

Citation: Usongo V, Berry C, Yousfi K, Doualla-Bell F, Labbé G, Johnson R, et al. (2018) Impact of the choice of reference genome on the ability of the core genome SNV methodology to distinguish strains of *Salmonella enterica* serovar Heidelberg. PLoS ONE 13(2): e0192233. https://doi.org/ 10.1371/journal.pone.0192233

Editor: Joël Mossong, Laboratoire National de Santé, LUXEMBOURG

Received: August 24, 2017

Accepted: January 18, 2018

Published: February 5, 2018

Copyright: © 2018 Usongo et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

Funding: The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

RESEARCH ARTICLE

Impact of the choice of reference genome on the ability of the core genome SNV methodology to distinguish strains of *Salmonella enterica* serovar Heidelberg

Valentine Usongo^{1,2}, Chrystal Berry³, Khadidja Yousfi¹, Florence Doualla-Bell¹, Genevieve Labbé⁴, Roger Johnson⁴, Eric Fournier¹, Celine Nadon³, Lawrence Goodridge², Sadjia Bekal^{1,5}*

1 Laboratoire de Santé Publique du Québec, Sainte-Anne-de-Bellevue, Québec, Canada, 2 Department of Food Science and Agricultural Chemistry, Food Safety and Quality Program, McGill University, Ste Anne de Bellevue, Québec, Canada, 3 Division of Enteric Diseases, National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Manitoba, Canada, 4 National Microbiology Laboratory at Guelph, Public Health Agency of Canada, Guelph, Ontario, Canada, 5 Département de Microbiologie, Infectiologie et Immunologie, Université de Montréal, Montréal, Québec, Canada

* sadjia.bekal@inspq.qc.ca

Abstract

Salmonella enterica serovar Heidelberg (*S*. Heidelberg) is one of the top serovars causing human salmonellosis. The core genome single nucleotide variant pipeline (cgSNV) is one of several whole genome based sequence typing methods used for the laboratory investigation of foodborne pathogens. SNV detection using this method requires a reference genome. The purpose of this study was to investigate the impact of the choice of the reference genome on the cgSNV-informed phylogenetic clustering and inferred isolate relationships. We found that using a draft or closed genome of *S*. Heidelberg as reference did not impact the ability of the cgSNV methodology to differentiate among 145 *S*. Heidelberg isolates involved in foodborne outbreaks. We also found that using a distantly related genome such as *S*. Dublin as choice of reference led to a loss in resolution since some sporadic isolates were found to cluster together with outbreak isolates. In addition, the genetic distances between outbreak isolates as well as between outbreak and sporadic isolates were overall reduced when *S*. Dublin was used as the reference genome as opposed to *S*. Heidelberg.

Introduction

Nontyphoidal *Salmonella* (NTS) *enterica* serovars are the most important causes of bacterial gastroenteritis. Among the NTS serovars, Heidelberg is ranked as the second and third most frequent serovar recovered from clinical cases in Québec and Canada respectively [1]. In Québec between 2004 and 2014, 23% of S. Heidelberg clinical isolates were from blood specimens, compared to 7% for *S. enterica* serovar Enteritidis and 5% for *S. enterica* serovar Typhimurium, suggesting an increased capacity of this serovar to cause invasive systemic disease [2].

Pulsefield gel electrophoresis (PFGE) has been the gold standard method used by PulseNet Canada (PNC) since the 1990s for the molecular typing of *Salmonella* during outbreak investigations. However, a major drawback with the use of PFGE in outbreak investigation is the low resolution power of this technique that is further exacerbated when applied to *S*. Heidelberg typing owing to the extremely low genetic diversity of this serovar. For example, 70% of *S*. Heidelberg isolated in Québec belonged to pulsovar 2 [2].

Whole genome sequence (WGS) based methods owing to their growing availability and high genomic resolution are rapidly replacing traditional typing methods such as PFGE within major public health laboratories including PNC [3]. Two popular methods that are increasingly applied in the field of bacterial genomic epidemiology are: the gene-by-gene methods which is basically an extension of the 7 gene MLST typing technique to encompass the entire genome (whole genome MLST, wgMLST) or just the core genome (core genome MLST, cgMLST) [4, 5] and the single nucleotide variant (SNV) methods which identifies single nucleotide variant sby comparing a population of target genomes against a reference [6, 7]. We recently found that the cgSNV method provided superior discriminatory power than traditional methods during outbreak investigations involving *Salmonella* Heidelberg [2].

The choice of the reference genome has been previously proposed as a potential consideration affecting core genome SNV (cgSNV)-based analysis and outcomes. For example, choosing a distantly related strain as a source of reference may tend to cluster isolates that are otherwise genetically distant. Another concern with the choice of reference genome is the sequencing status of the reference genome. It is generally perceived that high-quality complete genomes are preferred to ensure accurate and epidemiologically concordant phylogenetic analysis and outbreak investigation. These concerns were not addressed in our previous work on the cgSNV method [2]. Here using draft *de novo* assembled and completely sequenced or closed genomes of *S*. Heidelberg as well as a distantly related genome such as *S*. Dublin as references, we assessed the ability of the cgSNV methodology to differentiate amongst 145 *S*. Heidelberg strains involved in four distinct outbreaks and sporadic cases of salmonellosis in Québec.

Materials and method

Collection and characterization of bacterial isolates

The 145 S. Heidelberg clinical isolates described in this study were collected as part of the Quebéc surveillance program on human salmonellosis established since 2003 to ensure rapid detection of outbreaks. The food isolates were collected by the Ministère de l'Agriculture, des Pêcheries et de l'Alimentation du Québec (MAPAQ) during routine food-poisoning investigations. Isolates were grown on triple sugar iron agar at 37°C and stored at -80°C in trypticase soy spiked with 10% glycerol. PFGE and serotyping was performed at the Laboratoire de Santé Publique du Québec (LSPQ) following PNC guidelines.

Whole genome sequencing

Frozen bacterial isolates were cultured overnight at 37°C in brain heart infusion broth and genomic DNA was extracted using the Metagenomic DNA isolation Kit for Water (Epicentre, Madison, WI). Samples were prepared using Nextera XT chemistry (Illumina, Inc., San Diego, CA) and were sequenced using Illumina Miseq paired-end read technology using 300 base read lengths. Five strains were selected from the outbreak isolates to serve as references and their reads were *de novo* assembled using SPAdes v. 3.9 [8]. The complete genome counterparts for these strains have been reported in a previous work [9]. Two draft and three

	DNE
--	-----

Strain ID	Genome Status	NCBI Accession No	Source	Serovar	Genome used as reference in Tree number
ID117795	Draft	NA	Human	Heidelberg	1
ID117795	Completely sequenced	CP016507	Human	Heidelberg	2
ID128787	Draft	NA	Human	Heidelberg	3
ID128787	Completely sequenced	CP016586	Human	Heidelberg	4
ID128902	Draft	NA	Human	Heidelberg	5
ID128902	Completely sequenced	CP016579	Human	Heidelberg	6
ID134609	Draft	NA	Human	Heidelberg	7
ID134609	Completely sequenced	CP016581	Human	Heidelberg	8
ID135140	Draft	NA	Food	Heidelberg	9
ID135140	Completely sequenced	CP016510	Food	Heidelberg	10
SL486	Draft	NZ_ABEL00000000	Human	Heidelberg	11
CFSAN024776	Draft	NC_JWQE0000000	Human	Heidelberg	12
SL476	Completely sequenced	NC_011083	Ground turkey	Heidelberg	13
B182	Completely sequenced	CP003416	Bovine feces	Heidelberg	14
SL477	Completely sequenced	CP001144	Human	Dublin	NA

Table 1. General features of the strains used as reference genomes for the cgSNV analysis of 145 S. Heidelberg isolates.

NA, Not available.

https://doi.org/10.1371/journal.pone.0192233.t001

complete, unrelated reference genomes were also downloaded from NCBI and included in this analysis making a total of 15 assessed reference genomes (Table 1).

CgSNV typing

CgSNV analysis was performed using the SNVPhyl pipeline [10] v.1.0 integrated within the NML instance of the Galaxy platform [11]. Briefly, paired-end sequence reads from the 145 isolates were aligned against each of the 15 reference genomes using SMALT v.0.7.5 (http://www.sanger.ac.uk/science/tools/smalt-0). MUMmer v.3.23 [12] and PHAST [13] were used to identify repeat and prophage regions in each reference genome respectively and these regions were excluded from the analysis. Variants were called using two independent variant calling algorithms: FreeBayes v.0.9.20 and SAMtools [14] /BCFtools based on predefined criteria described elsewhere [10]. To infer the relationship between these isolates, minimum spanning trees were constructed from the SNVphyl output data using the geoBURST algorithm built into PHYLOViZ v2.0 [15].

Topological similarity

We assessed the topological similarity of the phylogenetic trees using the Robinson and Foulds (RF) test [16]. This test is a widely used tree metric for tree-to-tree distances and is defined as the minimum number of operations needed to transform one tree into the other. Briefly, new-ick tree files from the SNVphyl pipeline generated using each of the 14 genomes as references were concatenated and the resulting file was submitted to the online phylogenetic tool T-REX [17] to compute the topological distances between the trees.

Nucleotide sequence accession numbers

The sequence data supporting the results of this article have been deposited in the NCBI Sequence Read Archive under accession number SRP098783.

Results

Epidemiological characteristics and PFGE subtyping results of the 145 S. Heidelberg isolates

Epidemiologic and PFGE fingerprinting results of the 145 S. Heidelberg isolates used in this study are presented in Table 2.

Isolates from four distinct outbreaks that occurred in Quebéc between 2012 and 2016 were also included in the analysis. These outbreaks were designated as follows: outbreak 1, 2012 (n = 10; 8 human and 2 food isolates) outbreak 2, 2013 (n = 8 human isolates) outbreak 3, 2014 (n = 28; 12 human and 16 food isolates and outbreak 4, 2016 (n = 8 human isolates). All human cases and food items linked to these outbreaks were confirmed by epi-data. Outbreak 1 was linked to a wedding party, outbreak 2 and 3 were traced to separate restaurants and outbreak 4 was associated with a daycare catering service. In addition to the outbreak isolates (n = 54) we also added 91 sporadic clinical isolates collected in Quebéc between 2007 and 2016 into the analysis.

Whole genome sequencing results

An average of 983,919 reads was obtained per isolate (range, 339,270–2,974,917) for the set of 145 S. Heidelberg isolates, corresponding to an estimated average genome coverage of 121x (range, 42x -365x). The number of SPAdes-assembled contigs for the five outbreak isolates that were selected to act as the reference ranged from 24–27 with all the isolates assembled into fewer than 27 contigs (Table 3). The completely sequenced genome equivalent of these isolates have been published in a previous study [9].

Core genome single nucleotide analysis

After removing repeats and prophages as well as SNV-dense regions from all the reference genomes, an average of 4,008,254 genomic positions (range 3,681,444–4,049,343) representing an average of 86% of the reference genomes (range 84.63–86.72%) had sufficient coverage (\geq 15x) across all 145 isolates for reference mapping. For all the *S*. Heidelberg reference genomes, an average of 769 high-quality consensus SNVs (range 751–819) were identified by both variant callers as common to all isolates and used for subsequent phylogenetic clustering. For the distantly related *S*. Dublin genome, 18,155 high-quality core genome SNVs were used to construct the phylogeny. In total, 15 minimum spanning trees were generated with 14 of these trees representing the 14 *S*. Heidelberg reference genomes. All outbreak isolates formed distinct clusters with all the 14 *S*. Heidelberg reference genomes and the topologies of these trees were highly similar (Fig 1A and 1B).

Using S. Dublin as the reference genome led to a loss in resolving power. In fact, sporadic isolates clustered with outbreak 1 isolates (Fig 2).

The phylogenetic features revealed by the minimum spanning trees were nearly identical across all the 14 S. Heidelberg reference genomes (Table 4).

