View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Crossref

Webb et al. BMC Microbiology (2015) 15:94
DOI 10.1186/s12866-015-0426-4

BMC Microbiology

METHODOLOGY ARTICLE Open Access

Development of a comparative genomic fingerprinting assay for rapid and high resolution genotyping of *Arcobacter butzleri*

Andrew L Webb[1,2], Peter Kruczkiewicz[3], L Brent Selinger[2], G Douglas Inglis[1*] and Eduardo N Taboada[3*]

Abstract

Background: Molecular typing methods are critical for epidemiological investigations, facilitating disease outbreak detection and source identification. Study of the epidemiology of the emerging human pathogen *Arcobacter butzleri* is currently hampered by the lack of a subtyping method that is easily deployable in the context of routine epidemiological surveillance. In this study we describe a comparative genomic fingerprinting (CGF) method for high-resolution and high-throughput subtyping of *A. butzleri*. Comparative analysis of the genome sequences of eleven *A. butzleri* strains, including eight strains newly sequenced as part of this project, was employed to identify accessory genes suitable for generating unique genetic fingerprints for high-resolution subtyping based on gene presence or absence within a strain.

Results: A set of eighty-three accessory genes was used to examine the population structure of a dataset comprised of isolates from various sources, including human and non-human animals, sewage, and river water (n=156). A streamlined assay (CGF_{40}) based on a subset of 40 genes was subsequently developed through marker optimization. High levels of profile diversity (121 distinct profiles) were observed among the 156 isolates in the dataset, and a high Simpson's Index of Diversity (ID) observed (ID > 0.969) indicate that the CGF_{40} assay possesses high discriminatory power. At the same time, our observation that 115 isolates in this dataset could be assigned to 29 clades with a profile similarity of 90% or greater indicates that the method can be used to identify clades comprised of genetically similar isolates.

Conclusions: The CGF_{40} assay described herein combines high resolution and repeatability with high throughput for the rapid characterization of *A. butzleri* strains. This assay will facilitate the study of the population structure and epidemiology of *A. butzleri*.

Keywords: Molecular epidemiology, Subtyping, Comparative genomics, *Arcobacter butzleri*, Genome sequencing

Background

Arcobacter butzleri is closely related to the pathogen *Campylobacter jejuni* [1], and it has been isolated from surface waters, livestock, and animal products [2-5]. The pathogenicity of *A. butzleri* has yet to be resolved [2,6]; although *A. butzleri* has been isolated from the stools of diarrheic human beings, which is highly suggestive of pathogenicity [7-9], it has also been obtained from non-diarrheic individuals, [10,11] suggesting that it is a commensal or that non-pathogenic strains or subtypes exist within the species.

An important facet in the study of pathogens is epidemiology-based analysis of their incidence and distribution. Molecular subtyping or genotyping, which allows the classification of a bacterial species into distinct strains or subtypes based on genetic variation [12,13], forms one of the pillars of molecular epidemiology, through which the identification of etiological agents, patterns of transmission, and potential outbreaks can be carried out with enhanced precision [14]. Until recently, the study of *A. butzleri* has been hampered by the lack of advanced methods for subtyping. A recently developed multi-locus sequence typing (MLST) scheme [15] provides excellent identification of subtypes and has

* Correspondence:
douglas.inglis@agr.gc.ca; eduardo.taboada@phac-aspc.gc.ca
[1]Agriculture and Agri-Food Canada, 5403 – 1st Avenue S, Lethbridge, AB, Canada
[3]Public Health Agency of Canada, Township Rd. 9-1, Lethbridge, AB, Canada
Full list of author information is available at the end of the article

BioMed Central

© 2015 Webb et al.; licensee BioMed Central. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

been utilized to examine genetic diversity in *A. butzleri* isolated from people, livestock, and animal products [2,6]. However, this method remains a resource-intensive and relatively low-throughput means of subtyping, which limits the number of isolates that can be analyzed by most research groups [16,17], as evidenced by the relatively small number of isolates that have been contributed to the MLST database for *A. butzleri* by the global research community (*n*=683, PubMLST accessed on October 21, 2014). More importantly, the lack of a highly deployable subtyping method suitable for use in routine surveillance has precluded the large-scale epidemiological surveys required to fully assess the potential role of *A. butzleri* as an emerging pathogen of humans.

Recent advances in sequencing technologies (i.e. next generation sequencing) and bioinformatics have made it possible to rapidly obtain draft whole genome sequence (WGS) data [18] and it is likely that methods based on WGS analysis, including whole-genome MLST (wgMLST), will eventually become the new standard for microbial subtyping in an epidemiological context [19,20]. However, until the resources required for WGS-based subtyping allow it to become practical enough to be deployed in large-scale epidemiological surveillance, there is a continuing need for methods that fulfill performance criteria such as discriminatory power and repeatability, and convenience criteria such as throughput, cost, and ease of use [14]. Recently, Taboada *et al.* [21] employed whole genome analysis to develop a comparative genomic fingerprinting (CGF) method for high-resolution subtyping of *C. jejuni* that was highly concordant with MLST but better suited to large-scale surveillance due to improved throughput and cost relative to MLST. Moreover, by targeting a large number of accessory genes (e.g. 40 loci), the CGF method showed improved discriminatory power compared to MLST, allowing the differentiation of closely related strains with distinct epidemiology [21,22].

The overall goal of the current study was to develop a highly discriminatory CGF assay for *A. butzleri* by employing the strategy described by Taboada *et al.* [21] for *C. jejuni*. Objectives were to: (i) select *A. butzleri* isolates for whole genome sequencing; (ii) utilize whole genome sequence data to identify candidate CGF target genes in the accessory genome; (iii) screen CGF targets against a panel of *A. butzleri* isolates to determine accessory gene frequency and assess accessory genome variability; (iv) select a subset of CGF targets for development of a 40-locus assay (CGF_{40}); and (v) evaluate the ability of the CGF_{40} assay to reliably discriminate *A. butzleri* strains. The development of highly deployable genotyping techniques that are suitable for use in routine surveillance will improve our ability to distinguish strains of *A. butzleri* and facilitate the study of its epidemiology.

Results

Whole genome sequencing and comparative genomic analysis of *A. butzleri* strains

In order to design a CGF assay for *A. butzleri* it was necessary to perform a comparative genomic analysis of strains representing diverse sources and genetic backgrounds. Twenty-two *A. butzleri* isolates from various sources were genotyped using amplified fragment length polymorphism (AFLP) analysis [23,24] and eight strains representing highly diverse AFLP profiles were chosen for whole-genome sequencing (Additional file 1). The selected strains represented six of thirteen different clades observed in the set of twenty-two isolates analyzed by AFLP.

Illumina 100 bp read paired-end sequencing of *A. butzleri* isolates (*n*=8) produced an average of 132 ± 37.0 times coverage based on an assembly size of 2.27 Mbp ± 0.09, with a GC content of 27.3% ± 0.90 and 2.10 ± 1.70 ambiguous bases per 100 kbp. The *de novo* assemblies contained 444 ± 146 contigs and $2.28 \times 10^3 \pm 129$ predicted ORFs. In total, 2.47×10^4 coding sequences were identified from the assembled contigs, and 1.42×10^3 core and 1.63×10^3 unique accessory genes were identified by comparative genomic analysis of the eleven strains included in this study. After removing genes with biased population distribution, those with redundant patterns of presence and absence, or those presenting problems for subsequent polymerase chain reaction (PCR) primer design, a set of eighty-three candidate accessory genes was identified and used to design an expanded CGF assay aimed at examining the population structure of a large set of *A. butzleri* isolates (n=156) based on shared accessory genome content. Data from eleven accessory genes was discarded due to discordance between *in silico*-predicted CGF profiles and laboratory results on eight isolates sequenced *de novo* as part of this project. The reference CGF-based phylogeny was established from the remaining seventy-two accessory genes.

