GyrB versus 16s rRna sequencing for the identification of *Campylobacter jejuni*, *Campylobacter coli*, and *Campylobacter lari*

Nereus William Gunther IV,[1]
Jonnee Almond,[1] Xianghe Yan,[1]
David S. Needleman[2]

[1]Molecular Characterization of Foodborne Pathogens Research Unit, U.S. Department of Agriculture, Agricultural Research Service, Eastern Regional Research Center, Wyndmoor, PA, USA; [2]Integrated Biomolecular Resource Core Facility, U.S. Department of Agriculture, Agricultural Research Service, Eastern Regional Research Center, Wyndmoor, PA, USA

Abstract

Species of the genus *Campylobacter* are responsible for the largest number of bacterial food-borne illness cases occurring yearly in the developed world. The majority of disease is caused by three of the thermotolerant *Campylobacter* species: *Campylobacter jejuni*, *Campylobacter coli*, and *Campylobacter lari*. An inability to differentiate these three species using the commonly employed *16S rRNA* sequencing procedure has led to the development of alternative methods to identify these bacteria. Some of these methods include the utilization of the *gyrB* gene. A reliable method was developed for the differentiation of *C. jejuni*, *C. coli*, and *C. lari* employing the *gyrB* gene. It involves amplification and sequencing of a species-variable region of the gene with a single pair of DNA primers. The method works well for the separation and organization of the three *Campylobacter* strains as well as satisfying the suggested guidelines for sequence based identification for most strains investigated.

Introduction

Campylobacter species are the major causative agent of food-borne gastrointestinal bacterial infections in the developed world.[1-3] The genus *Campylobacter* is comprised of 16 different species.[4] *Campylobacter* jejuni accounts for the majority of the disease attributable to *Campylobacter* with C. coli and

C. lari responsible for most of the remainder.[5,6] Additionally, these three species comprise three of the four known thermotolerant species of *Campylobacter*. A commonly used method for the differentiation of bacterial species is *16S rRNA* sequencing.[7] However, in the case of *C. jejuni*, C. coli and *C. lari* differentiation by the *16S rRNA* method is not sufficiently discriminating given the high degree of similarity found between the *16S rRNA* gene sequences for these three species.[8] To address this issue other researchers have utilized the *gyrB* gene, which encodes the β subunit of DNA gyrase, to more fully differentiate bacterial species where the *16S rRNA* sequences are too well conserved.[9] Previous research in other laboratories used multiple primer pair PCR-based amplification and restriction fragment length polymorphism analysis (RFLP) of the *gyrB* gene to differentiate *Campylobacter* species.[10] Using *gyrB* we have developed a method capable of reliably differentiating the three main human disease causing *Campylobacter* species, *C. jejuni*, *C. coli* and *C. lari* while supplying additional sequence data with potential use in taxonomic studies. This method utilizes a single pair of degenerate DNA primers for both the amplification and subsequent sequencing of an 811-bp fragment of the *gyrB* gene from the bacterial strains to be identified. This method was found to provide excellent discrimination between strains of *C. jejuni*, *C. coli*, and *C. lari*. The advantage of this method with regards to species identification and differentiation is readily apparent when compared to results derived from the *16S rRNA* method.

Materials and Methods

Bacterial strains

In total, 25 different reference strains of *Campylobacter* species were used in this study. The *Campylobacter* strains chosen for this work were well characterized having been identified to the species level by multiple methods. Additionally, the *C. jejuni*, *C. coli*, and *C. lari* strains utilized were isolated from both human and animal sources and minimally passaged. *C. coli*: RM1403, RM4796, RM1529, RM4803, RM4780, RM3232, RM1840, RM2221, RM1182, RM1182; *C. lari*: RM2820, RM3600, RM4110, RM2099, RM2100, and *C. jejuni*: RM5032, RM1188, RM1449, RM1464, RM3194, RM1247, RM3405, RM1221, RM1246 strains were kindly provided by Dr. Robert Mandrell (USDA, Agricultural Research Service, Western Regional Research Center, Albany, CA, USA). *C. coli* strain ATCC33559 and *C. lari* strain ATCC35221 were purchased from ATCC (Manassas, VA, USA). Strains were