The genetic distances observed between outbreak isolates as well as between outbreak and sporadic isolates were nearly identical across the 14 *S*. Heidelberg reference genomes. Using *S*. Dublin as the reference genome led to a significant reduction in genetic distances between sporadic and outbreak isolates with some sporadic isolates having 0 and 3 SNV difference with outbreak 1 and 3 respectively whereas the genetic distances between isolates in these outbreaks ranged from 0–3 SNVs (Table 5).

Isolate No.	Source	Isolation date	Outbreak code	Pulsotype	Phage type	NCBI accession no.
ID117793	Human	05-2012	1	2	19	SH12-001
ID117794	Human	05-2012	1	2	19	SH12-002
ID117795	Human	05-2012	1	2	19	SH12-003
ID117796	Human	05-2012	1	2	19	SH12-004
ID117797	Human	05-2012	1	2	19	SH12-005
ID117798	Human	05-2012	1	2	19	SH12-006
ID117799	Human	05-2012	1	2	19	SH12-007
ID117800	Human	05-2012	1	2	19	SH12-008
ID118040	Food	05-2012	1	2	19	SH12-009
ID117870	Food	05-2012	1	2	19	SH12-010
ID128696	Human	11-2013	2	2	26	SH13-001
ID128783	Human	11-2013	2	2	26	SH13-002
ID128786	Human	11-2013	2	2	26	SH13-003
ID128787	Human	11-2013	2	2	26	SH13-004
ID128808	Human	11-2013	2	2	26	SH13-005
ID128902	Human	11-2013	2	2	26	SH13-006
ID128908	Human	11-2013	2	2	26	SH13-007
ID128910	Human	11-2013	2	2	26	SH13-008
ID134557	Human	07-2014	3	2	19	SH14-001
ID134930	Human	08-2014	3	2	19	SH14-002
ID134612	Human	08-2014	3	2	19	SH14-003
ID134421	Human	08-2014	3	2	19	SH14-004
ID134719	Human	08-2014	3	2	19	SH14-005
ID134608	Human	08-2014	3	2	19	SH14-006
ID135122	Human	08-2014	3	2	19	SH14-007
ID134610	Human	08-2014	3	2	19	SH14-008
ID134609	Human	08-2014	3	2	19	SH14-009
ID134565	Human	08-2014	3	2	17	SH14-010
ID134559	Human	08-2014	3	2	17	SH14-011
ID134929	Human	08-2014	3	2	ATHE-35	SH14-012
ID134879	Food	08-2014	3	2	19	SH14-013
ID134880	Food	08-2014	3	2	19	SH14-014
ID134881	Food	08-2014	3	2	19	SH14-015
ID134882	Food	08-2014	3	2	19	SH14-016
ID134883	Food	08-2014	3	2	19	SH14-017
ID134884	Food	08-2014	3	2	19	SH14-018
ID134885	Food	08-2014	3	2	19	SH14-019
ID134886	Food	08-2014	3	2	19	SH14-020
ID134887	Food	08-2014	3	2	19	SH14-021
ID134888	Food	08-2014	3	2	19	SH14-022
ID134889	Food	08-2014	3	2	19	SH14-023
ID134890	Food	08-2014	3	2	19	SH14-024
ID135137	Food	08-2014	3	2	19	SH14-025
ID135138	Food	08-2014	3	2	19	SH14-026
ID135139	Food	08-2014	3	2	19	SH14-027
ID135140	Food	08-2014	3	2	19	SH14-028
ID148030	Human	03-2016	4	2	19	SRR5228105

Table 2. Epidemiologic and subtyping results of the 145 S. Heidelberg clinical and food isolates used in this study.

(Continued)

Table 2. (Continued)

Isolate No.	Source	Isolation date	Outbreak code	Pulsotype	Phage type	NCBI accession no.
ID148149	Human	03–2016	4	2	19	SRR5228097
ID148230	Human	03–2016	4	2	19	SRR5228082
ID148231	Human	03–2016	4	2	19	SRR5228079
ID148280	Human	03–2016	4	2	19	SRR5228104
ID148286	Human	03–2016	4	2	19	SRR5228087
ID148337	Human	03–2016	4	2	19	SRR5228078
ID148338	Human	03–2016	4	2	19	SRR5228091
ID094525	Human	12-2007	NA	2	19	SRR5227118
ID095996	Human	04-2008	NA	3	11	SRR5227171
ID097320	Human	07-2008	NA	2	19	SRR5227121
ID099254	Human	10-2008	NA	2	29	SRR5227124
ID099787	Human	12-2008	NA	2	19	SRR5228101
ID100344	Human	01-2009	NA	2	19	SRR5227119
ID100753	Human	02-2009	NA	2	19	SRR5227155
ID101488	Human	04-2009	NA	122	16	SRR5227148
ID102666	Human	07–2009	NA	2	26	SRR5227120
ID102743	Human	08-2009	NA	2	19	SRR5227166
ID102860	Human	08-2009	NA	17	19	SRR5227126
ID102963	Human	08-2009	NA	2	26	SRR5228093
ID103472	Human	09-2009	NA	2	19	SRR5227163
ID103849	Human	10-2009	NA	1	2	SRR5227117
ID103978	Human	10-2009	NA	138	16	SRR5227128
ID104279	Human	11-2009	NA	140	1	SRR5227146
ID104398	Human	12-2009	NA	1	2	SRR5227169
ID105089	Human	02-2010	NA	6	32	SRR5227122
ID105144	Human	02-2010	NA	2	19	SRR5227127
ID106827	Human	06-2010	NA	87	32	SH12-013
ID107176	Human	07-2010	NA	2	29	SRR5228100
ID107454	Human	07-2010	NA	1	2	SRR5227152
ID108191	Human	08-2010	NA	2	19	SH10-001
ID108221	Human	08-2010	NA	2	26	SH10-014
ID108759	Human	09-2010	NA	86	26	SRR5227162
ID108677	Human	09-2010	NA	2	26	SH10-015
ID110275	Human	01-2011	NA	165	35	SRR5227139
ID110331	Human	01-2011	NA	107	22	SRR5227156
ID110403	Human	01-2011	NA	2	19	SRR5227141
ID110674	Human	02-2011	NA	168	atypical	SRR5227174
ID110801	Human	02-2011	NA	2	26	SH11-002
ID111466	Human	04-2011	NA	2	19	SRR5227167
ID113160	Human	08-2011	NA	- 66	29	SRR5227130
ID113273	Human	08-2011	NA	175	47	SRR5227140
ID113787	Human	09-2011	NA	178	atypical	SRR5227173
ID114520	Human	10-2011	NA	2	29	SRR5227135
ID114593	Human	11-2011	NA	2	29	SRR5227157
ID115377	Human	12-2011	NA	2	19	SRR5227165
ID115568	Human	01-2012	NA	- 86	29	SRR5227133
ID116136	Human	02_2012	NA	2	29	SRR5227164
10130	iiuiiiaii	02 -2012	1111	-	27 27	0111022/101

(Continued)

Table 2. (Continued)