A 'reference phylogeny' for a sample population of *A. butzleri* isolates

A reference phylogeny for a comprehensive set of *A. butzleri* isolates (n=156) recovered from river water, raw and treated sewage, diarrheic and non-diarrheic people, and non-human animals was derived from the binary (i.e. presence and absence) data for the expanded CGF assay. The phylogenetic distribution of twelve genome-sequenced strains, which includes four previously sequenced strains and eight strains sequenced as part of this study, shows that all but two strains (149 and 151) belong to distinct CGF clades. Moreover, the *in silico* MLST data (Additional file 1) is consistent with the CGF results because strains 149 and 151 share the same, albeit novel, sequence type while the remaining strains

are from diverse sequence types. An average of ten distinct alleles were observed at each of the seven MLST loci, and the lack of shared alleles suggests significant genetic diversity among the twelve WGS strains. Although this dataset does not represent a comprehensive sampling of the *A. butzleri* population, a comparative genomic analysis of these isolates would be expected to capture significant accessory genome diversity. The reference phylogeny contained a total of 31 multi-isolate clades when a ≥90% isolate similarity threshold was applied (Figure 1). The largest clade (Clade 5) comprised 12 isolates from four human diarrheic stool samples. Clade 31 contained all of the isolates recovered from two non-diarrheic human stools. Isolates from non-human animals clustered together and distinctly from other isolates. Although human isolates clustered with water isolates (clades 2 and 31, respectively), there were no clades that contained isolates from both diarrheic and non-diarrheic human beings. None of the four previously sequenced strains included in this dataset clustered at the 90% similarity level with the *A. butzleri* isolates from Southwestern Alberta.

Analysis of CGF_{40} concordance with reference phylogeny

After 1.0×10^4 iterations, CGF Optimizer [25] retrieved 40 accessory genes for CGF_{40} that had an Adjusted Wallace Coefficient (AWC) of 1.0 with respect to the reference phylogeny. Analysis of the 156 *A. butzleri* isolates yielded high Simpson's ID (Table 1) and AWC (Table 2) values for both assays at 90% and 95% similarity thresholds. In addition, direct comparison showed that clusters in the reference and CGF_{40} phylogenies were highly concordant (Figure 2). At 90% similarity, isolates from 29 of the 31 clades identified in the reference phylogeny also clustered together when analysed using the CGF_{40} assay. Moreover, of the 54 isolates that shared identical CGF_{40} profiles, 45 also shared identical profiles when analysed with the expanded set of 72 markers.

Analysis of CGF_{40} reproducibility

To assess assay reproducibility, the CGF_{40} analysis was repeated for 24 *A. butzleri* isolates on separate occasions. Concordance analysis revealed that 907 of the 920 data points (98.6%) had identical presence/absence patterns in both runs.

Discussion

Enteritis is inflammation of the alimentary canal (i.e. enteron) that is often characterized by diarrhea, abdominal pain, dehydration, loss of appetite, fever, and nausea [26]. We chose to target Southwestern Alberta as a study site (i.e. a model agro-ecosystem) as this region possesses high rates of enteritis [27], which has been attributed to dense livestock populations in the region [28,29]. *Arcobacter butzleri* is closely related to *C. jejuni* and it is considered by some to be an emerging pathogen [3,7,30] because it has been isolated from diarrheic people [11,31]. However, its pathogenicity and reservoirs/pathways of transmission for potentially pathogenic genotypes have yet to be elucidated. In order to understand the relationship between *A. butzleri* and human illness a method is required for the rapid and accurate genotyping of *A. butzleri* strains to facilitate epidemiological studies.

A number of subtyping methods have recently been used to examine genetic diversity of *Arcobacter* and to compare genotypes between sources [6,30,32]. Douidah *et al*. have recently proposed a two-stage approach using enterobacterial repetitive intergenic consensus PCR and pulsed-field gel electrophoresis for subtyping of human and animal *Arcobacter* isolates [32]. A scheme for MLST, a leading method for related organisms such as *C. jejuni* and *H. pylori*, has recently been developed for *A. butzleri* [6]. No *A. butzleri* sequence types have been directly linked to human illness, but given the relative paucity of data both in the literature and within the global MLST database [33] it is difficult to assess whether the *A. butzleri* MLST data generated so far is representative of large-scale population or epidemiological trends. Moreover, despite the demonstrated ability of MLST to accurately distinguish subtypes of *A. butzleri* and other bacteria, the resources required to generate MLST data for the substantial numbers of *A. butzleri* isolates that are necessary for comparative epidemiological investigations may be prohibitive for many research groups.

Comparative genomic fingerprinting provides a high-resolution and high-throughput alternative to MLST that is also deployable in the context of large-scale epidemiological surveillance [21,22]. The CGF method identifies intraspecies relationships by targeting accessory loci that are representative of genetic variation throughout the genome. The phylogenetic signal in accessory genome content variation has been examined in several bacterial species and was shown to be highly concordant with that contained in other forms of genetic variation ([34-36]). Such loci are binary (i.e. present or absent) and determination of their allelic status does not require sequencing, with assessment possible by PCR amplification. In addition, CGF assays target sufficient loci to distinguish between closely related strains that may be indistinguishable by other methods [21] while generating phylogenetic signal that is consistent with that of MLST [37]. Previous work has shown the CGF assay for *C. jejuni* to be highly predictive of MLST, and although each method clustered strains similarly, CGF provided additional discrimination within those groups [21,22]. In Canada, the CGF method is being used to analyze *C. jejuni* isolates generated through several large-scale surveillance networks, which will facilitate