cultured from frozen stocks (-80°C) directly onto Brucella (Becton Dickinson, Sparks, MD, USA) agar plates (1.5% agarose) and incubated in a microaerobic chamber (MACS-VA, Don Whitley, UK) (5% O_2, 10% CO_2, 85% N_2) at 42°C for 24 h. Additionally, four Escherichia coli strains, four Salmonella strains and a *Campylobacter showae* and *concisus* were chosen randomly from our lab's collections and utilized to test the specificity of the primers used in the *gyrB* based identification method.

DNA extraction

Bacterial DNA was extracted from a single *Campylobacter* colony using 25 μL of nuclease-free water and 25 μL of PrepMan Ultra reagent (Applied Biosystems, Foster City CA, USA) placed in a 1.5 mL micro centrifuge tube. The samples were heated in boiling water for 15 minutes, allowed to cool to room temperature and centrifuged at 16000 X g for 2 min. The supernatant (containing the DNA) was transferred to a clean 1.5 ml microcentrifuge tube.

16S rRNA and gyrB amplification

PCR amplification of a 1540-bp fragment of the *16S rRNA* gene used primer pair EubA-(AAGGAGGTGATCCANCCRCA) and EubB-(AGAGTTTGATCMTGGCTCAG)[11] (Integrated DNA Technologies, Coralville, IA) and amplification of a 811-bp fragment of the *gyrB* gene

used primer pair JCL forward(GHCAA-GAATTTTCAGAAGGWAAAGT) and JCL reverse-(GGATCTTTACTTTGACAATCAGCTA).

Primers JCL forward and reverse were designed by aligning multiple *C. jejuni*, *C. coli*, and *C. lari gyrB* sequences collected from sequences deposited in GenBank (NCBI, Bethesda, MD, USA). Areas of conserved sequences surrounding areas of species variable sequences were identified. Sequence areas that would produce PCR products between 600-1000 bp in length were selected and from these two promising sequences that could be amplified by primer pairs with similar characteristics were used to design two sets of primers. The primers were tested through PCR amplification and one primer pair was judged to be the most promising labeled JCL forward and reverse. PCR amplification was performed using a PCR kit (Invitrogen, Carlsbad, CA, USA) in a total volume of 50 μL containing 5 μL of 10x reaction buffer, 1 μL of dNTPs, 5 μL of each primers (10 μM), 1μL of template DNA, 0.25 μL of Taq DNA (5 u/μL) and 32.75 μL of PCR water to make up the final volume. Amplification was performed using an iCycler thermocycler (BioRad, Hercules, CA, USA). Thermal cycling conditions for *16S rRNA* was as follows: Initial denaturation at 95°C for 90 s, followed by 30 cycles of 95°C for 30 s, 55°C for 45 s, and 72°C for 60 s with a final extension at 72°C for 5 min. Thermal cycling conditions for *gyrB* gene were as follows: initial denaturation at 95°C for 90 s, followed by 30 cycles of 95°C for 30 s, 50°C for 45 s, and 72°C for 60 s with a final extension at 72°C for 5 min. Eight microliters of PCR product were used for electrophoresis in a 1.0% agarose gel at 120 V for 30-45 min.

DNA Sequencing

Sequencing of the *16S rRNA* PCR fragment utilized the amplification primers EubA and EubB and the nested primers 519F-(CAGCMGCCGCGGTAATWC) and 519R-(GWATTACCGCGGCKGCTG) in order to completely sequence the 1540 bp PCR fragment. To sequence of the 811-bp *gyrB* PCR fragment we were able to use the same primers, JCL forward and JCL reverse, that we used for amplification of the fragment. The sequencing reaction included the following components: 7 μL of 2.5x buffer, 1 μL of primers (3.2 μM), 1 μL of ABI's Big Dye Terminator (Foster City, CA, USA), 1.2 μL of cleaned PCR product and ddH$_2$O to make up the final volume. Cycle sequencing was performed using 30 cycles of 95°C for 10 s, 55°C for 5 s and 60°C for 4 min. After a purification step the sequencing reaction products were analyzed using an Applied Biosystem (Foster City) 3730 DNA sequencer.