DinestHuma090.2NA8792.0SH0-01DinesoHuma090.2NA21SR52:17-1Din1790Huma090.2NA292.0SR52:17-1Din1780Huma0-20.2NA292.0SR52:17-1Din1780Huma0-20.2NA291.0SR12:07-1Din1780Huma0-20.2NA291.0SR52:17-1Din1780Huma0-20.2NA2SR52:17-1SR52:17-1Din1820Huma0-20.2NA20SR52:17-1Din1820Huma0-20.2NA20SR52:17-1Din1820Huma0-20.2NA20SR52:17-1Din1820Huma0-20.2NA20SR52:17-1Din1820Huma0-20.2NA20SR52:17-1Din1820Huma0-20.2NA20SR52:17-1Din1820Huma0-20.2NA20SR52:17-1Din1820Huma0-20.2NA20SR52:17-1Din1820Huma0-20.2NA20SR52:17-1Din1820Huma0-20.2NA20SR52:17-1Din1820Huma0-20.2NA20SR52:17-1Din1820Huma0-20.2NA20SR52:17-1Din1820Huma0-20.2NA2NASR5	Isolate No.	Source	Isolation date	Outbreak code	Pulsotype	Phage type	NCBI accession no.
D116210Human0-1012NA107NTRE-10SR522719D117200Human0-2012NA29SR522714D117804Human0-2012NA219SR522714D117870Human0-2012NA219SR522714D117870Human0-2012NA29SR522714D117870Human0-2012NA29SR522713D118206Human0-2012NA29SR522713D118207Human0-2012NA2101SR522715D118207Human0-2012NA210SR522715D118208Human0-2012NA210SR522715D118309Human0-2012NA210SR522715D118301Human0-2012NA210SR522715D118304Human0-2012NA210SR522715D119304Human0-2012NA210SR522714D119304Human0-2012NA210SR522714D119304Human0-2012NA210SR522714D119304Human0-2012NA210SR522714D119304Human0-2012NA210SR522714D119304Human0-2012NA210SR522714D119304Human0-2012NA210	ID116271	Human	02-2012	NA	87	32	SH10-014
D1171101Human04-2012NA29SR822713D117980Human04-2012NA219SH10-002D11784Human04-2012NA219SR822714D11787Human05-2012NA219SR822714D117820Human05-2012NA45SR822715D118230Human05-2012NA410SR822715D118240Human05-2012NA10SR822715D118240Human05-2012NA10SR822715D118250Human06-2012NA210SR822715D11879Human06-2012NA210SR822715D11879Human06-2012NA210SR822715D11879Human07-2012NA210SR822715D11879Human07-2012NA210SR822714D11894Human07-2012NA210SR822714D11894Human07-2012NA210SR822714D11894Human06-2012NA210SR822714D11894Human08-2012NA210SR822714D11894Human08-2012NA210SR822714D11894Human08-2012NA210SR822714D11894Human08-2012NA210SR822714 </td <td>ID116824</td> <td>Human</td> <td>03-2012</td> <td>NA</td> <td>107</td> <td>ATHE-10</td> <td>SRR5227129</td>	ID116824	Human	03-2012	NA	107	ATHE-10	SRR5227129
D117960Human04-2012NA29SR822712D117568Human04-2012NA219SR822713D117787Human05-2012NA219SR12-011D118200Human05-2012NA45SR822713D118204Human05-2012NA45SR822713D118304Human05-2012NA107ATFE-10SR822713D118305Human05-2012NA5210SR822713D118304Human06-2012NA5210SR822713D118305Human06-2012NA5210SR822713D118704Human06-2012NA210SR822713D118704Human07-2012NA210SR822713D118704Human07-2012NA210SR822714D11924Human07-2012NA210SR822714D11924Human06-2012NA210SR822714D119304Human06-2012NA210SR822714D119305Human06-2012NA210SR822714D119304Human06-2012NA210SR822714D119304Human06-2012NA210SR822714D119305Human06-2012NA210SR822716D119304Human06-2012NA<	ID117211	Human	04–2012	NA	2	29	SRR5227151
Di117950Human0-2012NA219SH0202D117958Human0-2012NA219SH3227143D1181250Human0-2012NA45SR822713D118230Human0-2012NA82SR822713D118230Human0-2012NA5202SR822713D118531Human0-2012NA5210SR822713D118530Human0-2012NA5210SR822715D118790Human0-2012NA218SR822715D118791Human0-2012NA29SR822715D118794Human0-2012NA29SR822715D118794Human0-2012NA1610SR822714D119874Human0-2012NA219SR822714D119874Human0-2012NA219SR822714D119874Human0-2012NA219SR822716D119874Human0-2012NA219SR822716D119874Human0-2012NA219SR822716D120876Human0-2012NA219SR822716D120876Human0-2012NA219SR822716D120876Human0-2012NA219SR822716D120876Human0-2013NA219 <td>ID117095</td> <td>Human</td> <td>04-2012</td> <td>NA</td> <td>2</td> <td>29</td> <td>SRR5227172</td>	ID117095	Human	04-2012	NA	2	29	SRR5227172
ID11680Human0-2012NA29SRR.22712ID11800Human0-2012NA45SRR.22713ID11820Human0-2012NA8620SRR.22713ID11820Human0-2012NA210NILLSRR.22713ID11850Human0-2012NA220100SRR.22713ID11850Human0-2012NA210101SRR.22713ID11870Human0-2012NA210101SRR.22713ID11870Human0-2012NA219SRR.22713ID11870Human0-2012NA219SRR.22713ID11870Human0-2012NA2101SRR.22714ID11944Human0-2012NA2101SRR.22714ID11953Human0-2012NA2101SRR.22714ID11954Human0-2012NA2101SRR.22714ID11954Human0-2012NA210SRR.22716ID11954Human0-2012NA210SRR.22714ID12056Human0-2012NA210SRR.22716ID12056Human0-2012NA210SRR.22713ID12056Human0-2012NA210SRR.22716ID12056Human0-2014NA210SRR.22716ID12056Human0-2014<	ID117506	Human	04-2012	NA	2	19	SH10-002
ID117520Human0-2012NA29NIL201ID11820Human0-2012NA862SR852713ID11820Human0-2012NA107NIL101SR852715ID11830Human0-2012NA210SR852715ID11851Human0-2012NA210SR852715ID11857Human0-2012NA210SR852715ID11879Human0-2012NA218SR12.012ID11879Human0-2012NA218SR852715ID11879Human0-2012NA210SR852715ID11942Human0-2012NA210SR852714ID11945Human0-2012NA210SR852714ID11959Human0-2012NA210SR852714ID11964Human0-2012NA210SR852714ID11964Human0-2012NA210SR852714ID11964Human0-2012NA210SR852714ID12054Human0-2012NA219SR852713ID12054Human0-2012NA219SR852713ID12054Human0-2012NA219SR852714ID12054Human0-2012NA219SR852714ID12054Human0-2014NA219<	ID117683	Human	04–2012	NA	2	19	SRR5227143
ID118290Human6>0202NA4.05.000SRR5227154ID118200Human05-2012NA107ATHE-10SRR5227156ID11851Human06-2012NA5210.0SRR5227157ID11852Human06-2012NA5210.0SRR5227157ID118570Human06-2012NA5210.0SRR5227157ID118797Human06-2012NA210.0SRR5227157ID118797Human07-2012NA29SRR5227157ID119244Human07-2012NA210.0SRR5227145ID119454Human08-2012NA217SRR5227145ID11959Human08-2012NA210.0SRR5227147ID11959Human08-2012NA210.0SRR5227145ID11959Human08-2012NA210.0SRR5227160ID11959Human08-2012NA210.0SRR5227160ID11959Human09-2012NA210.0SRR5227160ID12950Human09-2012NA210.0SRR5227160ID12950Human11-2012NA210.0SRR5227160ID12950Human09-2012NA210.0SRR5227160ID12950Human09-2012NA210.0SRR5227160ID12950Human09-2013NA210.0 <td>ID117578</td> <td>Human</td> <td>05–2012</td> <td>NA</td> <td>2</td> <td>19</td> <td>SH12-011</td>	ID117578	Human	05–2012	NA	2	19	SH12-011
ID118230Human6-2012NA8629MS222154ID118551Human0-2012NA5210SR8227136ID11852Human0-2012NA5210SR822715ID11870Human0-2012NA216SR822715ID11879Human0-2012NA216SR822715ID11879Human0-2012NA216SR822715ID11972Human0-2012NA29SR822713ID11984Human0-2012NA216SR822713ID11994Human0-2012NA29SR822713ID11994Human0-2012NA216SR822714ID11996Human0-2012NA216SR822714ID11986Human0-2012NA216SR822714ID12040Human0-2012NA216SR822714ID12047Human0-2012NA216SR822714ID12047Human0-2012NA216SR822714ID12048Human0-2012NA216SR822714ID12049Human0-2013NA216SR822714ID12049Human0-2014NA216SR822714ID12049Human0-2014NA216SR822714ID12049Human0-2014NA216<	ID118209	Human	05–2012	NA	4	5	SRR5227123
ID118320Human0PNA107NTHE-10SR822713ID11852Human0Col21NAS210SR822713ID11870Human0NAS210SR822713ID11870Human0Col21NAS210SR822713ID11870Human07012NA210SR822713ID11924Human07012NA210SR822713ID11924Human07012NA210SR822713ID11944Human080201NA210SR822714ID11953Human080212NA210SR822714ID11964Human080212NA210SR822714ID11958Human080212NA210SR822714ID1964Human0NA210SR822714ID1965Human0NA210SR822714ID1965Human0NA210SR822713ID12056Human0NA210SR822713ID12057Human0NA210SR822714ID12058Human0NA210SR822713ID12059Human0NA210SR822713ID12059Human0NA210SR822713ID12050Human<	ID118236	Human	05-2012	NA	86	29	SRR5227154
ID118521Human0012NA5210SR522715ID118700Human0012NA5210SR522715ID18707Human0012NA218SR522715ID118240Human0012NA29SR522715ID119242Human0-7012NA29SR522715ID119254Human0-7012NA1861010SR522713ID119254Human0-7012NA1861010SR522714ID119360Human0-8012NA219SR522714ID11954Human0-8012NA219SR522714ID11957Human0-8012NA219SR522714ID11957Human0-8012NA219SR522716ID11967Human0-9012NA219SR522714ID12058Human0-9012NA219SR522713ID12059Human0-9012NA219SR522714ID12059Human0-9012NA219SR522714ID12059Human0-9012NA219SR522714ID12050Human0-9013NA219SR522714ID12050Human0-9014NA219SR522714ID12050Human0-9014NA219SR522714ID12051Human0-9014NA2<	ID118280	Human	05-2012	NA	107	ATHE-10	SRR5227150
In1832Human6-2012NA5210RR522715D118790Human6-2012NA5210SR522718D118790Human0-7012NA219SR522718D119240Human0-7012NA29SR522718D119360Human0-7012NA29SR522718D119360Human0-7012NA217SR522718D119360Human6-2012NA217SR522718D119470Human6-2012NA219SR522714D119470Human6-2012NA219SR522714D119480Human6-2012NA219SR522716D119490Human6-2012NA219SR522716D119490Human6-2012NA219SR522716D120490Human6-2012NA219SR522716D120490Human6-2012NA219SR522716D120490Human6-2012NA219SR522714D120491Human6-2012NA219SR522716D120491Human6-2012NA219SR522714D120492Human6-2013NA219SR522714D120492Human6-2014NA219SR522714D120493Human6-2014NA219<	ID118551	Human	06-2012	NA	52	10	SRR5227136
ID118700Human06-2012NA5210NRR5227158ID11879Human07-2012NA218SR112-012ID11924Human07-2012NA219SRR5227155ID119366Human07-2012NA16610SRR5227145ID119364Human08-2012NA217SRR5227145ID119559Human08-2012NA219SRR5227147ID119574Human08-2012NA219SRR5227161ID119674Human08-2012NA219SRR5227161ID119674Human08-2012NA219SRR5227161ID119674Human09-2012NA219SRR5227161ID12039Human09-2012NA219SRR5227163ID120404Human09-2012NA219SRR522713ID12047Human09-2012NA219SRR5227163ID12047Human09-2012NA219SRR522713ID12047Human09-2012NA219SRR522713ID12048Human09-2012NA219SRR5227143ID12049Human09-2013NA219SRR5227143ID12049Human09-2013NA219SRR5227143ID124054Human09-2013NA219SRR5227143ID12405	ID118532	Human	06-2012	NA	52	10	SRR5227175
ID118759Human060212NA218SH12-0121D118294Human07-012NA219SLR85271591D119246Human07-012NA210SLR85271451D119464Human08-012NA217SLR85271451D119464Human08-012NA219SLR85271471D119539Human08-012NA219SLR527161D11984Human08-2012NA219SLR527161D11984Human08-2012NA219SLR527161D11984Human09-2012NA219SLR5271321D12058Human09-2012NA219SLR5271321D12059Human10-2012NA219SLR5271321D12058Human12-012NA219SLR5271321D12195Human12-012NA219SLR527161D12195Human04-013NA219SLR527161D12404Human04-013NA219SLR5271421D12405Human04-013NA219SLR5271421D12405Human06-013NA219SLR5271431D12405Human06-013NA219SLR5271431D12405Human06-013NA219SLR5271441D12405 </td <td>ID118700</td> <td>Human</td> <td>06-2012</td> <td>NA</td> <td>52</td> <td>10</td> <td>SRR5227158</td>	ID118700	Human	06-2012	NA	52	10	SRR5227158
ID118979Human07-2012NA29SRR5227189ID119264Human07-2012NA229SRR527145ID119464Human08-2012NA8617SRR527145ID11959Human08-2012NA219SRR527147ID11959Human08-2012NA219SRR527147ID11984Human08-2012NA219SRR527147ID11986Human08-2012NA219SRR527144ID11986Human09-2012NA219SRR527144ID12059Human09-2012NA219SRR52713ID12058Human09-2012NA219SRR52713ID12054Human12-2012NA219SRR52713ID12054Human12-2012NA219SRR52713ID12054Human04-2013NA219SRR52713ID12054Human04-2013NA219SRR52713ID12054Human04-2013NA219SRR52713ID12054Human06-2013NA219SRR52713ID12578Human06-2014NA219SRR52713ID12578Human06-2013NA219SRR52713ID12579Human06-2014NA219SRR52713ID12670Human06-2014	ID118759	Human	06-2012	NA	2	18	SH12-012
ID119224Human07-2012NA229SRR527125ID119366Human07-2012NA18610SRR527184ID119464Human08-2012NA217SRR527145ID11953Human08-2012NA219SRR527147ID119674Human08-2012NA219SRR527147ID11988Human08-2012NA219SRR527147ID119867Human08-2012NA219SRR527144ID120403Human09-2012NA219SRR527142ID12047Human09-2012NA219SRR527132ID120475Human09-2012NA219SRR527142ID120476Human11-2012NA219SRR527149ID121356Human11-2012NA219SRR527149ID12356Human04-2013NA219SRR527142ID12404Human04-2013NA219SRR527142ID124305Human06-2013NA219SRR527142ID124357Human06-2013NA219SRR527142ID12677Human08-2013NA219SRR5227142ID12677Human08-2013NA219SRR5227142ID12677Human08-2013NA219SRR5227142ID12677Hum	ID118979	Human	07-2012	NA	2	19	SRR5227159
ID119366Human07-2012NA18610SRR52718ID11944Human08-2012NA217SRR527147ID119539Human08-2012NA219SRR527170ID119674Human08-2012NA219SRR527170ID119888Human08-2012NA5210SRR527141ID119967Human08-2012NA219SRR527141ID12003Human09-2012NA219SRR527132ID120598Human09-2012NA219SRR527132ID120747Human09-2012NA219SRR527131ID12058Human11-2012NA219SRR527149ID121056Human01-2012NA219SRR527149ID124054Human03-2013NA219SRR527163ID124054Human04-2013NA219SRR527142ID124054Human06-2013NA219SRR527142ID12495Human06-2013NA219SRR527142ID12677Human06-2013NA219SRR527142ID12677Human08-2013NA219SRR527142ID12677Human08-2013NA219SRR527142ID12677Human08-2013NA219SRR527142ID12677Human<	ID119224	Human	07–2012	NA	2	29	SRR5227125
ID119464Human08-2012NA217SRR5227145ID119539Human08-2012NA219SRR5227167ID11967Human08-2012NA5210SRR5227161ID11988Human08-2012NA219SRR5227161ID119967Human09-2012NA219SRR5227160ID120493Human09-2012NA219SRR5227132ID120474Human09-2012NA219SRR5227131ID120474Human09-2012NA219SRR5227132ID120474Human09-2012NA219SRR5227131ID121956Human12-2012NA219SRR5227166ID12405Human03-2013NA219SRR5227153ID12405Human04-2013NA219SRR5227153ID12405Human06-2013NA219SRR5227137ID12677Human06-2013NA219SRR522103ID12677Human08-2013NA219SRR5228081ID12677Human08-2013NA219SRR5228081ID12677Human08-2013NA219SRR5228081ID12677Human08-2013NA219SRR5228081ID12677Human08-2013NA219SRR5228081ID14704<	ID119366	Human	07–2012	NA	186	10	SRR5227138
ID19539Human08-2012NA219SRR527147ID119674Human08-2012NA219SRR5227160ID11988Human08-2012NA219SRR527144ID11967Human09-2012NA219SRR527144ID12059Human09-2012NA219SRR527132ID120747Human09-2012NA219SRR527131ID12058Human09-2012NA219SRR527131ID120574Human09-2012NA219SRR527149ID121956Human11-2012NA219SRR527149ID124024Human03-2013NA219SRR527168ID12405Human04-2013NA219SRR527153ID12405Human04-2013NA219SRR527134ID12405Human06-2013NA219SRR527137ID12677Human06-2013NA219SRR5228081ID12677Human08-2013NA219SRR5228083ID12676Human08-2013NA219SRR5228081ID12676Human08-2013NA219SRR5228081ID12676Human08-2013NA219SRR5228081ID12676Human08-2013NA219SRR5228081ID12676Human </td <td>ID119464</td> <td>Human</td> <td>08-2012</td> <td>NA</td> <td>2</td> <td>17</td> <td>SRR5227145</td>	ID119464	Human	08-2012	NA	2	17	SRR5227145