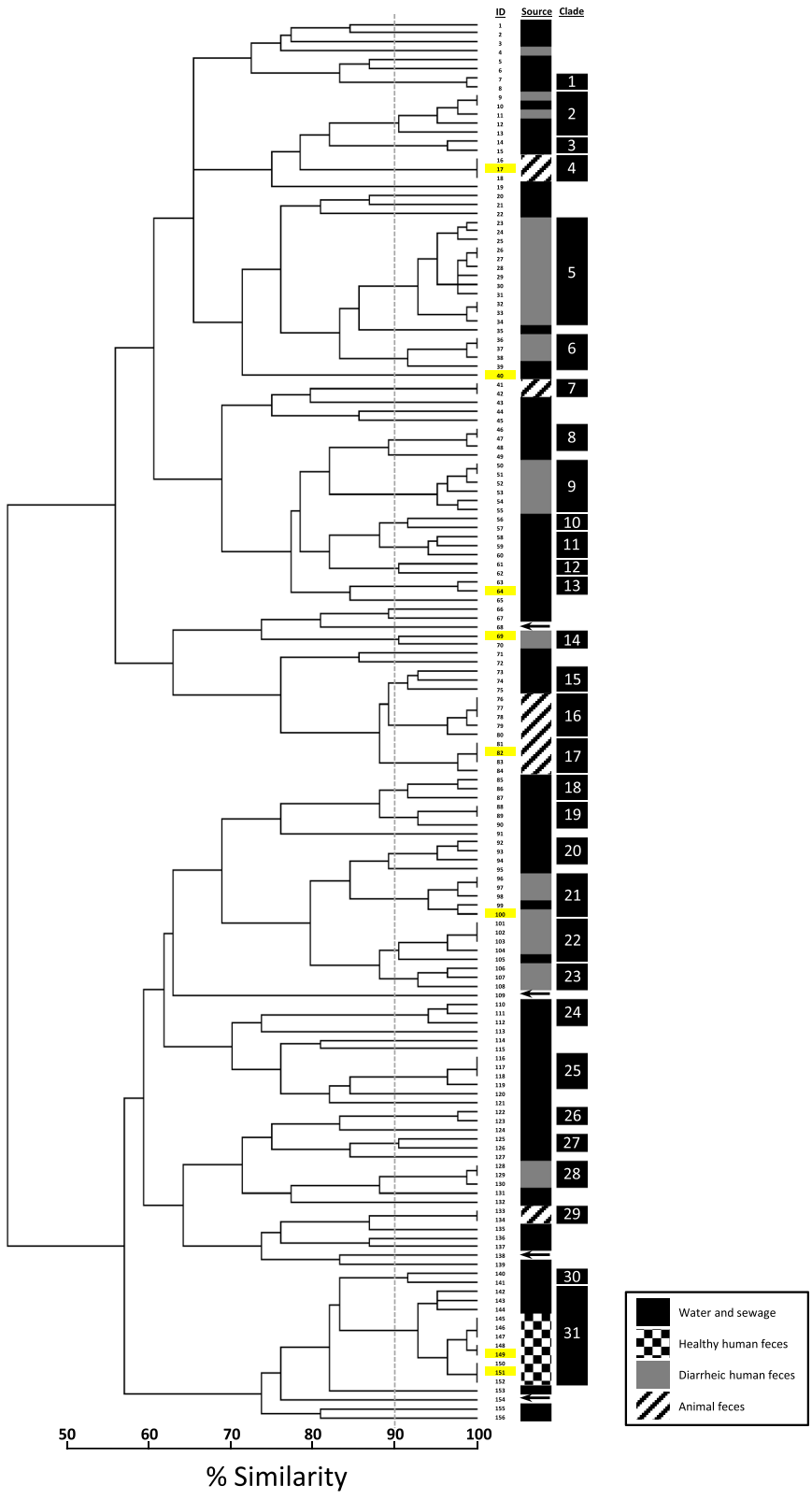

Figure 1 (See legend on next page.)

(See figure on previous page.)
Figure 1 Reference genealogy of *A. butzleri* isolates (*n*=156). Clusters were calculated by simple matching comparison of 72 accessory genes using pairwise coefficients and UPGMA analysis. The scale represents fingerprint similarity based on the total number of shared loci between isolate profiles and the total number of loci in the assay. Dashed grey line represents a 90% similarity threshold used for clade definition. Isolates sequenced as part of this study are highlighted in yellow; ID 17 (strain L353, PRJNA233527), ID 40 (strain L355, PRJNA233527), ID 64 (strain L348, PRJNA233527), ID 69 (strain L352, PRJNA233527), ID 82 (strain L354, PRJNA233527), ID 100 (strain L349, PRJNA233527), ID 149 (strain L351, PRJNA233527), ID 151 (strain L350, PRJNA233527). Published reference *A. butzleri* strains are designated with arrows and include ID 68 (strain 7h1h, PRJNA200766), ID 109 (strain JV22, PRJNA61483), ID 138 (strain RM4018, PRJNA58557), ID 154 (strain ED-1, PRJNA158699).

the study of campylobacteriosis through the holistic comparison of *C. jejuni* subtypes collected from a diverse range of sources and infection cases [38].

In addition to being a close phylogenetic relative of *C. jejuni*, two features of the *A. butzleri* pan-genome identified through our comparative genomic analysis suggested that it would be an excellent species for the development of a CGF-based genotyping assay. The *A. butzleri* strains showed significant variability in accessory genome content, which allows for with a high level of discriminatory power when comparing gene carriage. Moreover, we observed that the *A. butzleri* pan-genome contained a significant number of accessory genes, which allowed us to focus on those of greatest genotyping potential. The CGF_{40} assay is based on a marker optimization process that yielded phylogenetic clusters that were highly concordant with those observed in the reference phylogeny, and it provided a high discriminatory power for differentiation of isolates from diverse sources. In addition, the majority of isolates that were identical by CGF_{40} analysis also proved to be identical or highly similar using the larger number of markers. This suggests that the 40 loci that we selected during marker optimization were appropriate for high resolution genotyping of *A. butzleri* strains, and that there may be an "efficiency plateau" above which additional loci do not sufficiently increase discriminatory power to justify their inclusion in the assay. The CGF_{40} assay should be easily deployable, and we found that 32 isolates could be processed (i.e. from stock to digital phylogeny) during a typical workday by a single individual using one thermal cycler and capillary electrophoresis system; more than 1000 *A. butzleri* isolates have been genotyped by one individual in our research group over the course of a 12-month period.

Previous efforts to characterize *A. butzleri* have identified a high degree of genetic variation but have failed to associate specific genotypes in a geographic or temporal context [6,39-41]. In total, we observed 29 clades within the CGF_{40} phylogeny when compared at a similarity threshold of 90% or greater and 121 distinct (i.e. non-identical) CGF_{40} profiles were observed among the 156 isolates analyzed. Of interest, each of the four previously genome-sequenced strains in the public databases formed their own clades in both CGF-based phylogenies. Taken together, these results suggest that the density of marker sampling targeted by the CGF assay described herein provides sufficient power for discriminating isolates at a high level of resolution. At the same time, our observation that 115 of the 156 isolates in this dataset could be assigned to clades with a profile similarity of 90% or greater suggests that this level of discriminatory power does not compromise the ability to identify clades comprised of genetically similar isolates. It is noteworthy that although we developed the CGF_{40} assay with isolates primarily obtained from Southwestern Alberta, the dataset used for the comparative genomic analysis to identify potential CGF markers also included several genome-sequenced isolates from international sources. Moreover, it is our intention to further validate the CGF_{40} assay by examining *A. butzleri* populations in a pan-Canadian and an international context.

Arcobacter butzleri were isolated from the stools of diarrheic and non-diarrheic human beings living in Southwestern Alberta, as well as from river and sewage samples throughout Southwestern Alberta during 2008 and 2009. The clustering of isolates from human beings with isolates from river and sewage waters throughout Southwestern Alberta over the same time period suggests that it may be possible for *A. butzleri* strains to be transferred between people and their environment. Although we were unable to identify clades that included isolates from human beings and non-human animals in

Table 1 Simpson's Index of Diversity[a] for *A. butzleri* isolates (*n*=152) genotyped by CGF_{40}

Partitioning method	Assay	Partitions[b]	Simpson's ID	CI (95%)	CINA (95%)
Binary Pairwise Similarity (UPGMA)	Reference	87	0.984	0.978-0.991	0.977-0.992
	CGF_{40}	86	0.987	0.983-0.992	0.982-0.992

[a]Simpson's Index of Diversity (ID), confidence intervals (CI), and non-approximated confidence intervals (CINA) were calculated using the online tool of the Comparing Partitions Website (http://darwin.phyloviz.net/ComparingPartitions/index.php?link=Tool).
[b]Partitions were denoted at the 95% similarity level, which was calculated using the simple matching coefficient in BioNumerics (version 6.6, Applied Maths, Austin, TX).