Sequence analysis, alignment, and construction of phylogenetic tree for *gyrB* and *16S rRNA* data

After sequence trimming, the sequence data were then combined and assembled to form contiguous stretches of sequence contigs with the Sequencher program, V. 4.9 (Gene Codes Corp, Ann Arbor, MI, USA). The Sequencher software was also used for gel trace analysis, contig assembly verification, contig sequence comparisons, and Phred quality analysis. The alignments of all obtained nucleotide sequences of *gyrB* and *16S rRNA* were generated by using the multiple sequence alignment program, CLUSTALX version 1.83[12] with the default parameters and adjusted manually. Phylogram trees were generated from the cluster analysis using the Bootstrap Neighbor Joining method with 1000 random samplings and a seed number of 111. The resulting phylogram trees were displayed using the TreeView X program.[13] The sequence similarity and divergence for individual tested strains were

calculated by MegAlign of the DNASTAR program (DNASTAR Inc., Madison, WI, USA).

Results

16S rRNA sequences

Sequencing of the 1540-bp *16S rRNA* PCR fragments generated 1418 bp of sequence with an average Phred score of ≥20 (≥99% accurate) for each of the 25 *Campylobacter* strains. When these sequences were aligned and displayed as a phylogenetic tree, they formed an indistinct heterogeneous tree with no preferential grouping by species for the three separate *Campylobacter* species investigated (Figure 1).

The sequences of all of the 25 *Campylobacter* strains regardless of the originating species differed from one another by no more then 2.9%. Additionally, the *16S rRNA* sequences of the majority of the *C. jejuni* and
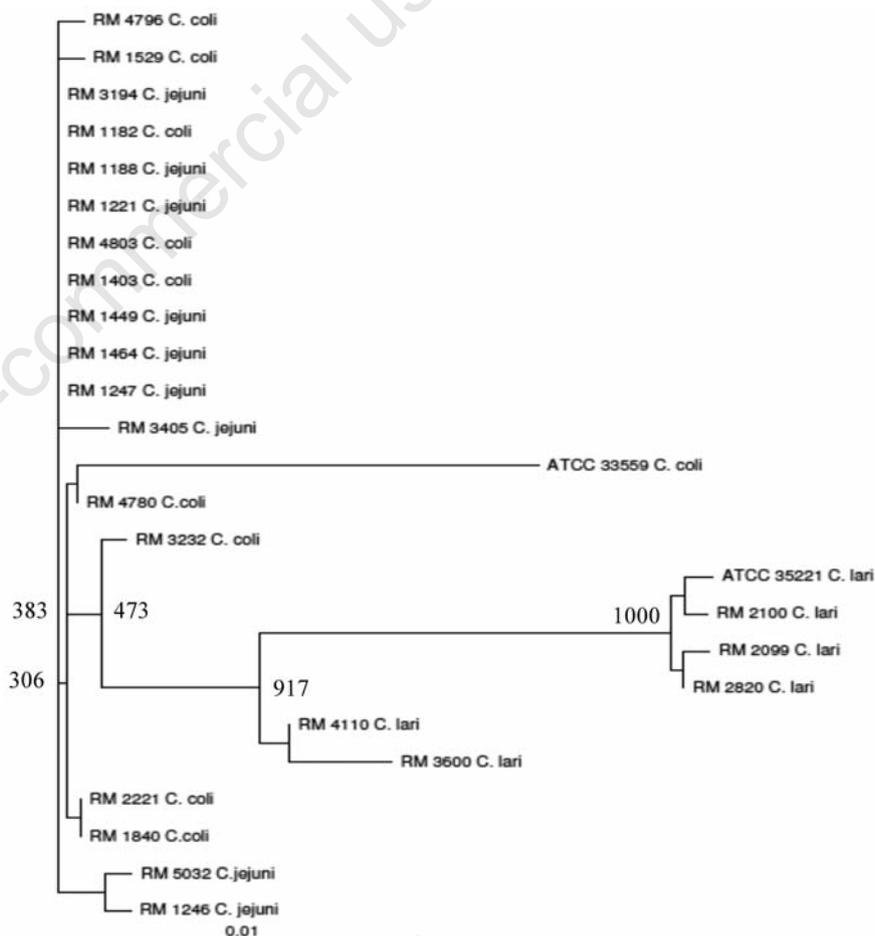

