possessed all 14 target sites.

During the evaluation study, non-specific amplification for the non-targeted base was noted in 160 (15.2%) of the 1050 reactions, occurring with 13 of the 14 SNP sites, at within-site frequencies ranging from 0 to 100% for individual primers. Such non-specific amplification of the target sequence containing the alternate SNP base could not be entirely eliminated by our efforts in primer design and optimization. However, we observed that in some cases a 0.5–2 °C higher annealing and extension temperature reduced the intensity of non-specific amplicons.

These observations are consistent with those of Dobosy et al., 2011 regarding the error rate of RNase H2, who reported that up to 15% of RNase H2 cleavage can occur when there is a mismatch in the target sequence. Given this frequency, it seems that non-specific amplification may compromise assay performance. However, the non-specific amplicons in our study were readily identifiable visually by their low relative intensity.

With respect to reducing the time and cost for the rhPCR genotyping assay procedure, an approach to decrease the turnaround time could be the use of real-time rhPCR in combination with fluorescent probes to detect the presence of amplicons (Broccanello et al., 2018). The use of real-time rhPCR would also decrease the cost per sample, compared to the pre-cast gel imaging detection that was used for its development and evaluation in this study. Further optimization of the rhPrimer design may be necessary to minimize non-specific amplification and the production of primer dimers when using fluorescent probes, as the formation of primer dimers occurred in a few of PCR reactions performed in this study (see bottom of lane 250,467 in Fig. 3B and 4C, and upper portion of gel in Fig. S2a, below the lowest molecular weight markers).

This assay offers increased resolution for subtyping Canadian *S.* Heidelberg isolates compared to the traditional PFGE and phage typing methods, and the rhPCR results can be fully correlated with WGS, as opposed to the results from the traditional methods. This type of assay could be used as a rapid, low cost, viable alternative to WGS to track specific lineages or clonal expansions of high priority clonal pathogens in laboratories that have limited resources, or for time-sensitive applications. For example, in a few hours, preliminary information can be gathered to help decide whether a full outbreak investigation should be initiated. If an outbreak investigation is underway, the results of this assay can quickly determine if novel isolates are related or unrelated to the outbreak cluster. The results of the evaluation demonstrate that this assay correctly resolves each strain into its predicted genotype based on the analysis of the WGS data for the 75 tested isolates. Other possible and perhaps developmentally less challenging applications of rhPCR assays include microbial source tracking, in which the target organisms have a small number of reliable unique SNPs.

Declaration of interest

The authors declare that there are no conflicts of interest.

Funding

This work was supported by the Genomics Research and Development Initiative (GRDI) of the Government of Canada [Grant ID 2256344].

Acknowledgements

We thank Nick Downey and Robert Setter from Integrated DNA Technologies for their technical support during the development of this assay. We also thank Saskatchewan Health and the Canadian Food Inspection Agency for providing isolates for this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mimet.2018.12.021.

References

Bekal, S., Berry, C., Reimer, A.R., Van Domselaar, G., Beaudry, G., Fournier, E., Tremblay, C., 2016. Usefulness of high-quality core genome single-nucleotide variant analysis for subtyping the highly clonal and the most prevalent *Salmonella enterica* serovar heidelberg clone in the context of outbreak investigations. J. Clin. Microbiol. 54, 289–295. https://doi.org/10.1128/JCM.02200-15. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4733192.

Broccanello, C., Chiodi, C., Funk, A., McGrath, J.M., Panella, L., Stevanato, P., 2018. Comparison of three PCR-based assays for SNP genotyping in plants. Plant Methods 14 (28). https://doi.org/10.1186/s13007-018-0295-6.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinforma. 10, 421. https://doi.org/10.1186/1471-2105-10-421.