ID19674Human08-2012NA219SRR5227170ID19888Human08-2012NA5210SRR5227161ID19967Human08-2012NA219SRR5227144ID12003Human09-2012NA219SRR5227132ID120747Human09-2012NA219SRR5227131ID120747Human09-2012NA219SRR5227131ID12075Human11-2012NA219SRR5227131ID12056Human03-2013NA219SRR5227166ID124365Human04-2013NA219SRR5227153ID124305Human04-2013NA229SRR5227137ID12637Human06-2013NA219SRR5227137ID12639Human06-2013NA219SRR5227137ID12639Human06-2013NA219SRR5227137ID12637Human06-2013NA219SRR5228083ID12677Human08-2013NA219SRR5228080ID12677Human08-2013NA219SRR5228081ID12676Human08-2013NA219SRR5228081ID12676Human08-2013NA219SRR5228081ID12670Human02-2016NA219SRR5228081ID147047 <td>ID119539</td> <td>Human</td> <td>08-2012</td> <td>NA</td> <td>2</td> <td>19</td> <td>SRR5227147</td>	ID119539	Human	08-2012	NA	2	19	SRR5227147
ID119888Human08-2012NA5210SRR5227161ID119967Human08-2012NA219SRR5227144ID120403Human09-2012NA219SRR5227160ID12058Human09-2012NA219SRR5227132ID120747Human09-2012NA219SRR5227131ID120747Human09-2012NA219SRR5227131ID120756Human11-2012NA18910SRR5227166ID124305Human03-2013NA229SRR5227168ID124305Human04-2013NA435SRR5227137ID124305Human04-2013NA435SRR5227137ID124305Human06-2013NA219SRS227137ID126378Human06-2013NA219SRS227134ID126771Human08-2013NA219SRS228083ID126776Human08-2013NA219SRS228083ID126776Human08-2013NA219SRS228083ID126776Human08-2013NA219SRS228083ID126776Human08-2013NA219SRS228083ID126776Human08-2013NA219SRS228083ID126776Human02-2016NA219SRS228084ID14	ID119674	Human	08-2012	NA	2	19	SRR5227170
ID119967Human08-2012NA219SRR5227144ID120403Human09-2012NA219SRR5227160ID12058Human09-2012NA219SRR5227132ID120747Human09-2012NA219SRR5227131ID12056Human11-2012NA18910SRR5227149ID122356Human12-2012NA219SRR5227166ID12404Human03-2013NA229SRR5227165ID12405Human04-2013NA435SRR5227142ID12438Human04-2013NA219SRR5227142ID12438Human06-2013NA219SRR5227142ID125378Human06-2013NA219SRR5227137ID126392Human08-2013NA219SRR522809ID12677Human08-2013NA219SRR5228083ID12676Human08-2013NA219SRR5228081ID12676Human08-2013NA219SRR5228085ID12676Human02-2016NA219SRR5228085ID14704Human02-2016NA219SRR5228085ID14705Human02-2016NA2119SRR5228085ID14704Human02-2016NA2119SRR5228086ID14705	ID119888	Human	08-2012	NA	52	10	SRR5227161
ID120403 Human 09-2012 NA 2 19 SRR5227160 ID120598 Human 09-2012 NA 2 19 SRR5227132 ID120747 Human 09-2012 NA 2 19 SRR5227132 ID120747 Human 11-2012 NA 189 10 SRR5227149 ID12356 Human 12-2012 NA 2 19 SRR5227168 ID124024 Human 03-2013 NA 2 19 SR85227168 ID124024 Human 04-2013 NA 2 19 SR85227142 ID124305 Human 04-2013 NA 4 35 SR85227142 ID125712 Human 06-2013 NA 2 19 SR85227137 ID126392 Human 08-2013 NA 2 19 SR8522049 ID126712 Human 08-2013 NA 2 19 SR85228089 ID126776 Human	ID119967	Human	08-2012	NA	2	19	SRR5227144
ID120598 Human 09-2012 NA 2 19 SRR5227132 ID120747 Human 09-2012 NA 2 19 SRR5227131 ID121956 Human 11-2012 NA 189 10 SRR5227149 ID122356 Human 03-2013 NA 2 19 SRR5227168 ID124024 Human 03-2013 NA 2 29 SRR5227168 ID124305 Human 04-2013 NA 4 35 SRR5227142 ID124392 Human 04-2013 NA 4 35 SRR5227142 ID126392 Human 06-2013 NA 2 19 SRR5228081 ID126712 Human 08-2013 NA 2 19 SRR5228081 ID12676 Human 08-2013 NA 2 19 SRR5228081 ID126666 Human 08-2013 NA 2 19 SRR5228081 ID126676 Human	ID120403	Human	09-2012	NA	2	19	SRR5227160
ID120747 Human 09-2012 NA 2 19 SRR5227131 ID121956 Human 11-2012 NA 189 10 SRR5227149 ID122356 Human 12-2012 NA 2 19 SRR5227166 ID124024 Human 03-2013 NA 2 19 SRR5227168 ID124036 Human 04-2013 NA 2 29 SRR5227153 ID124498 Human 04-2013 NA 4 35 SRR5227142 ID125378 Human 06-2013 NA 2 19 SRR5227137 ID126392 Human 07-2013 NA 2 19 SRS228099 ID126712 Human 08-2013 NA 2 19 SRS228081 ID12677 Human 08-2013 NA 2 19 SRS228081 ID12666 Human 08-2013 NA 2 19 SRS228081 ID12667 Human <td< td=""><td>ID120598</td><td>Human</td><td>09-2012</td><td>NA</td><td>2</td><td>19</td><td>SRR5227132</td></td<>	ID120598	Human	09-2012	NA	2	19	SRR5227132
ID121956 Human 11-2012 NA 189 10 SRR5227149 ID122356 Human 12-2012 NA 2 19 SRR5227176 ID124024 Human 03-2013 NA 2 19 SRR5227168 ID124035 Human 04-2013 NA 2 29 SRR5227153 ID12404 Human 04-2013 NA 4 35 SRR5227142 ID124708 Human 06-2013 NA 2 19 SRR5227137 ID126392 Human 06-2013 NA 2 19 SRR5227134 ID126712 Human 08-2013 NA 2 19 SRR5228089 ID126776 Human 08-2013 NA 2 19 SRR5228081 ID12676 Human 08-2013 NA 2 19 SRR5228081 ID12676 Human 08-2013 NA 2 19 SRR5228081 ID12670 Human	ID120747	Human	09-2012	NA	2	19	SRR5227131
ID122356 Human 12-2012 NA 2 19 SRR5227176 ID124024 Human 03-2013 NA 2 19 SRR5227168 ID124035 Human 04-2013 NA 2 29 SRR5227153 ID124498 Human 04-2013 NA 4 35 SRR5227142 ID125378 Human 06-2013 NA 2 19 SRR5227137 ID126392 Human 07-2013 NA 2 19 SRR5227137 ID126712 Human 08-2013 NA 2 19 SRR5228099 ID126777 Human 08-2013 NA 2 19 SRR5228083 ID126776 Human 08-2013 NA 2 19 SRR5228081 ID126825 Human 08-2013 NA 2 29 SRR5228081 ID147047 Human 02-2016 NA 32 SRR5228081 ID14712 ID147010 Human	ID121956	Human	11-2012	NA	189	10	SRR5227149
ID124024 Human 03-2013 NA 2 19 SRR5227168 ID124305 Human 04-2013 NA 2 29 SRR5227153 ID124498 Human 04-2013 NA 4 35 SRR5227142 ID125378 Human 06-2013 NA 2 19 SRR5227137 ID126392 Human 07-2013 NA 2 19 SRR5227134 ID126772 Human 08-2013 NA 2 19 SRR5228099 ID126777 Human 08-2013 NA 2 19 SRR5228083 ID126776 Human 08-2013 NA 2 19 SRR5228081 ID126896 Human 08-2013 NA 2 19 SRR5228081 ID126892 Human 08-2013 NA 2 29 SRR5228081 ID142047 Human 02-2016 NA 32 SRR5228081 ID14701 Human 02-2016	ID122356	Human	12-2012	NA	2	19	SRR5227176
ID124305 Human 04-2013 NA 2 29 SR5227153 ID124498 Human 04-2013 NA 4 35 SR5227142 ID125378 Human 06-2013 NA 2 19 SR5227137 ID126392 Human 07-2013 NA 2 19 SR5227134 ID126712 Human 08-2013 NA 2 19 SR5228089 ID126777 Human 08-2013 NA 2 19 SR5228083 ID126776 Human 08-2013 NA 2 19 SR5228080 ID126696 Human 08-2013 NA 2 19 SR5228081 ID126825 Human 08-2013 NA 2 29 SR5228081 ID147047 Human 02-2016 NA 4 5 SR5228081 ID147047 Human 02-2016 NA 231 32 SR5228084 ID147047 Human 02-2	ID124024	Human	03-2013	NA	2	19	SRR5227168
ID124498 Human 04-2013 NA 4 35 SRR5227142 ID125378 Human 06-2013 NA 2 19 SRR5227137 ID126392 Human 07-2013 NA 2 19 SRR5227134 ID126392 Human 08-2013 NA 2 19 SRR5227134 ID126712 Human 08-2013 NA 2 19 SR5228099 ID126777 Human 08-2013 NA 2 19 SR5228083 ID126776 Human 08-2013 NA 2 19 SR5228080 ID126825 Human 08-2013 NA 2 29 SR5228081 ID147047 Human 02-2016 NA 4 5 SR5228085 ID147047 Human 02-2016 NA 21 19 SR5228084 ID147091 Human 02-2016 NA 2 19 SR5228088 ID14729 Human 02	ID124305	Human	04-2013	NA	2	29	SRR5227153
ID125378 Human 06-2013 NA 2 19 SRR5227137 ID126392 Human 07-2013 NA 2 19 SRR5227134 ID126712 Human 08-2013 NA 2 19 SRR5228099 ID126777 Human 08-2013 NA 2 19 SRR5228083 ID126776 Human 08-2013 NA 2 19 SRR5228080 ID126776 Human 08-2013 NA 2 19 SRR5228081 ID126825 Human 08-2013 NA 2 29 SRR5228081 ID147047 Human 02-2016 NA 4 5 SRR5228081 ID147047 Human 02-2016 NA 231 32 SRR5228084 ID14701 Human 02-2016 NA 21 19 SRR5228084 ID147091 Human 02-2016 NA 22 19 SRR5228086 ID147129 Human	ID124498	Human	04-2013	NA	4	35	SRR5227142
ID126392 Human 07-2013 NA 2 19 SRR5227134 ID126712 Human 08-2013 NA 2 19 SRR5228099 ID126777 Human 08-2013 NA 2 19 SRR5228083 ID126776 Human 08-2013 NA 2 19 SRR5228080 ID126696 Human 08-2013 NA 2 19 SRR5228081 ID126696 Human 08-2013 NA 2 29 SRR5228081 ID126825 Human 08-2013 NA 2 29 SRR5228085 ID147047 Human 02-2016 NA 4 5 SRR5228084 ID14701 Human 02-2016 NA 21 10 SRR5228084 ID14701 Human 02-2016 NA 2 19 SRR5228086 ID147091 Human 02-2016 NA 2 19 SRR5228086 ID14725 Human <	ID125378	Human	06-2013	NA	2	19	SRR5227137
ID126712Human08-2013NA219SRR5228099ID126777Human08-2013NA219SRR5228083ID126776Human08-2013NA219SRR5228080ID126696Human08-2013NA219SRR5228081ID126825Human08-2013NA229SRR5228085ID147047Human02-2016NA45SRR5228084ID147120Human02-2016NA23132SRR5228084ID147091Human02-2016NA5210SRR5228088ID147129Human02-2016NA219SRR5228088ID147253Human02-2016NA219SRR5228092ID147255Human02-2016NA22910SRR5228090ID147457Human02-2016NA22910SRR5228090ID147457Human02-2016NA22519SRR5228094ID147462Human02-2016NA21429SRR5228094ID147900Human03-2016NA5210SRR5228094ID14796Human03-2016NA5210SRR522806ID14796Human03-2016NA5210SRR522806ID14796Human03-2016NA5210SRR522806	ID126392	Human	07-2013	NA	2	19	SRR5227134
ID126777Human08-2013NA219SRR5228083ID126776Human08-2013NA219SRR5228080ID126696Human08-2013NA219SRR5228081ID126825Human08-2013NA229SRR5228085ID147047Human02-2016NA45SRR5228084ID147100Human02-2016NA23132SRR5228084ID147091Human02-2016NA5210SRR5228088ID147129Human02-2016NA219SRR5228086ID147253Human02-2016NA219SRR5228086ID147255Human02-2016NA22910SRR5228092ID147457Human02-2016NA22910SRR5228090ID147457Human02-2016NA22519SRR5228094ID147462Human02-2016NA21429SRR5228094ID147900Human03-2016NA5210SRR5228106ID147766Human03-2016NA5210SRR5228106	ID126712	Human	08-2013	NA	2	19	SRR5228099
ID126776Human08-2013NA219SRR5228080ID126696Human08-2013NA219SRR5228081ID126825Human08-2013NA229SRR5228085ID147047Human02-2016NA45SRR5228102ID147120Human02-2016NA23132SRR5228084ID147017Human02-2016NA5210SRR5228084ID147091Human02-2016NA5210SRR5228088ID147129Human02-2016NA219SRR5228092ID147253Human02-2016NA19429SRR5228086ID147255Human02-2016NA22910SRR5228090ID147457Human02-2016NA22910SRR5228090ID147452Human02-2016NA22519SRR5228092ID147462Human02-2016NA21429SRR5228094ID147900Human03-2016NA5210SRR5228095ID147990Human03-2016NA5210SRR5228106ID147796Human03-2016NA5210SRR5228106	ID126777	Human	08-2013	NA	2	19	SRR5228083
ID126696Human08-2013NA219SRR5228081ID126825Human08-2013NA229SRR5228085ID147047Human02-2016NA45SRR5228085ID147120Human02-2016NA23132SRR5228084ID147091Human02-2016NA5210SRR5228088ID147129Human02-2016NA5210SRR5228088ID147129Human02-2016NA219SRR5228086ID147253Human02-2016NA22910SRR5228090ID147255Human02-2016NA22910SRR5228090ID147457Human02-2016NA22519SRR5228089ID147462Human02-2016NA21429SRR5228094ID147990Human03-2016NA5210SRR5228095	ID126776	Human	08-2013	NA	2	19	SRR5228080
ID126825Human08-2013NA229SRR5228085ID147047Human02-2016NA45SRR5228102ID147120Human02-2016NA23132SRR5228084ID147091Human02-2016NA5210SRR5228088ID147129Human02-2016NA219SRR5228088ID147129Human02-2016NA219SRR5228086ID147253Human02-2016NA22910SRR5228086ID147255Human02-2016NA22519SRR5228089ID147457Human02-2016NA21429SRR5228094ID147990Human03-2016NA5210SRR5228094ID147796Human03-2016NA5210SRR5228106	ID126696	Human	08-2013	NA	2	19	SRR5228081
ID147047Human02-2016NA45SRR5228102ID147120Human02-2016NA23132SRR5228084ID147091Human02-2016NA5210SRR5228088ID147129Human02-2016NA219SRR5228092ID147253Human02-2016NA19429SRR5228086ID147255Human02-2016NA22910SRR5228090ID147457Human02-2016NA22519SRR5228089ID147462Human02-2016NA21429SRR5228094ID147990Human03-2016NA5210SRR5228106ID147296Human03-2016NA5210SRR5228106	ID126825	Human	08-2013	NA	2	29	SRR5228085
D147120 Human D2-2016 NA 231 32 SRR5228084 ID147120 Human 02-2016 NA 52 10 SRR5228088 ID147091 Human 02-2016 NA 52 10 SRR5228088 ID147129 Human 02-2016 NA 2 19 SRR5228092 ID147253 Human 02-2016 NA 194 29 SRR5228086 ID147255 Human 02-2016 NA 229 10 SRR5228090 ID147257 Human 02-2016 NA 225 19 SRR5228089 ID147457 Human 02-2016 NA 214 29 SRR5228094 ID147462 Human 02-2016 NA 214 29 SRR5228094 ID147990 Human 03-2016 NA 52 10 SRR5228106 ID147796 Human 03-2016 NA 52 10 SRR5228106	ID147047	Human	02-2016	NA	4	5	SRR5228102
ID147091 Human 02-2016 NA 52 10 SRR5228088 ID147129 Human 02-2016 NA 2 19 SRR5228092 ID147253 Human 02-2016 NA 194 29 SRR5228086 ID147253 Human 02-2016 NA 194 29 SRR5228086 ID147255 Human 02-2016 NA 229 10 SRR5228089 ID147255 Human 02-2016 NA 225 19 SRR5228089 ID147457 Human 02-2016 NA 214 29 SRR5228094 ID147462 Human 02-2016 NA 214 29 SRR5228094 ID147990 Human 03-2016 NA 52 10 SRR5228106 ID147796 Human 03-2016 NA 52 10 SRP5228005	ID147120	Human	02-2016	NA	231	32	SRR5228084
ID147051 Human 02-2016 NA 2 10 SRR5228092 ID147129 Human 02-2016 NA 2 19 SRR5228092 ID147253 Human 02-2016 NA 194 29 SRR5228086 ID147255 Human 02-2016 NA 229 10 SRR5228090 ID147457 Human 02-2016 NA 225 19 SRR5228089 ID147462 Human 02-2016 NA 214 29 SRR5228094 ID147990 Human 03-2016 NA 52 10 SRR5228106 ID147796 Human 03-2016 NA 52 10 SRR5228106	ID147091	Human	02-2016	NA	52	10	SRR5228088
ID147253 Human 02-2016 NA 194 29 SRR5228086 ID147255 Human 02-2016 NA 229 10 SRR5228090 ID147257 Human 02-2016 NA 225 19 SRR5228089 ID147457 Human 02-2016 NA 214 29 SRR5228094 ID147462 Human 02-2016 NA 214 29 SRR5228094 ID147990 Human 03-2016 NA 52 10 SRR5228106 ID147796 Human 03-2016 NA 52 10 SRR5228095	ID147129	Human	02-2016	NA	2	19	SRR5228092
ID147255 Human 02-2016 NA 229 10 SRR5228090 ID147457 Human 02-2016 NA 225 19 SRR5228089 ID147462 Human 02-2016 NA 214 29 SRR5228094 ID147990 Human 03-2016 NA 52 10 SRR5228106 ID147786 Human 03-2016 NA 52 10 SRR5228005	ID147253	Human	02-2016	NA	194	29	SRR5228086
ID147457 Human 02-2016 NA 225 19 SRR5228089 ID147462 Human 02-2016 NA 214 29 SRR5228094 ID147990 Human 03-2016 NA 52 10 SRR5228106 ID147796 Human 03-2016 NA 52 10 SRR5228106	ID147255	Human	02-2016	NA	229	10	SRR5228090
ID147462 Human 02-2016 NA 214 29 SRR5228094 ID147990 Human 03-2016 NA 52 10 SRR5228106 ID147796 Human 03-2016 NA 52 10 SRR5228106	ID147457	Human	02-2016	NA	225	19	SRR5228089
ID147990 Human 03-2016 NA 52 10 SRR5228106 ID147796 Human 03-2016 NA 52 10 SRR5228106	ID147462	Human	02-2016	NA	214	29	SRR5228094
Initial Initia	ID147990	Human	03-2016	NA	52	10	SRR5228106
$10^{14}/70$ Huilall $0^{-2}010$ INA 5^{-1} 10 $5^{-1}8^{-1}8^{-1}$	ID147796	Human	03-2016	NA	52	10	SRR5228095