Table 2 Adjusted Wallace Coefficient values[a] of CGF_{40} compared to the reference phylogeny for *A. butzleri* isolates (*n*=152)

Partitions[b]	Reference (90% Similarity)	Reference (95% Similarity)
CGF_{40}	0.88	0.62
(90% Similarity)	(0.83-0.93)	(0.53-0.71)
CGF_{40}	0.92	0.87
(95% Similarity)	(0.89-0.95)	(0.83-0.91)

[a]Adjusted Wallace Coefficient values were calculated using the online tool of the Comparing Partitions Website (http://darwin.phyloviz.net/ComparingPartitions/index.php?link=Tool).
[b]Partitions were denoted by 90% and 95% accessory gene pairwise similarity, which were calculated using the binary simple matching algorithm in BioNumerics (version 6.6, Applied Maths).

this dataset, this may be due to the non-concurring sampling of both reservoirs since non-human animal stools were sampled only towards the end of 2009. We hope to examine the possible links between *A. butzleri* found in human beings and animals through concurrent and comprehensive sampling in future studies; the rapid and inexpensive characterization of isolates using our CGF_{40} method for will be very useful in this regard.

To our knowledge no studies conducted to date have examined the carriage and shedding of *A. butzleri* strains in diarrheic and non-diarrheic human beings, and although *Arcobacter* species have been detected in and occasionally isolated from the stools of non-diarrheic individuals [10,11], this is the first time that *A. butzleri* have been isolated from stools of a non-diarrheic human being sampled on two separate occasions. Individuals were sampled six months apart and periodic shedding of the same *A. butzleri* genotype suggests that strains of this bacterium may chronically colonize people without inciting disease. Colonization of healthy human beings by *A. butzleri* may occur in a similar manner to the closely related pathogen *C. jejuni*, which has been shown to colonize healthy people more frequently in areas with endemic rates of infection [23,42,43]. Thus, it may be possible to relate genotypes to endemic disease rates by characterizing *A. butzleri* isolates from diarrheic and non-diarrheic human beings.

Conclusions

We used WGS and comparative genomic analysis of *A. butzleri* isolated from diverse sources and demonstrated that accessory gene variation among strains can be used for high-throughput, high-resolution, and reproducible subtyping of this bacterium. Although WGS analysis will eventually become the gold standard in epidemiological genotyping of pathogenic bacteria, until WGS data are routinely deployed for surveillance of highly prevalent pathogens, the CGF_{40} assay described herein will allow the scientific community to address key knowledge gaps about the epidemiology of arcobacteriosis toward the prevention and mitigation of enteric disease. Furthermore, the developed CGF_{40} assay is highly deployable and will allow researchers and clinicians to efficiently compare the genetic diversity, persistence, and prevalence of *A. butzleri* subtypes in different sources, and to rapidly and efficiently identify relevant strains as candidates for WGS analysis.

Methods

Ethics statement

Scientific and ethics approval to isolate *A. butzleri* from diarrheic and non-diarrheic human beings (i.e. healthy volunteers) was obtained by GDI from the Regional Ethics Committee of the former Chinook Health Region (CHR) and from the University of Lethbridge Human Subject Research Committee. The requirement for informed written consent was waived by the CHR Regional Ethics Committee and the University of Lethbridge Human Subject Research Committee for subsamples of stools submitted by diarrheic people as the samples were submitted for the detection of enteric pathogens at the Chinook Regional Hospital and the identities of patients was not disclosed. Informed written consent as mandated by the University of Lethbridge Human Subject Research Committee was obtained from all healthy volunteers in advance of the submission of stool samples for the isolation of *A. butzleri* and other enteric bacteria.

A. butzleri isolation and DNA extraction

Arcobacter butzleri were isolated from one stool sample per person for eleven diarrheic humans, and from two stools obtained from one non-diarrheic human, as well as from non-human animal feces, sewage, and river water collected in Southwestern Alberta during 2008 and 2009. Isolates were streaked for purity and stored at -80°C in Columbia broth (Difco and BBL Microbiology, Lawrence, KS) with 30% glycerol. Isolates from glycerol stocks were grown on Columbia agar (Difco and BBL Microbiology) amended with 10% sheep blood in a microaerobic atmosphere (5% O_2, 3% H_2, 10% CO_2, and 82% N_2) at 37°C for 24-48 hr, and biomass was collected from the surface of the agar medium. An automated system (Model 740, Autogen, Holliston, MA) was used to extract genomic DNA. Putative *A. butzleri* isolates were identified by PCR amplification using an *Arcobacter* PCR-multiplex assay [9].

Whole genome sequencing and assembly

For WGS analysis, DNA was extracted using a DNEasy Blood and Tissue Kit (Qiagen Inc, Toronto, ON). To minimize possible genetic bias amongst strains selected

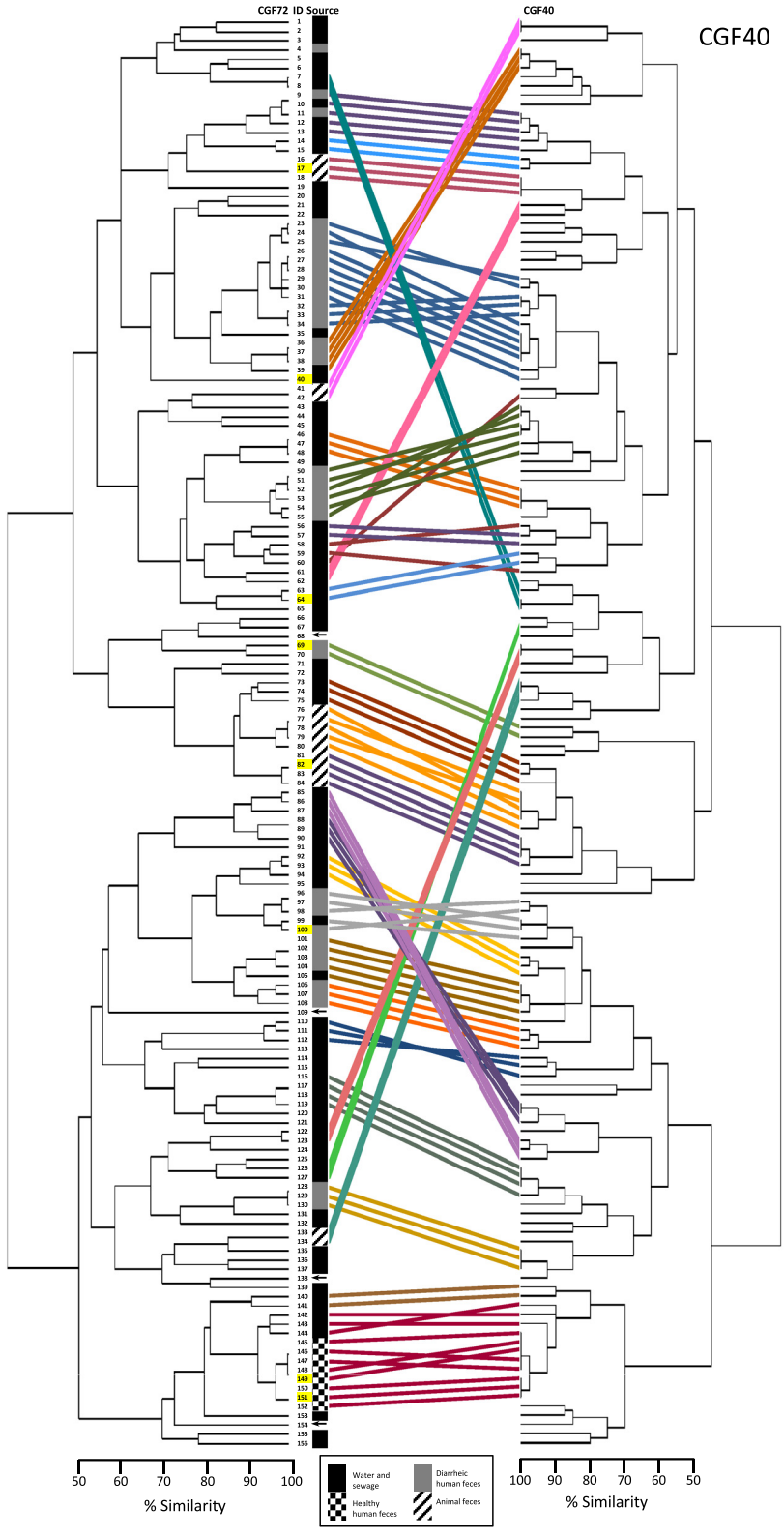

Figure 2 (See legend on next page.)