Figure 1. Phylogram tree of 16S rRNA sequences of 25 *Campylobacter* strains of the species *C. jejuni*, *C. coli*, and *C. lari*. Bootstrap values are included for the main branches of the tree.

C. coli strains differed from one another by less then 1% with some differing by only 0.1%.

gyrB sequences

Sequencing of the 811-bp *gyrB* PCR fragment generated 703 bp of sequence with an average Phred score of ≥ 20 ($\geq 99\%$ accurate) for each of the 25 *Campylobacter* strains. The sequences were aligned and displayed in tree form resulting in the strains organizing into three distinct groups (Figure 2).

Group A was comprised exclusively of all ten *C. coli* strains sequenced for this study. Similarly, group B contained only the six *C. lari* strains utilized in the study. Finally, group C was restricted to only the nine *C. jejuni* strains that had been selected for this research. There were only slight differences observed between the *C. coli* strains comprising group A with no more than 2% differences between the *C. coli* *gyrB* sequences. Similarly, in group C there was no more than a 2% difference between the *C. jejuni* *gyrB* sequences produced. However, the *C. lari* strains' *gyrB* sequences that account for group B showed a higher level of strain to strain sequence variations with percent differences in sequences of about 9.5%. This is similar to the differences in sequences observed between the *C. coli* *gyrB* sequences in group A and the *C. jejuni* *gyrB* sequences in group C which range from 9-11.2%. The sequence differences between the *C. jejuni* *gyrB* sequences of group C and the *C. lari* *gyrB* sequences of group B ranged from 11.5-14.4%. Finally the sequence differences between the *C. lari* *gyrB* sequences of group B and the *C. coli* *gyrB* sequences of group A ranged from 13.2-15.8%.

BLAST results from 16S rRNA and gyrB sequences

When the individual sequences derived from the *16S rRNA* or *gyrB* genes of the *Campylobacter* strains were queried against NCBI's bacterial sequence databases using the BLAST (blastn) web-based analysis tool, the *gyrB* sequences produced superior specificity with regards to the species identities of the homologous sequences derived from the database searches.[14] All ten of the *C. coli* strain *gyrB* sequences when compared against the bacterial sequence database produced a collection of *C. coli* sequences that matched the query sequence with identity scores of 99-97%. The first match from the database that was not a *C. coli* sequence only had a maximum identity of 91% (*data not shown*). Conversely, for the *16S rRNA* sequences, the *C. coli* sequence queries returned a list of matching sequences that were a mixture of *C. jejuni* and *C. coli* sequences all possessing identity scores of 99-98%. The nine *C. jejuni* strain *gyrB* query sequences produced a list of *C. jejuni* strain

sequences from the database with 99-98% matching identities. The first non *C. jejuni* sequence match returned from the database comparison had identity scores of no more then 90% (*data not shown*). Again the results of the database searches of the *16S rRNA C. jejuni* sequences were much less specific and returned a mixture of *C. jejuni*, *C. coli*, and *C. lari* sequences with identity scores in the 99-97% range. Finally, the six *C. lari* *gyrB* sequences queried against the bacterial database returned a list of *C. lari* sequences with identities to the query sequences of between 99-90%. The first non- *C. lari* sequences from the lists resulting from the database comparisons had percent identity scores of no greater than 88% to the query sequences (data not

shown). The BLAST results from a comparison of the *16S rRNA C. lari* sequences compared against the bacterial database returned a mixture of *C. lari* and non-*C. lari* strains, with percentage identity scores of 99-97% to the query sequences.