Dobosy, J.R., Rose, S.D., Beltz, K.R., Rupp, S.M., Powers, K.M., Behlke, M.A., Walder, J.A., 2011. RNase H-dependent PCR (rhPCR): improved specificity and single nucleotide polymorphism detection using blocked cleavable primers. BMC Biotechnol. 11, 1472–6750. https://doi.org/10.1186/1472-6750-11-80.

Fricke, W.F., Mammel, M.K., McDermott, P.F., Tartera, C., White, D.G., Leclerc, J.E., Ravel, J., Cebula, T.A., 2011. Comparative genomics of 28 *Salmonella enterica* isolates: evidence for CRISPR-mediated adaptive sublineage evolution. J. Bacteriol. 193, 3556–3568. https://doi.org/10.1128/JB.00297-11.

Khodakov, D., Wang, C., Zhang, D.Y., 2016. Diagnostics based on nucleic acid sequence variant profiling: PCR, hybridization, and NGS approaches. Adv. Drug Deliv. Rev. 105, 3–19. https://doi.org/10.1016/j.addr.2016.04.005.

Labbé, G., Edirmanasinghe, R., Ziebell, K., Nash, J.H., Bekal, S., Parmley, E.J., Mulvey, M.R., Johnson, R.P., 2016a. Complete Genome and Plasmid Sequences of Three Canadian Isolates of *Salmonella enterica* subsp. *enterica* Serovar Heidelberg from Human and Food Sources. Genome Announc. 14 (4). https://doi.org/10.1128/genomeA.01526-15. pii: e01526–15.

Labbé, G., Ziebell, K., Bekal, S., Macdonald, K.A., Parmley, E.J., Agunos, A., Desruisseau, A., Daignault, D., Slavic, D., Hoang, L., Ramsay, D., Pollari, F., Robertson, J., Nash, J.H., Johnson, R.P., 2016b. Complete genome sequences of 17 Canadian isolates of *Salmonella enterica* subsp. *enterica* Serovar Heidelberg from human, animal, and food sources. Genome Announc. 4 (5). https://doi.org/10.1128/genomeA.00990-16. (pii: e00990–16).

McAllister, C.H., Fortier, C.E., St Onge, K.R., Sacchi, B.M., Nawrot, M.J., Locke, T., Cooke, J.E.K., 2018. A novel application of RNase H2-dependent quantitative PCR for detection and quantification of *Grosmannia clavigera*, a mountain pine beetle fungal symbiont, in environmental samples. Tree Physiol. 38, 485–501. https://doi:10.1093/treephys/tpx147.

Nadon, C., Van Walle, I., Gerner-Smidt, P., Campos, J., Chinen, I., Concepcion-Acevedo, J., Gilpin, B., Smith, A.M., Kam, K.M., Perez, E., Trees, E., Kubota, K., Takkinen, J., Nielsen, E.M., Carleton, H., FWD-NEXT Expert Panel, 2017. PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. Euro Surveill 22https://doi.org/10.2807/1560-7917.ES.2017.22.23.30544. (pii = 30544).

Nakamura, T., Yamada, K.D., Tomii, K., Katoh, K., 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics 34, 2490–2492. https://doi.org/10.1093/bioinformatics/bty121.

Niemz, A., Ferguson, T.M., Boyle, D.S., 2011. Point-of-care nucleic acid testing for infectious diseases. Trends Biotechnol. 29, 240–250. https://doi.org/10.1016/j.tibtech.2011.01.007.

Robertson, J., Yoshida, C., Kruczkiewicz, P., Nadon, C., Nichani, A., Taboada, E.N., Nash, J.H.E., 2018. Comprehensive assessment of the quality of *Salmonella* whole genome sequence data available in public sequence databases using the *Salmonella in silico* Typing Resource (SISTR). Microb Genom. 4. https://doi.org/10.1099/mgen.0.000151.

Zanoli, L.M., Spoto, G., 2013. Isothermal amplification methods for the detection of nucleic acids in microfluidic devices. Bios 3, 18–43. https://doi.org/10.3390/bios3010018.