(See figure on previous page.)
Figure 2 Tanglegram of CGF_{72} (reference) and CGF_{40} genealogies for *A. butzleri* isolates ($n = 156$). Coloured lines represent isolates within clusters in the reference phylogeny that are ≥90% similar to one or more other isolates. Scales represent fingerprint similarity based on the total number of shared loci between isolate profiles and the total number of loci in the assay. Coloured lines also indicate the location of the same isolate in the CGF_{72} and the CGF_{40} cladograms. Scales represent fingerprint similarity based on the total number of shared loci between isolate profiles and the total number of loci in the assay. Isolates sequenced as part of this study are highlighted in yellow; ID 17 (strain L353, PRJNA233527), ID 40 (strain L355, PRJNA233527), ID 64 (strain L348, PRJNA233527), ID 69 (strain L352, PRJNA233527), ID 82 (strain L354, PRJNA233527), ID 100 (strain L349, PRJNA233527), ID 149 (strain L351, PRJNA233527), ID 151 (strain L350, PRJNA233527). Published reference A. butzleri strains are designated with arrows and include ID 68 (strain 7h1h, PRJNA200766), ID 109 (strain JV22, PRJNA61483), ID 138 (strain RM4018, PRJNA58557), ID 154 (strain ED-1, PRJNA158699).

for WGS, *A. butzleri* isolates from diverse sources were genotyped using Amplified Fragment Length Polymorphism (AFLP) analysis as described previously [24,44], and eight strains selected to represent highly diverse AFLP profiles were chosen for sequencing (Table 3). The identity of isolate DNA was tested by sequencing approximately 1000 bp of the 16S rRNA gene and by comparing the results with *A. butzleri* sequences within the National Centre for Biotechnology Information (NCBI) genetic database [45,46]. The DNA for isolates to be sequenced was quantified by spectrophotometry (A_{600}) (Ultrospec 3100 pro, GE Healthcare Life Sciences, Baie d'Urfe, QC). Isolates were sequenced as paired-end, 100 bp reads on a HiSeq platform (Illumina Inc., San Diego, CA) with Phred30 (99.9%) base-calling accuracy [47], and reads were *de novo* assembled into contigs using ABySS [48] with specifications for short paired-end reads. Sequencing data for the *A. butzleri* isolates were accessioned in the NCBI genetic sequence database as a single bioproject (PRJNA233527).

Detection and identification of coding sequences

Rapid Annotation Using Subsystem Technology [49] was used to identify open reading frames (ORF) for the eight sequenced *A. butzleri* genomes, as well as three previously available genome assemblies (RM4018 - PRJNA58557, ED1 - PRJNA158699, JV22 - PRJNA61483). The genome assembly for a fourth strain, 7h1h (PRJNA200766), was not available at the time that the comparative genomic analysis was performed, however we were able to utilize the four published WGS strains for all subsequent *in silico* CGF analyses.

To identify core and accessory genes, the ORFs from each genome were searched against the eleven genome assemblies using the program BLASTP from the Basic Local Alignment Search Tool [45,46], with filtering to remove redundant results from likely orthologous genes. ORFs present in all assemblies were identified as core, and all non-redundant ORFs absent from one or more strains were designated as accessory.

Identification of candidate accessory genes for CGF assay development

To simplify CGF assay design, accessory genes with limited genotypic potential due to a highly biased population distribution (i.e. present in greater than 80% of strains or present in fewer than 20% of strains) were eliminated from further consideration as candidate markers. Moreover, for groups of accessory genes that presented redundant patterns of presence and absence in the dataset (i.e. genes that are typically linked and provide limited additional discrimination), only one representative gene from each unique pattern was considered as a candidate marker for CGF development. Short genes (i.e. <300 bp) and/or those containing nucleotide gaps or polymorphisms that might affect PCR primer design were

Table 3 Isolates of *A. butzleri* from diverse sources for whole genome sequence analysis

Strain	Source	Location	Collection date
L348	Sewage outfall	Lethbridge, Alberta, CA	07/May/2008
L349	Diarrheic human stool	Blairmore, Alberta, CA	30/Jul/2008
L350	Non-diarrheic human stool	Lethbridge, Alberta, CA	30/Sep/2008
L351	Non-diarrheic human stool	Lethbridge, Alberta, CA	01/Apr/2009
L352	Diarrheic human stool	Lethbridge, Alberta, CA	27/Apr/2009
L353	Horse feces	Diamond City, Alberta, CA	02/Jul/2009
L354	Pig feces	Lethbridge, Alberta, CA	12/Aug/2009
L355	Raw sewage	Lethbridge, Alberta, CA	08/Mar/2009

also discarded. Accessory genes meeting the above criteria were identified and used to design an expanded CGF assay (i.e. the reference assay) to examine the population structure of a diverse collection of *A. butzleri* isolates (n=152) based on accessory genome variability. Data from these isolates, which were recovered from river water, raw and treated sewage, diarrheic and non-diarrheic human beings, and non-human animals in Southwestern Alberta was used in conjunction with *in silico*-derived [50] CGF data from four published genome-sequenced strains (RM4018 - PRJNA58557, ED1 - PRJNA158699, JV22 - PRJNA61483, 7h1h - PRJNA200766). CGF profiles were also generated *in silico* using the program MIST [50] for the eight isolates sequenced *de novo* to allow for comparison with PCR-derived CGF data, thus facilitating assessment of marker performance. A dendrogram representing an estimate for a 'reference phylogeny' was constructed from the binary (i.e. presence and absence) data for those genes that generated data fully concordant with *in silico*-predicted CGF profiles (n=72). Hierarchical clustering was performed by the unweighted pair group method with arithmetic mean (UPGMA) using the *hclust* function in R [51] and the simple matching coefficient of genetic similarity.

Optimization of markers for development of final CGF assay

The program CGF Optimizer [25], which calculates the AWC and the Robinson-Foulds Symmetric Distance (SymD) [52-55] to assess the concordance between clustering results from sets of prospective CGF markers and a reference phylogeny, was used to identify a subset of accessory genes yielding high concordance to the reference phylogeny generated using the expanded CGF assay. Briefly, CGF Optimizer was used to subsample sets of candidate accessory genes and to compute the AWC of each set to the reference phylogeny; the 40 loci that were most concordant with the reference phylogeny (i.e. the set with the highest AWC) were selected for the final CGF_{40} assay.