Comparing the *gyrB* sequences produced by this research against NCBI's bacterial sequence database allowed for the successful identification of twenty-five different *C. jejuni*, *C. coli* and *C. lari* strains with zero resulting misidentifications or false-positives. Additionally, ten other bacterial strains were included in the *gyrB* sequencing assay to address concerns about the potential for non-specific cross reactions occurring in this method. A collection of *E. coli*, *Salmonella* and non-*jejuni*, *coli* and *lari*
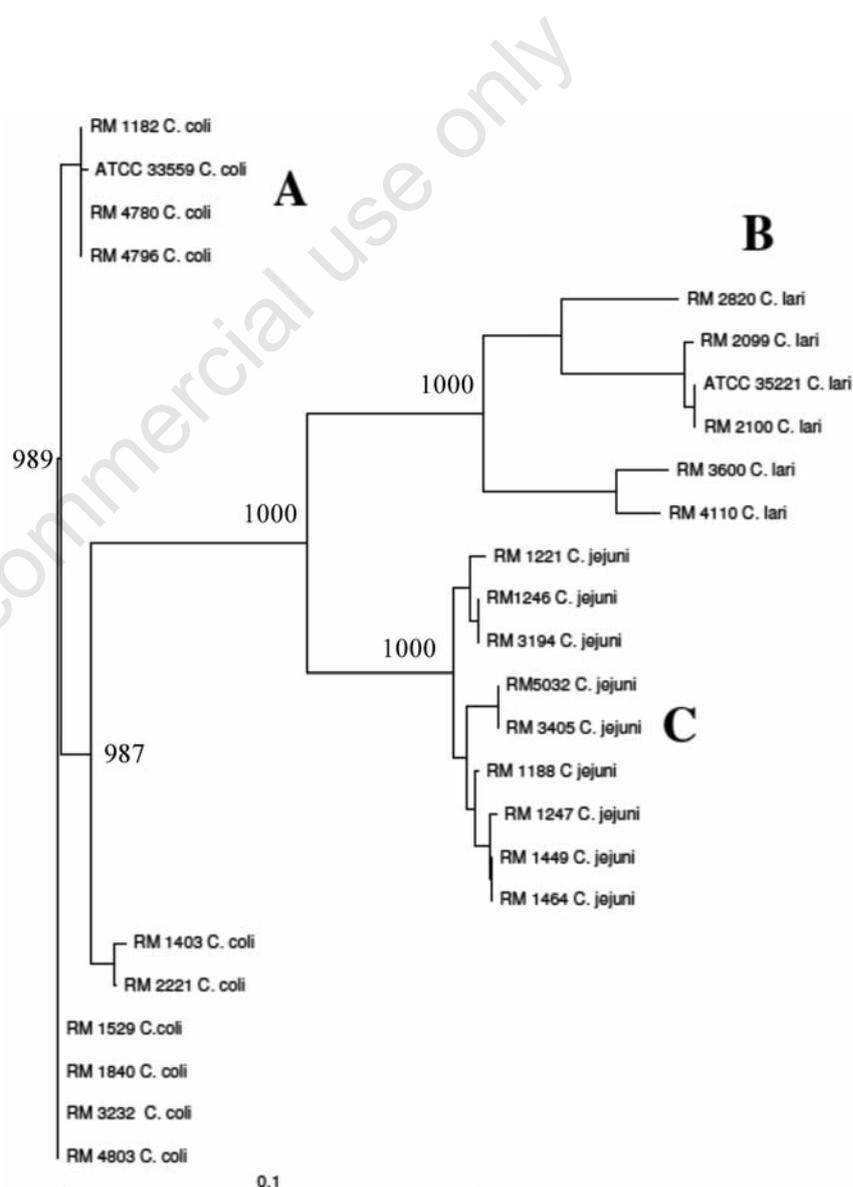

Figure 2. Phylogram tree of *gyrB* sequences of 25 *Campylobacter* strains of the species *C. jejuni*, *C. coli*, and *C. lari*. Bootstrap values are included for the main branches of the tree.