(Continued)

Table 2. (Continued)

Isolate No.	Source	Isolation date	Outbreak code	Pulsotype	Phage type	NCBI accession no.
ID147816	Human	03-2016	NA	2	19	SRR5228103
ID147910	Human	03-2016	NA	214	29	SRR5228077
ID147841	Human	03-2016	NA	2	19	SRR5228096
ID148066	Human	03-2016	NA	2	19	SRR5228098

NA, Not available.

https://doi.org/10.1371/journal.pone.0192233.t002

Topological similarity of the phylogenetic trees

To confirm the similarities in tree topologies, we performed the RF test. The computed RF topological distances between the 14 trees ranged from 0 to 24 (Table 6).

Discussion

In this study we assessed the impact of the choice of the reference genome on the resolution clustering of *S*. Heidelberg outbreak and sporadic isolates using the cgSNV methodology. Our results revealed that using a draft or completely sequenced *S*. Heidelberg genomes as references did not affect the ability of the cgSNV method to distinguish between four epidemiologically well characterized *S*. Heidelberg outbreak isolates and to separate these isolates from sporadic or background strains. In fact, all outbreak and sporadic isolates were clustered on distinct branches (Fig 1A and 1B). On the contrary, using *S*. Dublin as reference choice resulted in a tree with less resolution (Fig 2). Sporadic isolates clustered together with outbreak 1 isolates and in addition, the genetic distances observed within outbreak as well as between outbreak and sporadic isolates was overall reduced when *S*. Dublin was used as reference choice as opposed to *S*. Heidelberg reference genomes (Figs 1 and 2, Table 5). This finding is in agreement with a recent study on *Salmonella* which found that using a distantly related genome as a choice of reference failed to cluster *S*. Enteritidis outbreak strains concordantly [18]. These observations emphasize the need to choose an appropriate reference genome during laboratory investigations of foodborne outbreaks involving reference based methods.