CGF assay development

Primer3 [56] was employed to design PCR primers for genes selected for CGF assays (Additional file 2). The programs MultiPLX [57] and CGF Multiplexer [25] were used to arrange primers with compatible thermodynamic properties into multiplex pools that would generate amplicons differing by at least 100 bp to facilitate unambiguous scoring of marker presence or absence. The CGF profiles obtained *in silico* [50] and by multiplex PCR amplification for the sequenced strains were compared to ascertain primer sensitivity and specificity, and primer pair concentrations within each multiplex were adjusted to optimize product amplification (Table 4). In addition, the reproducibility of the final CGF_{40} assay was tested by running duplicate PCR reactions for a set of 24 *A. butzleri* isolates (23 test isolates plus 1 control). To generate a CGF profile, eight PCR reactions targeting five loci per reaction were performed for each *A. butzleri* isolate. Individual PCR reactions (25 μl) contained 2.0 μl of genomic DNA, 2.5 μl of 10X incubation mix without $MgCl_2$ (MP Biomedicals, Solon, OH; 1X), 2.5 μl of $MgCl_2$ (MP Biomedicals; 2.5 mM), 0.5 μl of a deoxynucleoside triphosphate pool (0.2 mM), 1.0 μl of the multiplex primer pool (0.4 μM), 0.2 μl Taq DNA Polymerase (MP Biomedicals; 1 U μl^{-1}), and 16.3 μl Optima water (Fisher Scientific, Ottawa, ON). PCR conditions consisted of 32 cycles of denaturation at 93°C for 30 s, annealing at 60°C for 90 s, and extension at 72°C for 60 s. After a final extension step at 72°C for 5 min, PCR products were stored at 4°C, and visualized using a QIAxcel automated capillary electrophoresis system (Qiagen Inc.) with a QIAxcel 2400 Sample DNA Screening Kit (Qiagen Inc.), QX 15-1000 bp alignment marker (Qiagen Inc.), and 30 ng μL^{-1} QX 50-800 bp Size Marker (Qiagen Inc.). Capillary electrophoresis lanes were scored for amplification of the five loci targeted (i.e. scored as present or absent) in each multiplex PCR, resulting in a 40-digit binary profile for each isolate. Isolate profiles were clustered using the simple matching coefficient in BioNumerics (version 6.6, Applied Maths, Austin, TX), and isolate similarity was visualized as an UPGMA dendrogram.

Assessment of CGF discrimination and concordance

PCR data for the reference and CGF_{40} assays was generated for the 152 *A. butzleri* isolates. The CGF profiles of four previously published genome-sequenced strains (RM4018, ED1, JV22, and 7h1h) were also obtained *in silico* [50]. To verify concordance between the expanded CGF and CGF_{40} assays, binary data from each assay was subjected to hierarchical clustering by UPGMA using the *hclust* function in R [51] and the simple matching coefficient of genetic similarity. The online 'Comparing Partitions' tool [52] was used to calculate the discriminatory power of each assay and the concordance between assays. The discriminatory power of each CGF assay was calculated using Simpson's ID [58], and the concordance was calculated as the AWC value between the CGF_{40} assay and the reference phylogeny. A "tanglegram" was generated using a custom R script to compare dendrograms for the CGF_{40} and the reference phylogeny. This script is available online at https://gist.github.com/peterk87/d92f81ae475063792f49. Briefly, the script generates the dendrograms from binary CGF_{40} and reference phylogeny data and rearranges the CGF_{40} dendrogram with respect to the reference phylogeny in order to maximize structural concordance or minimize entanglement of branches using the "untangle_step_rotate_1side"

Table 4 Primers for PCR amplification of CGF_{40} markers[a]

	Product size (bp)	Primer forward (5′ to 3′)	Primer reverse (5′ to 3′)	Concentration[b] (μM)
Multiplex 1	150	GCATCCTCTTCCTCCATCAT	TCGAATAAATCCCCTACCCTT	12
	250	ATACACCACCAGATGAGCTG	TAACGTACCGCATCCATTGA	10
	400	AGTGCCCGTTCTATTGGTAT	GCATAAAGAGCTTCTCCTCC	8
	500	ACTCTTCCCGAATCTGCAAT	TCTCCAATTCCTTGTCCTATTGT	10
	600	AGTCATGCAATCCTAACGAGA	AGGAGCCTACTATGTACCTCT	10
Multiplex 2	150	TTTTCATTGGGAAGAAGAATTTAGT	TCCAATTCATAAATATCTCTTGGTGA	12
	250	TCTTTTAAAGAAGACAGCTGTAGT	TTTTGCAACACCTAATCTTGC	18
	350	TGATACAGGAATTATAAGAAGTGTTCC	GCATGAACTTCAACTCCAGG	5
	450	TGGAAATGACAGAGGATGGT	AGTAACGGATGAGCTTTTAAATTT	8
	600	TTGGGCTATTATGTCCCCAG	TCGTACAACTGGCATAGCTT	7
Multiplex 3	200	CCTCAACTTCTAACAGCAGG	CTCACATCACCCAATCCACT	8
	300	TGGAATATCATAAACCAAAAATTGTTT	TTCATTGCAAATCCGCCTTT	10
	450	ACAGCATCCTTGATTCTAGCA	GTGTAATCATAGCCCAAATCCA	12
	550	TGAAATAATGAATGAACACAATAGCA	GTGCACAACCTAAAACCTCA	10
	700	GACAGGAACAGAGGGAAGTC	AGCATCTTTATTTGTCGCACT	10
Multiplex 4	200	TGATGAAACACTAGAAAATAAGGCT	CCAGTAAAACCTCTGTCAGC	11
	350	TCACTTTTAGGTACTCACGACT	GCTATAAAACTTGCACCTTTATCG	9
	450	CAAAGATTTCTACGGGAAATTTGT	ACATCCTTTGCCTCTTTAAAAGA	9
	550	TCGAGGACAAGCAGATTCAA	GCCATTTCTACTTCCATTGTGT	7
	700	ACAGCAGTAACATTACAGGG	TCAAAAGCAATTCCACCACT	11
Multiplex 5	150	TCTATAGGTGCTGACCCACT	GCCGCAATACTTCCAAAACT	9
	250	TTTACAGGAGCTTGGACATCA	TTTTACCATCATCTTCAACCCA	9
	400	CATCGTCCTTCAGTCGAATAT	GGAAACCATTTTCTTTTGCCA	9
	550	GTCATTTTTACACCACCTGCA	TCAAAACGCTTAGCCAAATCT	12
	700	ACTTTTTGCTTCTCAAAGTAGAAC	CCTCTGAAAAATTGAAATAATATACCC	10
Multiplex 6	150	GGTTGGGGAAAACTGCTTTT	TCTCTTGATTTTTAGTTTCAATCTCT	10
	250	TGCTATGGGTGCAATGGTTA	AAGATTCTAGCAACACCCGA	8
	400	TGGGGACATGAAAACTGGAA	TTCACATACTTTCTCAGGCATT	10
	550	ACTATGGCTATATATGCGAAGAAA	TCCATAAATGTTTCAACTCAGGA	10
	650	GGAATTGCCGAGTTTACACG	TGAGCTCCATGTTGTATTGGA	10
Multiplex 7	200	ACTCCATTTGTGCTTATTGGA	TCTTGAACTAGCCAAAAGTGC	10
	350	TCGAAATATCTTTTAGCTTCAAGAA	AAAACATCATTTTCTTTTGCCCA	10
	450	AGAGTTTGGATGGAAAACTGT	TGCAACTATTCCATCAAAACCA	10
	550	GGTTCAACACCAGGAACAAA	TGCAACACCTATCATCTCATTT	10
	700	GGAAAAGGCAAAGAATCCTCA	ACCATCGCCAGACTTCATTA	10
Multiplex 8	150	TGCAAGAAATGGTGGAACAA	CCTGTTGCAATAGTTGGTGT	10
	250	TGGTAGAAGAAACAATAAAAAGATTTG	AGTCTTGATTTATCGACAGTTCT	10
	350	TTTTGTTTGAAGCTTATTCGTGA	AGTCCATATCCTTTCTCTCTCA	8
	450	AGGAGCTGTTGAGATTTTCAA	GTCGTTGCTCATCTGCTTTT	7
	550	GATGCTGGATTTTGTATGGCT	AGCCAAGAAACTTTCAATATCTCT	10

[a]Primer pairs were selected and grouped into multiplexes using Primer3 [56], multiPLX [57], and CGF Multiplexer [25].
[b]Multiplex primer pair concentrations were optimised for $T_a = 60°C$.

function from the R package *dendextend* (https://github.com/talgalili/dendextend). It then uses the reference phylogeny to create color-coded linkage groups at a 90% cluster similarity level and plots the color-coded tanglegram.

Additional files

Additional file 1: Cluster analysis of *A. butzleri* strains (*n*=12) by multilocus sequence typing, amplified fragment length polymorphism, and comparative genomic fingerprinting.

Additional file 2: Identification of genes targeted by primer pairs in CGF_{40} multiplex PCR.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
ALW participated in study design, led laboratory experiments, participated in bioinformatic analysis, and drafted the manuscript. GDI conceived of the study, contributed to the collection of samples and the isolation of *A. butzleri* isolates, participated in the study design, and helped to draft the manuscript. PK participated in the study design and led various aspects of bioinformatic analysis. LBS participated in study design and helped to draft the manuscript. ENT participated in study design, bioinformatic analysis, and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements
We thank Dr. Valerie Boras and the staff of the Microbiology Diagnostic Laboratory at the Chinook Regional Hospital, Lethbridge, AB for providing stools from diarrheic individuals. We also thank human volunteers for contributing non-diarrheic stool samples. Collection, processing, and/or isolation of *A. butzleri* from the various samples were conducted by the following individuals at the Lethbridge Research Centre: Philip Kubara and Randy Wilde collected surface water samples; Randy Wilde, Amy Wright, and Skip Inglis collected fecal samples from non-human animals; Jenny Gusse, Philip Kubara and Xhevrie Qelag processed water samples; Kathaleen House processed human stool samples; and Greg Frick, Gwen Leusink, Rachel Poberznick, Xhevrie Qelag, Amy Wright, and Mitchel Stevenson completed preliminary identification of isolates. Illumina sequencing was conducted by the Michael Smith Genome Sciences Centre at the University of British Columbia, Vancouver.

Author details
[1]Agriculture and Agri-Food Canada, 5403 – 1st Avenue S, Lethbridge, AB, Canada. [2]Department of Biological Sciences, University of Lethbridge, Lethbridge, AB, Canada. [3]Public Health Agency of Canada, Township Rd. 9-1, Lethbridge, AB, Canada.

Received: 22 October 2014 Accepted: 15 April 2015
Published online: 07 May 2015

References
1. Vandamme P. Taxonomy of the family *Campylobacteraceae*. 2nd ed. Washington DC, USA: ASM Press; 2000.
2. Merga JY, Williams NJ, Miller WG, Leatherbarrow AJH, Bennett M, Hall N, et al. Exploring the diversity of *Arcobacter butzleri* from cattle in the UK using MLST and whole genome sequencing. PLoS Biol. 2013;8(2):12.
3. Nieva-Echevarria B, Martinez-Malaxetxebarria I, Girbau C, Alonso R, Fernández-Astorga A. Prevalence and genetic diversity of *Arcobacter* in food products in the north of Spain. J Food Prot. 2013;76(8):1447–50.
4. Revez J, Huuskonen M, Ruusunen M, Lindström M, Hänninen M-L. *Arcobacter* species and their pulsed-field gel electrophoresis genotypes in Finnish raw milk during summer 2011. J Food Prot. 2013;76(9):1630–2.
5. Rasmussen LH, Kjeldgaard J, Christensen JP, Ingmer H. Multilocus sequence typing and biocide tolerance of *Arcobacter butzleri* from Danish broiler carcasses. BMC Res Notes. 2013;6(322):7.
6. Miller WG, Wesley IV, On SLW, Houf K, Megraud F, Wang G, et al. First multi-locus sequence typing scheme for *Arcobacter* spp. BMC Microbiol. 2009;9(1):196.
7. de Boer RF, Ott A, Güren P, van Zanten E, van Belkum A, Kooistra-Smid AMD. Detection of *Campylobacter* species and *Arcobacter butzleri* in stool samples by use of real-time multiplex PCR. J Clin Microbiol. 2013;51(1):253–9.
8. Miller WG, Parker CT, Rubenfield M, Mendz GL, Wösten MMSM, Ussery DW, et al. The complete genome sequence and analysis of the Epsilonproteobacterium *Arcobacter butzleri*. PLoS Biol. 2007;2(12):e1358.
9. Douidah L, De Zutter L, Vandamme P, Houf K. Identification of five human and mammal associated *Arcobacter* species by a novel multiplex-PCR assay. J Microbiol Methods. 2010;80(3):281–6.
10. Samie A, Obi CL, Barrett LJ, Powell SM, Guerrant RL. Prevalence of *Campylobacter* species, *Helicobacter pylori* and *Arcobacter* species in stool samples from the Venda region, Limpopo, South Africa: Studies using molecular diagnostic methods. J Infection. 2007;54(6):558–66.
11. Houf K, Stephan R. Isolation and characterization of the emerging foodborne pathogen *Arcobacter* from human stool. J Microbiol Methods. 2007;68(2):408–13.
12. Collado L, Gutiérrez M, González M, Fernández H. Assessment of the prevalence and diversity of emergent Campylobacteria in human stool samples using a combination of traditional and molecular methods. Diagn Microbiol Infect Dis. 2013;75(4):3.
13. Fitzgerald C, Helsel LO, Nicholson MA, Olsen SJ, Swerdlow DL, Flahart R, et al. Evaluation of methods for subtyping *Campylobacter jejuni* during an outbreak involving a food handler. J Clin Microbiol. 2001;39(7):2386–90.
14. Van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, et al. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect. 2007;13:46.
15. Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A. 1998;95(6):3140–5.
16. Korczak BM, Zurfluh M, Emler S, Kuhn-Oertli J, Kuhnert P. Multiplex strategy for multilocus sequence typing, fla typing, and genetic determination of antimicrobial resistance of *Campylobacter jejuni* and *Campylobacter coli* isolates collected in Switzerland. J Clin Microbiol. 2009;47(7):1996–2007.
17. Lévesque S, Michaud S, Arbeit RD, Frost EH. High-resolution melting system to perform multilocus sequence typing of *Campylobacter jejuni*. PLoS ONE. 2011;6(1):e16167.
18. Miller JR, Koren S, Sutton G. Assembly algorithms for next-generation sequencing data. Genomics. 2010;95(6):315–27.
19. Cody AJ, McCarthy ND, Jansen Van Rensburg M, Isinkaye T, Bentley SD, Parkhill J, et al. Real-time genomic epidemiological evaluation of human *Campylobacter* isolates by use of whole-genome multilocus sequence typing. J Clin Microbiol. 2013;51(8):2526–34.
20. Struelens M, Brisse S. From molecular to genomic epidemiology: transforming surveillance and control of infectious diseases. Euro Surveill. 2013;18(4):e20386.
21. Taboada EN, Ross SL, Mutschall SK, MacKinnon JM, Roberts MJ, Buchanan CJ, et al. Development and validation of a comparative genomic fingerprinting method for high-resolution genotyping of *Campylobacter jejuni*. J Clin Microbiol. 2012;50(3):788–97.
22. Clark CG, Taboada E, Grant CCR, Blakeston C, Pollari F, Marshall B, et al. Comparison of molecular typing methods useful for detecting clusters of Campylobacter jejuni and C. coli isolates through routine surveillance. J Clin Microbiol. 2011;50(3):47.
23. Taylor DN, Perlman DM, Echeverria PD, Lexomboon U, Blaser MJ. *Campylobacter* immunity and quantitative excretion rates in Thai children. J Infect Dis. 1993;168:754–8.
24. Kokotovic B, On SLW. High-resolution genomic fingerprinting of *Campylobacter jejuni* and *Campylobacter coli* by analysis of amplified fragment length polymorphisms. FEMS Microbiol Lett. 1999;173(1):77–84.
25. Kruczkiewicz P. A comparative genomic framework for the *in silico* design and assessment of molecular typing methods using whole-genome sequence data with application to *Listeria monocytogenes*. *M.Sc*. Lethbridge, AB: University of Lethbridge; 2013.
26. Elliot EJ. Acute gastroenteritis in children. Brit Med J. 2007;334(7583):35–40.
27. Notifiable Diseases On-Line. Public Health Agency of Canada, Ottawa ON. [http://dsol-smed.phac-aspc.gc.ca/dsol-smed/ndis/index-eng.php]
28. Inglis GD, McAllister TA, Larney FJ, Topp E. Prolonged survival of *Campylobacter* species in bovine manure compost. Appl Environ Microbiol. 2010;76(4):1110–9.