The loss in resolution observed in this study can be due to the following reasons: Firstly, the SNVPhyl pipeline uses only core genome SNVs to build the phylogeny implying that a reference genome with high similarity with the isolates under investigation would have a larger core genome from which more SNVs can be produced to build a phylogeny as opposed to a dissimilar reference such as *S*. Dublin. However, the smallest core genome among all the reference genomes used in this study was equivalent to 84.76% of the total genome. Secondly, another issue to consider is that as the reference genome grows more distant from the sequences under analysis, more variation would be observed between the core regions of the reference genome and all other sequences leading to a long branch separating the reference

Strain	Total length (bp)	No of contigs	N50 (bp)	Coverage
ID117795	4,751,241	24	694,16	137
ID128787	4,747,971	27	363,273	43
ID128902	4,746,565	27	412,159	64
ID134609	4,853,519	27	412,096	130
ID135140	4,753,550	27	412,162	116

Table 3. Assembly statistics for the 5 S. Heidelberg isolates that served as reference genomes.

https://doi.org/10.1371/journal.pone.0192233.t003



Fig 1. Minimum spanning phylogenetic tree of the core genome of 145 S. Heidelberg sequenced isolates generated using A) draft genome or B) closed referenced genome (ID117795) as an example. Isolates in the same circle have 0 hqSNVs and the size of each circle is proportional to the number of isolates in the circle.

https://doi.org/10.1371/journal.pone.0192233.g001

PLOS ONE

genome from the rest of the other isolates. This was indeed what we observed using *S*. Dublin as reference (Fig 2) since the majority of the 18155 positions used to construct the phylogeny were indeed variations between the reference (17666 positions) and the other isolates.

The RF values for the trees constructed using *S*. Heidelberg draft genome and its corresponding closed genome equivalent was zero with the exception of the reference genome pair ID134609 (tree 7 vs. tree 8) whereby the RF topological distance was 24. This difference could be attributed to misidentification of SNVs linked to the presence of repetitive regions in the draft genome that were not properly detected. By nature, draft genomes are not entirely genomically accurate and for this reason, it is possible that the SPAdes assembly for this isolate may have collapsed repetitive regions larger than the read/read pair size into a single contig. Aligning reads to repetitive regions is problematic and has been reported to lead to potential



Fig 2. Minimum spanning phylogenetic tree of the core genome of 145 S. Heidelberg sequenced isolates generated using the distantly related reference S. Derby (SL477). Isolates in the same circle have 0 hqSNVs and the size of each circle is proportional to the number of isolates in the circle.

https://doi.org/10.1371/journal.pone.0192233.g002

PLOS ONE

misidentification of SNVs [19]. Although we used the software MUMmer to identify and filter the repetitive regions of the reference genome, for a closed genome, this method will be much more successful due to its intact, completely mapped sequence than for a draft genome, which may erroneously possess several copies of unmapped repeat regions. These unmapped repeats can lead to repeat reads mapping to a single contig resulting in the inclusion of additional SNVs and subsequent alignment issues and false SNV calls. Interestingly in this reference pair, 766 sites were used to generate the phylogeny using the draft genome as reference as opposed to 761 sites for the closed genome counterpart. Despite the slight differences in topology between this reference pair as well as in other tree pairs, both the draft and closed genomes

Dafarances salactad	from within	n the outbre	alr icolatae								Dafarancae		from NCB)		Distantly valated
														4	genome downloaded from NCBI
Minimum	ID117795		ID128787		ID128902		ID134609		ID135140		SL486	IWQE01.1	SL476	CP003416.1	CP001144
spanning tree features	Draft	Closed	Draft	Closed	Draft	Closed	Draft	Closed	Draft	Closed	Draft	Draft	Closed	Closed	Closed
Total # of nodes on the tree	92	92	92	92	92	92	93	93	93	93	92	93	93	93	86
Total # of nodes representing 1 isolate	79	79	79	79	79	79	80	80	80	80	78	79	80	80	72
Total # of nodes with more than one isolate	13	13	13	13	13	13	13	13	13	13	13	14	13	13	15
Total # of OB nodes	13	13	13	13	13	13	13	13	13	13	13	13	13	13	10
Total # of nodes with one isolate	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
Total # of OB nodes with more than one isolate	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Total # of sporadic nodes	79	79	79	79	79	79	80	80	80	80	78	79	80	80	76
Total # of sporadic nodes linked to main OB clusters	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
Number of sites used to generate phylogeny	762	762	761	760	762	760	766	761	764	763	777	799	751	819	18155
Number of genomic positions used for reference mapping	4,046,580	4,037,652	4,046,584	4,037,362	4,047,586	4,037,102	4,049,343	4,037,849	4,046,484	4,037,788	4,008,004	4,048,063	3,925,643	4,036,335	3,681,444
% of genomic positions representing the genome	86,04%	86,68%	86,07%	86,7%	86,08%	86,69%	85%	86,72%	86%	86,72%	85,95%	86,12%	85,72%	86,67%	84,63%
https://doi.org/10.137	71/journal.pc	one.0192233	1.t004												

PLOS ONE

DraftClosedDraftClosedDraftClosedDraftClosedI10-30-372-7471-7348-5148-5218-2118-2118-21272-7471-730066-6765-6780-8179-805348-5148-5266-6765-670-20-356-5856-590-10-11012878710-30-371-7371-7347-5047-5017-2017-2017-20271-7371-730066-6766-6780-8180-819347-5047-5066-6766-6766-6780-8180-819417-2017-2080-8180-8156-5856-5861-10-11012890210-30-372-7471-7348-5247-5017-2017-20272-7471-730066-6866-6781-8280-819348-5247-5066-6866-670-30-257-6056-5861012890210-30-372-7471-7348-5247-5017-2017-201012890210-30-372-7471-7348-5247-5017-2017-201013514010-30-372-7471-7348-5247-5017-2017-201013514010-30-372-7471-73 <td< th=""><th>Draft 1-93 4-105 6-86 10-100 1-92 4-105 6-86 10-100 1-92 4-105 6-86 10-100 1-92 4-105 6-86 10-101 1-93 4-105 6-87 10-100 1-93 4-105</th><th>Closed 1–93 4–104 6–87 10–100 1–92 4–105 6–86 10–100 1–92 4–105 6–86 10–100 1–92 4–105 6–87 10–100 1–92 4–105</th></td<>	Draft 1-93 4-105 6-86 10-100 1-92 4-105 6-86 10-100 1-92 4-105 6-86 10-100 1-92 4-105 6-86 10-101 1-93 4-105 6-87 10-100 1-93 4-105	Closed 1–93 4–104 6–87 10–100 1–92 4–105 6–86 10–100 1–92 4–105 6–86 10–100 1–92 4–105 6–87 10–100 1–92 4–105
ID117795 I O-3 O-3 72-74 71-73 48-51 48-52 18-21 18-21 1 2 72-74 71-73 0 0 66-67 65-67 80-81 79-80 5 4 18-21 18-21 80-81 79-80 56-58 56-59 0-1 0-1 0 10128787 1 0-3 0-3 71-73 71-73 47-50 47-50 17-20 17-20 2 71-73 71-73 0 0 66-67 66-67 80-81 80-81 5 4 17-20 17-20 80-81 80-81 56-58 56-58 0-1 0-1 1 0 1 0-1 1 0-1 1 0-1 1 0-1 1 0-1 1 0-1 1 0-1 1 0 1 0-1 1 0 1 0 1 0 1 0 1 0 1 0	1-93 1-93 4-105 6-86 10-100 1-92 4-105 6-86 10-100 1-92 4-105 6-86 10-100 1-92 4-105 6-86 10-101 1-93 4-105 6-87 10-100 1-93 4-105	$\begin{array}{c} 1-93\\ 1-93\\ 4-104\\ 6-87\\ 10-100\\ 1-92\\ 4-105\\ 6-86\\ 10-100\\ 1-92\\ 4-105\\ 6-86\\ 10-100\\ 1-92\\ 4-105\\ 6-87\\ 10-100\\ 1-92\\ 4-105\\ 4-105\\ \end{array}$
1 0 0 1 0 0 0 66-67 65-58 56-59 0 1 0<-1	4-105 6-86 10-100 1-92 4-105 6-86 10-100 1-92 4-105 6-86 10-101 1-93 4-105 6-87 10-100 1-93 4-105	$\begin{array}{c} 1 & 93 \\ \hline 1 & 93 \\ \hline 4 - 104 \\ \hline 6 - 87 \\ \hline 10 - 100 \\ \hline 1 - 92 \\ \hline 4 - 105 \\ \hline 6 - 86 \\ \hline 10 - 100 \\ \hline 1 - 92 \\ \hline 4 - 105 \\ \hline 6 - 86 \\ \hline 10 - 100 \\ \hline 1 - 92 \\ \hline 4 - 105 \\ \hline 6 - 87 \\ \hline 10 - 100 \\ \hline 1 - 92 \\ \hline 4 - 105 \\ \hline \end{array}$
3 48-51 14-52 66-67 65-67 0-2 0-3 56-58 56-59 0 4 18-21 18-21 80-81 79-80 56-58 56-59 0-1 0-1 10128787 1 0-3 0-3 71-73 71-73 47-50 47-50 17-20 17-20 17-20 2 71-73 71-73 0 0 66-67 66-67 66-67 80-81 80-81 56-58 56-57 56-58 56-58 56-57 56	6-86 10-100 1-92 4-105 6-86 10-100 1-92 4-105 6-86 10-101 1-93 4-105 6-87 10-100 1-93 4-105 6-87 10-100 1-93 4-105	$\begin{array}{c} -87\\ 6-87\\ 10-100\\ 1-92\\ 4-105\\ 6-86\\ 10-100\\ 1-92\\ 4-105\\ 6-86\\ 10-100\\ 1-92\\ 4-105\\ 6-87\\ 10-100\\ 1-92\\ 4-105\\ \end{array}$
A 18-21 18-21 80-81 79-80 56-58 56-59 0-1 0-1 ID128787 1 0-3 71-73 71-73 71-73 47-50 47-50 17-20 17-20 17-20 2 71-73 71-73 0 0 66-67 66-67 80-81 80-81 90-1 3 47-50 47-50 66-67 66-67 0-2 0-2 56-58 56-58 65-58 65-58 65-58 65-58 65-58 65-58 0-1 0-1 1 10128902 1 0-3 0-3 72-74 71-73 48-52 47-50 17-20 17-20 17-20 3 48-52 47-50 66-68 66-67 0-3 0-2 57-60 56-58 0-1 0-1 1D135140 1 0-3 0-3 72-74 71-73 48-52 47-51 17-20 17-20 17-20 1D135609 1 0-3 0-3 72	10-100 1-92 4-105 6-86 10-100 1-92 4-105 6-86 10-101 1-93 4-105 6-87 10-100 1-93 4-105	$\begin{array}{c} 10 - 100 \\ 1 - 92 \\ 4 - 105 \\ 6 - 86 \\ 10 - 100 \\ 1 - 92 \\ 4 - 105 \\ 6 - 86 \\ 10 - 100 \\ 1 - 92 \\ 4 - 105 \\ 6 - 87 \\ 10 - 100 \\ 1 - 92 \\ 4 - 105 \\ \end{array}$
ID128787 I 0-3 0-3 71-73 71-73 47-50 47-50 47-50 47-50 17-2	1-92 1-92 4-105 6-86 10-100 1-92 4-105 6-86 10-101 1-93 4-105 6-87 10-100 1-93 4-105	$\begin{array}{c} 1 - 92 \\ 1 - 92 \\ 4 - 105 \\ 6 - 86 \\ 10 - 100 \\ 1 - 92 \\ 4 - 105 \\ 6 - 86 \\ 10 - 100 \\ 1 - 92 \\ 4 - 105 \\ 6 - 87 \\ 10 - 100 \\ 1 - 92 \\ 4 - 105 \end{array}$
1 1	4-105 6-86 10-100 1-92 4-105 6-86 10-101 1-93 4-105 6-87 10-100 1-93 4-105	$\begin{array}{c} 4-105\\ 6-86\\ 10-100\\ 1-92\\ 4-105\\ 6-86\\ 10-100\\ 1-92\\ 4-105\\ 6-87\\ 10-100\\ 1-92\\ 4-105\\ 4-105\\ \end{array}$
3 47-50 47-50 66-67 66-67 0-2 0-2 56-58 56-58 1 ID128902 1 0-3 0-3 72-74 71-73 48-52 47-50 17-20 17-20 2 72-74 71-73 0 0 66-68 66-67 81-82 80-81 56-58 56-58 17-20 17-20 17-20 3 48-52 47-50 66-68 66-67 0-3 0-2 57-60 56-58 0-1 0-1 4 17-20 17-20 81-82 80-81 57-60 56-58 0-1 0-1 10135140 1 0-3 0-3 72-74 71-73 48-52 47-51 17-20 17-20 3 48-52 47-51 66-68 66-68 0-3 0-3 58-60 58-59 0-1 0-1 10135609 1 0-3 0-3 72-74 71-73 48-52 47-50 56-58 0-1 0-1	6-86 10-100 1-92 4-105 6-86 10-101 1-93 4-105 6-87 10-100 1-93 4-105	6-86 10-100 1-92 4-105 6-86 10-100 1-92 4-105 6-87 10-100 1-92 4-105
4 17-20 17-20 80-81 80-81 56-58 56-58 0-1 0-1 ID128902 1 0-3 0-3 72-74 71-73 48-52 47-50 17-20	10-100 1-92 4-105 6-86 10-101 1-93 4-105 6-87 10-100 1-93 4-105	10-100 1-92 4-105 6-86 10-100 1-92 4-105 6-87 10-100 1-92 4-105
ID128902 1 0-3 0-3 72-74 71-73 48-52 47-50 17-20 17-20 17-20 2 72-74 71-73 0 0 66-68 66-67 81-82 80-81 7 3 48-52 47-50 66-68 66-67 0-3 0-2 57-60 56-58 0-1 0-1 4 17-20 17-20 81-82 80-81 57-60 56-58 0-1 0-1 10135140 1 0-3 0-3 72-74 71-73 48-52 47-51 17-20 17-20 2 72-74 71-73 0 0 66-68 66-68 61-68 81-82 80-81 7 3 48-52 47-51 66-68 66-68 0-3 0-3 58-60 58-50 0-1 0-1 10135609 1 0-3 0-3 72-74 71-73 48-52 47-50 17-20 17-20 10135609 1	1-92 4-105 6-86 10-101 1-93 4-105 6-87 10-100 1-93 4-105	1-92 4-105 6-86 10-100 1-92 4-105 6-87 10-100 1-92 4-105
2 72-74 71-73 0 0 66-68 66-67 81-82 80-81 7 3 48-52 47-50 66-68 66-67 0-3 0-2 57-60 56-58 0-1 0-1 0 4 17-20 17-20 81-82 80-81 57-60 56-58 0-1 0-1 0 2 72-74 71-73 0 0 66-68 66-68 81-82 87-81 7 2 72-74 71-73 0 0 66-68 66-68 81-82 80-81 7 3 48-52 47-51 66-68 66-68 0-3 0-3 58-60 58-50 0-1 0-1 0 4 17-20 17-20 81-82 80-81 58-60 58-50 0-1 0-1 10135609 1 0-3 0.3 72-74 71-73 0 0 66-67 0-3 0-2 57-60 56-58 0-1 0-1	4-105 6-86 10-101 1-93 4-105 6-87 10-100 1-93 4-105	4-105 6-86 10-100 1-92 4-105 6-87 10-100 1-92 4-105
3 48-52 47-50 66-68 66-67 0-3 0-2 57-60 56-58 0 10135140 1 0-3 0-3 72-74 71-73 48-52 47-51 17-20 10-1	6-86 10-101 1-93 4-105 6-87 10-100 1-93 4-105	6-86 10-100 1-92 4-105 6-87 10-100 1-92 4-105
4 17-20 17-20 81-82 80-81 57-60 56-58 0-1 0-1 ID135140 1 0-3 0-3 72-74 71-73 48-52 47-51 17-20 17-20 17-20 17-20 17-20 17-20 17-20 17-20 17-20 17-20 17-20 17-20 17-20 66-68 66-68 66-68 66-68 81-82 80-81 40-3 9-3 58-60 58-50 0-1 0-1 17-20 <th1< th=""><td>10-101 1-93 4-105 6-87 10-100 1-93 4-105</td><td>10-100 1-92 4-105 6-87 10-100 1-92 4-105</td></th1<>	10-101 1-93 4-105 6-87 10-100 1-93 4-105	10-100 1-92 4-105 6-87 10-100 1-92 4-105
ID135140 1 0-3 0-3 72-74 71-73 48-52 47-51 17-20 17-20 1 2 72-74 71-73 0 0 66-68 66-68 81-82 80-81 4 3 48-52 47-51 66-68 66-68 0-3 0-3 58-60 58-59 0-1 0-1 4 17-20 17-20 81-82 80-81 58-60 58-59 0-1 0-1 1D135609 1 0-3 0-3 72-74 71-73 48-52 47-50 17-20 17-20 17-20 2 72-74 71-73 0 0 66-68 66-67 81-82 80-81 7 3 48-52 47-50 66-68 66-67 0-3 0-2 57-60 56-58 0-1 0-1 417-20 ND 71-73 ND 46-49 ND 17-20 ND 46-49 ND 17-20 ND 46-49 ND	1-93 4-105 6-87 10-100 1-93 4-105	1-92 4-105 6-87 10-100 1-92 4-105
2 72-74 71-73 0 0 66-68 66-68 81-82 80-81 4 3 48-52 47-51 66-68 66-68 0-3 0-3 58-60 58-59 0-1 0-1 4 17-20 17-20 81-82 80-81 58-60 58-59 0-1 0-1 0-1 1D135609 1 0-3 0-3 72-74 71-73 48-52 47-50 17-20 17-20 17-20 2 72-74 71-73 0 0 66-68 66-67 81-82 80-81 0 3 48-52 47-50 66-68 66-67 0-3 0-2 57-60 56-58 0-1 0-1 SL486 1 0-3 ND 71-73 ND 46-49 ND 17-20 ND 65-66 0-2 55-57 0-1 0 JWQE01.1 1 0-3 ND 71-73 ND 46-49 ND 17-20 <t< th=""><td>4-105 6-87 10-100 1-93 4-105</td><td>4-105 6-87 10-100 1-92 4-105</td></t<>	4-105 6-87 10-100 1-93 4-105	4-105 6-87 10-100 1-92 4-105
3 48-52 47-51 66-68 66-68 0-3 0-3 58-60 58-59 0 4 17-20 17-20 81-82 80-81 58-60 58-59 0-1 0-1 0 1D135609 1 0-3 0-3 72-74 71-73 48-52 47-50 17-20 17-20 17-20 2 72-74 71-73 0 0 66-68 66-67 0-3 0-2 57-60 56-58 0-1 0-1 3 48-52 47-50 66-68 66-67 0-3 0-2 57-60 56-58 0-1 0-1 44 17-20 17-20 81-82 80-81 57-60 56-58 0-1 0-1 51486 1 0-3 ND 71-73 ND 46-49 ND 17-20 ND 4 17-20 80-81 55-57 0-1 0 0 65-66 0-2 55-57 0-1 0 0 0	6-87 10-100 1-93 4-105	6-87 10-100 1-92 4-105
4 17-20 17-20 81-82 80-81 58-60 58-59 0-1 0-1 ID135609 1 0-3 0-3 72-74 71-73 48-52 47-50 17-20 10-1	10-100 1-93 4-105	10-100 1-92 4-105
ID135609 1 0-3 0-3 72-74 71-73 48-52 47-50 17-20 <th17-20< th=""> <th17-20< th=""> <th17-20< <="" th=""><td>1–93 4–105</td><td>1–92 4–105</td></th17-20<></th17-20<></th17-20<>	1–93 4–105	1–92 4–105
2 72-74 71-73 0 0 66-68 66-67 81-82 80-81 4 3 48-52 47-50 66-68 66-67 0-3 0-2 57-60 56-58 0 4 17-20 17-20 81-82 80-81 57-60 56-58 0-1 0-1 SL486 1 0-3 ND 71-73 ND 46-49 ND 17-20 ND 46-49 3 46-49 0 65-66 0-2 55-57 0-1 0 4 17-20 ND 71-73 ND 46-49 ND 17-20 ND 46-49 4 17-20 80-81 55-57 0-1 0 JWQE01.1 1 0-3 ND 71-73 ND 46-49 ND 17-20 ND 46-49 JWQE01.1 1 0-3 ND 71-73 ND 46-49 ND 17-20 ND 46-49 0-1 <t< th=""><td>4-105</td><td>4-105</td></t<>	4-105	4-105
3 48-52 47-50 66-68 66-67 0-3 0-2 57-60 56-58 6 4 17-20 17-20 81-82 80-81 57-60 56-58 0-1 0-1 0 SL486 1 0-3 ND 71-73 ND 46-49 ND 17-20 ND 65-66 0-2 55-57 0 0 65-66 0-2 55-57 0		
4 17-20 17-20 81-82 80-81 57-60 56-58 0-1 0-1 SL486 1 0-3 ND 71-73 ND 46-49 ND 17-20 ND 46-49 3 46-49 65-66 0-2 55-57 0-1 0 4 17-20 80-81 55-57 0-1 0 66-67 0-2 55-57 0-1 0 JWQE01.1 1 0-3 ND 71-73 ND 46-49 ND 17-20 ND 46-49 JWQE01.1 1 0-3 ND 71-73 ND 46-49 ND 17-20 ND 46-49 3 46-49 0 66-67 0-2 55-57 0-1 0 3 46-49 0-3 ND 68-70 0 66-67 79-80 55-57 0-1 0 \$SL476 1 ND 0-3 ND 68-70 ND 46-49 ND <td>6-86</td> <td>6-86</td>	6-86	6-86
SL486 1 0-3 ND 71-73 ND 46-49 ND 17-20 ND 2 71-73 0 65-66 0-2 55-57 0-1 3 46-49 80-81 55-57 0-1 ND 17-20 ND 17-20 ND 4 17-20 80-81 55-57 0-1 ND 17-20 17-20 17-20 17-20 17-20 17-20 17-20 17-20 17-20 17-20 <th< th=""><td>10-101</td><td>10-100</td></th<>	10-101	10-100
2 71-73 0 65-66 80-81 3 46-49 65-66 0-2 55-57 0-1 4 17-20 80-81 55-57 0-1 ND JWQE01.1 1 0-3 ND 71-73 ND 46-49 ND 17-20 ND 2 70-72 0 66-67 0-2 55-57 0-1 ND 46-49 ND 17-20 ND 46-49 ND 17-20 ND 46-49 0-2 55-57 0-1 40 3 46-49 66-67 0-2 55-57 0-1 40	1–92	ND
3 46-49 65-66 0-2 55-57 0-1 JWQE01.1 1 0-3 ND 71-73 ND 46-49 ND 17-20 ND 2 70-72 0 66-67 0-2 55-57 0-1 ND 1 3 46-49 ND 71-73 ND 46-49 ND 17-20 ND 4 4 17-20 79-80 55-57 0-1 4 4 17-20 1	4-105	1
4 17-20 80-81 55-57 0-1 I JWQE01.1 1 0-3 ND 71-73 ND 46-49 ND 17-20 ND 1 2 70-72 0 66-67 0-2 55-57 0-1 1	6-85	1
JWQE01.1 1 0-3 ND 71-73 ND 46-49 ND 17-20 ND 4 2 70-72 0 66-67 0-2 55-57 0-1 10000 1000 10000	13-100	
2 70-72 0 66-67 79-80 79-80 3 46-49 66-67 0-2 55-57 0-1 1 4 17-20 79-80 55-57 0-1 17-20 SL476 1 ND 0-3 ND 68-70 0 65-66 80-81 3 46-49 65-66 0-2 55-57 0-1 17-20 4 17-20 80-81 55-57 0-1 17-20	1–91	ND
3 46-49 66-67 0-2 55-57 0-1 4 17-20 79-80 55-57 0-1 1 SL476 1 ND 0-3 ND 68-70 ND 46-49 ND 17-20 3 46-49 65-66 0-2 55-57 0-1 17-20 4 17-20 80-81 55-57 0-1 17-20	4-105]
4 17-20 79-80 55-57 0-1 I SL476 1 ND 0-3 ND 68-70 ND 46-49 ND 17-20 3 46-49 46-49 0 65-66 0-2 55-57 0-1 4 17-20 80-81 55-57 0-1 17-20	6-86	1
SL476 1 ND 0-3 ND 68-70 ND 46-49 ND 17-20 2 68-70 0 65-66 0-2 55-57 55-57 4 17-20 80-81 55-57 0-1	10-99	
2 68-70 0 65-66 80-81 3 46-49 65-66 0-2 55-57 4 17-20 80-81 55-57 0-1	ND	1-92
3 46-49 65-66 0-2 55-57 4 17-20 80-81 55-57 0-1		4-105
4 17–20 80–81 55–57 0–1		6-85
		13-100
CP003416.1 1 ND 0-3 ND 72-74 ND 47-50 ND 17-20	ND	1-92
2 72–74 0 67–68 81–82		4-106
3 47-50 67-68 0-2 56-58		6-86
4 17-20 81-82 56-58 0-1	1	10-100
CP001144 1 ND 0-2 ND 45-47 ND 28-31 ND 10-12		0-60
2 45-47 0 39-40 49-49	ND	2-67
3 28-31 39-40 0-2 32-34	ND	1 - 07
4 10-12 49-49 32-34 0	ND	3-53

Table 5. Comparison of the number of high quality SNVs between 145 S. Heidelberg sporadic and outbreak isolates using a draft, closed and distantly related reference genomes.