29. Hannon SJ, Taboada EN, Russell ML, Allan B, Waldner C, Wilson HL, et al. Genomics-based molecular epidemiology of *Campylobacter jejuni* isolates from feedlot cattle and from people in Alberta Canada. J Clin Microbiol. 2009;47(2):410–20.
30. Collado L, Figueras MJ. Taxonomy, epidemiology, and clinical relevance of the genus *Arcobacter*. Clin Microbiol Rev. 2011;24(1):174–92.
31. Douidah L, de Zutter L, Baré J, De Vos P, Vandamme P, Vandenberg O, et al. Occurrence of putative virulence genes in *Arcobacter* species isolated from humans and animals. J Clin Microbiol. 2012;50(3):735–41.
32. Douidah L, De Zutter L, Baré J, Houf K. Towards a typing strategy for *Arcobacter* species isolated from humans and animals and assessment of the *in vitro* genomic stability. Foodborne Pathog Dis. 2014;11(4):272–80.
33. Jolley KA, Maiden MCJ. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11:595.
34. Taboada EN, MacKinnon JM, Luebbert CC, Gannon VPJ, Nash JHE, Rahn K. Comparative genomic assessment of Multi-Locus Sequence Typing: rapid accumulation of genomic heterogeneity among clonal isolates of *Campylobacter jejuni*. BMC Evol Biol. 2008;8(1):229.
35. Dagerhamn J, Blomberg C, Browall S, Sjostrom K, Morfeldt E, Henriques-Normark B. Determination of accessory gene patterns predicts the same relatedness among strains of *Streptococcus pneumoniae* as sequencing of housekeeping genes does and represents a novel approach in molecular epidemiology. J Clin Microbiol. 2008;46(3):863–8.
36. Deng X, Phillippy A, Li Z, Salzberg S, Zhang W. Probing the pan-genome of *Listeria monocytogenes*: new insights into intraspecific niche expansion and genomic diversification. BMC Genomics. 2010;11(1):500.
37. Carrillo CD, Kruczkiewicz P, Mutschall S, Tudor A, Clark C, Taboada EN. A framework for assessing the concordance of molecular typing methods and the true strain phylogeny of *Campylobacter jejuni* and *C. coli* using draft genome sequence data. Front Cell Infect Microbiol. 2012;2(57):12.
38. Taboada EN, Clark CG, Sproston EL, Carrillo CD. Current methods for molecular typing of *Campylobacter* species. J Microbiol Methods. 2013;95(1):24–31.
39. On SLW, Atabay HI, Amisu KO, Coker AO, Harrington CS. Genotyping and genetic diversity of *Arcobacter butzleri* by amplified fragment length polymorphism (AFLP) analysis. Lett Appl Microbiol. 2004;39(4):347–52.
40. Merga JY, Leatherbarrow AJH, Winstanley C, Bennett M, Hart CA, Miller WG, et al. Comparison of *Arcobacter* isolation methods, and diversity of *Arcobacter* spp. in Cheshire, United Kingdom. Appl Environ Microbiol. 2011;77(5):1646–50.
41. Hume ME, Harvey RB, Stanker LH, Droleskey RE, Poole TL, Zhang H-B. Genotypic variation among *Arcobacter* isolates from a farrow-to-finish swine facility. J Food Prot. 2001;64(5):645–51.
42. Inglis GD, Boras VF, Houde A. Enteric campylobacteria and RNA viruses associated with healthy and diarrheic humans in the Chinook health region of southwestern Alberta Canada. J Clin Microbiol. 2011;49(1):209–19.
43. Blaser MJ. Epidemiologic and clinical features of *Campylobacter jejuni* infections. J Infect Dis. 1997;176(2):S103–5.
44. On SLW, Harrington CS, Atabay HI. Differentiation of *Arcobacter* species by numerical analysis of AFLP profiles and description of a novel *Arcobacter* from pig abortions and turkey faeces. J Appl Microbiol. 2003;95(5):1096–105.
45. Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, et al. The NCBI BioSystems database. Nucleic Acids Res. 2010;38 suppl 1:D492–6.
46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
47. Ewing B, Green P. Base-calling of automated sequencer traces using Phred II Error probabilities. Genome Res. 1998;8(3):186–94.
48. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Biral I. ABySS: A parallel assembler for short read sequence data. Genome Res. 2009;19:1117–23.
49. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST server: rapid annotations using subsystems technology. BMC Genomics. 2008;9(75):15.
50. Kruczkiewicz P, Mutschall S, Barker D, Thomas J, Van Domselaar G, Gannon VPJ, et al. MIST: a tool for rapid in silico generation of molecular data from bacterial genome sequences. https://bitbucket.org/peterk87/microbialinsilicotyper/wiki/Home#markdown-header-paper.
51. R: A language and environment for statistical computing [http://www.R-project.org]
52. Carriço JA, Costa-Silva C, Melo-Cristino J, Pinto FR, de Lencastre H, Almeida JS, et al. Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant *Streptococcus pyogenes*. J Clin Microbiol. 2006;44(7):2524–32.
53. Robinson DF, Foulds LR. Comparison of phylogenetic trees. Math Biosci. 1981;53(1–2):131–47.
54. Severiano A, Pinto FR, Ramirez M, Carriço JA. Adjusted Wallace Coefficient as a measure of congruence between typing methods. J Clin Microbiol. 2011;49(11):3997–4000.
55. Rodrigo AG, Kelly-Borges M, Bergquist PR, Bergquist PL. A randomisation test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree. N Z J Bot. 1993;31(3):257–68.
56. Rozen S, Skaletsky HJ. Primer3 on the WWW for general users and for biologist programmers. Totowa, NJ: Humana Press Inc; 2000.
57. Kaplinski L, Andreson R, Puurand T, Remm M. MultiPLX: automatic grouping and evaluation of PCR primers. Bioinformatics. 2005;21(8):1701–2.
58. Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol. 1988;26(11):2465–6.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central