^aThe reference genomes were obtained from: *S.* Heidelberg outbreak isolates [ID117795, ID128787, ID128902, ID135140, and ID135609]; publicly-available *S.* Heidelberg references from NCBI [SL486, JWQE01.1, SL476, CP003416.1] and distantly-related *S.* Dublin reference from NCBI [CP001144]. ND, Not Determined.

https://doi.org/10.1371/journal.pone.0192233.t005

•			0				·	•					
	Tree 2	Tree 3	Tree 4	Tree 5	Tree 6	Tree 7	Tree 8	Tree 9	Tree 10	Tree 11	Tree 12	Tree 13	Tree 14
Tree 1	0	20	20	20	20	20	20	20	20	7	11	15	11
Tree 2		20	20	20	20	20	20	20	10	7	11	15	11
Tree 3			0	0	0	24	0	24	24	13	11	15	11
Tree 4				0	0	24	0	24	24	13	11	15	11
Tree 5					0	24	0	24	24	13	11	15	11
Tree 6						24	0	24	24	13	11	15	11
Tree 7							24	0	0	13	15	19	15
Tree 8								24	24	13	11	15	11
Tree 9									0	13	15	19	15
Tree 10										13	15	19	15
Tree 11											4	8	4
Tree 12												4	0
Tree 13													4

Table 6. Robinson-Foulds topological distances between trees generated with 145 S. Heidelberg sequenced isolates using draft and closed genomes as references during cgSNV analysis.

https://doi.org/10.1371/journal.pone.0192233.t006

were still able to differentiate the outbreak from sporadic isolates in concordance with epidemiological data. In agreement with our observations, a recent study on *Listeria monocytogenes* also demonstrated that phylogenetic clustering based on SNV analysis using a *de novo* assembled draft genome selected from within the group was similar to the phylogeny using a closely related closed genome [6].

Despite the increasing accessibility of WGS-based technologies and decreasing costs, the implementation of WGS for routine surveillance and outbreak investigation may still present many challenges as sequencing a genome to completion remains a costly and time-consuming endeavour and for many public-health laboratories, this is not a viable option. [20]. The results of our study indicates that draft genomes can be relied upon as a suitable reference choice during laboratory investigations of foodborne outbreaks using the cgSNVphyl pipeline. In fact, a recent study evaluated the quality score of 32000 genomes located in public repositories and concluded that most of these genomes were of sufficient quality to perform analysis on and only 10% of draft genomes were of poor quality and unsuitable for downstream analysis [21]. In conclusion, our results provide strong evidence that the choice of the reference genome does not impact the ability of the cgSNV methodology to distinguish between S. Heidelberg isolates involved in foodborne outbreaks. Our results also demonstrate that using a distantly related genome as reference could lead to a loss in resolution during cgSNV analysis. Although wgMLST was recently recommended as the primary subtyping tool moving forward by PNC and other foodborne surveillance networks [22], it is important to note that the cgSNV approach still remains an important method in the PNC molecular tool box. In fact, this method was recently used by PNC in collaboration with provincial public health laboratories to identify the source of a multi-provincial outbreak of S. Enteritidis [23]. Despite the advantages provided by cg/wgMLST approaches such as curability and standardization, the development and validation of schemas for each organism still remains a daunting task both financially and technically for public health laboratories operating with limited resources. The implementation of the cgSNV methodology described here could be a

viable alternative for monitoring *S*. Heidelberg. Whether this method applies to other *Salmo-nella* serovars remains to be determined.

Acknowledgments

We express our gratitude to members of the genomics analysis and bioinformatics platforms at the National Microbiology Laboratory and the Bacteriology Unit staff at the Laboratoire de Santé Publique du Québec. Lawrence Goodridge, Céline Nadon and Sadjia Bekal are funded by Genome Canada and by the Génome Québec provincial genome center.

Author Contributions

Conceptualization: Valentine Usongo, Sadjia Bekal.

Formal analysis: Valentine Usongo, Chrystal Berry, Khadidja Yousfi.

Funding acquisition: Celine Nadon, Lawrence Goodridge, Sadjia Bekal.

Investigation: Valentine Usongo.

Methodology: Valentine Usongo, Chrystal Berry, Khadidja Yousfi, Genevieve Labbé, Roger Johnson, Eric Fournier, Celine Nadon.

Project administration: Florence Doualla-Bell, Lawrence Goodridge, Sadjia Bekal.

Resources: Lawrence Goodridge, Sadjia Bekal.

Supervision: Lawrence Goodridge, Sadjia Bekal.

Writing - original draft: Valentine Usongo, Chrystal Berry, Sadjia Bekal.

Writing – review & editing: Valentine Usongo, Chrystal Berry, Florence Doualla-Bell, Genevieve Labbé, Roger Johnson, Celine Nadon, Lawrence Goodridge, Sadjia Bekal.

References

- The National Microbiology Laboratory (NML), Centre for Food-borne, Environmental and Zoonotic Infectious Diseases (CFEZID), Public Health Agency of Canada, Provincial Public Health Microbiology Laboratories. 2014. National Enteric Surveillance Program (NESP): annual summary 2012. Public Health Agency of Canada, Ottawa, Ontario, Canada.
- Bekal S, Berry C, Reimer AR, Van Domselaar G, Beaudry G, Fournier E, et al. 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. 2016; 54:289–295. https://doi.org/10.1128/JCM.02200-15 PMID: 26582830
- Allard MW, Strain E, Melka D, Bunning K, Musser SM, Brown EW, et al. Practical Value of Food Pathogen Traceability through Building a Whole Genome Sequencing Network and Database. J Clin Microbiol. 2016; 54: 1975–1983. https://doi.org/10.1128/JCM.00081-16 PMID: 27008877
- Maiden MC, Rensburg VMJ, Bray JE, Earle SG, Ford SA, Jolley KA, et al. MLST revisited: the gene-bygene approach to bacterial genomics. Nat Rev Microbiol. 2013; 11:728–736. <u>https://doi.org/10.1038/</u> nrmicro3093 PMID: 23979428
- Moura A, Criscuolo A, Pouseele H, Maury MM, Leclercq A, Tarr C, et al. Whole genome-based population biology and epidemiological surveillance of *Listeria monocytogenes*. Nat Microbiol. 2016; 10/10 2:16185.
- Kwong JC, Mercoulia K, Tomita T, Easton M, Li HY, Bulach DM, et al. Prospective Whole-Genome Sequencing Enhances National Surveillance of *Listeria monocytogenes*. J Clin Microbiol. 2016; 54:333–342. https://doi.org/10.1128/JCM.02344-15 PMID: 26607978
- Bertels F, Silander OK, Pachkov M, Rainey PB, Nimwegen VE. Automated reconstruction of wholegenome phylogenies from short-sequence reads. Mol Biol Evol. 2014; 31:1077–1088. https://doi.org/ 10.1093/molbev/msu088 PMID: 24600054

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its application to single cell sequencing. J Comput Biol. 2012; 19:455–477. https://doi.org/10.1089/cmb.2012.0021 PMID: 22506599
- Labbé G, Ziebell K, Bekal S, Macdonald KA, Parmley EJ, Agunos A, et al. Complete genome sequence of 17 Canadian isolates of *Salmonella enterica subsp.* enterica serovar Heidelberg from human, animal and food sources. Genome Announce. 2016; 15: pii: e00990-16. <u>https://doi.org/10.1128/genomeA.</u> 00990-16 PMID: 27635008
- Petkau A, Mabon P, Sieffert C, Knox NC, Cabral J, Iskander M, et al. SNVPhyl: A Single Nucleotide Variant Phylogenomics pipeline for microbial genomic epidemiology. BioRixV. 2016;bioRixv. <u>https://doi.org/10.1101/092940</u>.
- Afgan E, Baker D, Beek VDM, Blankenberg D, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucl Acids Res. 2016; 44: W3–W10. https://doi.org/10.1093/nar/gkw343 PMID: 27137889
- Olson ND, Lund SP, Colman RE, Foster JT, Sahl JW, Schupp JM, et al. Best practices for evaluating single nucleotide variant calling methods for microbial genomics. Front Genet. 2015; 6:235. <u>https://doi.org/10.3389/fgene.2015.00235 PMID: 26217378</u>
- 13. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucl Acids Res. 2016; 44:W16–21. <u>https://doi.org/10.1093/nar/gkw387</u> PMID: 27141966
- Li H, Handsaker B, Wysoker A, Fennel T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079 https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943
- Francisco AP, Vaz C, Monteiro PT, Cristino MJ, Ramirez M, Carrico JA. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinform. 2012; 13: 87.
- Robinson DF, Foulds LR. Comparison of phylogenetic trees. Mathematical Biosciences. 1981; 53:131–147. https://doi.org/10.1016/0025-5564(81)90043-2
- Boc A, Diallo AB, Makarenkov V. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. Nucl Acids Res. 2012; 40:W573–W579. <u>https://doi.org/10.1093/nar/gks485</u> PMID: 22675075
- 18. Leekitcharoenphon P, Nielsen EM, Kaas RS, Lund O, Aarestrup FM. Evaluation of whole genome sequencing for outbreak detection of *Salmonella enterica*. PLoS One. 2014; 9:e87791.
- Olson ND, Lund SP, Colman RE, Foster JT, Sahl JW, Schupp JM, et al. Best practices for evaluating single nucleotide variant calling methods for microbial genomics. Front Genet. 2015; 6:235. <u>https://doi.org/10.3389/fgene.2015.00235</u> PMID: 26217378
- Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, Ahn TH, et al. Insights from 20 years of bacterial genome sequencing. Funct Integr Genomics. 2015; 15:141–161. <u>https://doi.org/10.1007/s10142-015-0433-4 PMID: 25722247</u>
- 21. Land M, Hyatt D, Jun SR, Kora G, Hauser L, Lukjancenko O, et al. Quality scores for 32,000 genomes. Stand Genomic Sci. 2014; 9.
- Nadon C, Van Walle I, Smidt GP, Campos J, Chinen I, Acevedo CJ, et al. PulseNet International: Vision for the implementation of whole genome sequencing (wgs) for global food-borne disease surveillance. Euro Surveill. 2017; 22:30544. <u>https://doi.org/10.2807/1560-7917.ES.2017.22.23.30544</u> PMID: 28662764
- Croxen MA, Macdonald KA, Walker M, deWith N, Zabek E, Peterson C, et al. Multi-provincial Salmonellosis outbreak related to newly hatched chicks and poults. A genomics perspective. PLoSCurr.2017; 1. PMID: 29034124