Discussion

We have described a method for the PCR amplification and sequencing of a fragment of the *gyrB* gene capable of reliably identifying and differentiating the three main human disease causing *Campylobacter* strains: *C. jejuni*, *C. coli*, and *C. lari*. This method has the advantage of being able to use the same primers for both the amplification and sequencing of the *gyrB* fragment. An 811-bp fragment of the *gyrB* gene was selected for this assay since it allowed for a single sequencing run covering the entire length of the resulting gene fragment. Therefore, the two sequencing primers would be sufficient to produce overlapping sequences the entire length of the *gyrB* fragment resulting in confirmatory data to increase the overall confidence in the reported sequence. The resulting 703 bp of high quality sequence proved sufficient to separate the collection of *Campylobacter* strains into three distinct groups which correlated perfectly with their species identities. Additionally BLAST based comparisons to NCBI's bacterial sequence database for many of the individual sequences produced sequence matches with ≥99% sequence homology to bacterial sequences from the same species. Suggested guidelines in the literature describing the cut-off levels for sequence homology (percent identity) for proper identification of bacterial species ranging from ≥97% to ≥99%.[15,16] The *C. jejuni* and *C. coli* strain *gyrB* sequences investigated in this manuscript satisfy these guidelines. However, it should be noted that some of the *C. lari* strains failed to meet this standard. Based on the phylogenetic tree results there appears to be a large amount of *C. lari* strain to strain variation with regards to the *gyrB* sequence. This leads to a less tight grouping of the *C. lari* strains in the phylogenetic tree but does not affect the overall separation of the *C. lari* strains from the more tightly grouped *C. jejuni* or *C. coli* strains. Therefore, the sequence of the *gyrB* gene remains a good tool for separating strains of *C. coli*, *C. jejuni*, and *C. lari* from one another at the species level. It is also a good tool for identification of *C. coli* and *C. jejuni* strains through BLAST comparisons to sequence databases. However, the variations in the *gyrB* sequences between strains of *C. lari* make it difficult to use those sequences alone for the identification of *C. lari* strains based on the currently accepted identity guidelines.

This problem with *C. lari* may be compounded by the fact that there are considerably fewer *gyrB* sequences in the nucleotide database for *C. lari* as compared to *C. coli* and *C. jejuni*. *C. lari* has been shown to be a genetically diverse species, and the *C. lari* strains used in our study may be members of different subgroups.[17,18] By increasing the number of *C. lari* *gyrB* genes sequenced and available in bacterial sequence databases it maybe possible that distinct *gyrB* subgroups of *C. lari* will emerge and allow for superior sequence based identification of *C. lari* strains.

The sequence data produced by this method provides for the possibility of additional taxonomic studies similar to the one just mentioned regarding determining if there might be subgroupings of *C. lari* strains. This is an advantage that a sequence based identification system has over multiplex PCR identification techniques that are generally quicker and simpler to perform. Additionally, the *gyrB* gene is not the only target gene with the power to identify pathogenic *Campylobacter* species. Other research has shown that the *lpxA* and *GTPase* genes can also clearly identify and separate many of the *Campylobacter* species.[19,20] Therefore, it may be advantageous to also utilize these gene targets in addition to this *gyrB*-based method in any future taxonomic studies focusing on identifying *Campylobacter* sub-groups.

References

1. CDC. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food--10 States, United States, 2005. MMWR Morb Mortal Wkly Rep 2006;55:392-5.
2. Fields PI, Swerdlow DL. Campylobacter jejuni. Clin Lab Med 1999;19:489-504.
3. Park SF. The physiology of Campylobacter species and its relevance to their role as foodborne pathogens. Int J Food Microbiol 2002;74:177-88.
4. Mandrell RE, Harden LA, Bates A, et al. Speciation of Campylobacter coli, C. jejuni, C. helveticus, C. lari, C. sputorum, and C. upsaliensis by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl Environ Microbiol 2005;71: 6292-307.
5. Moore JE, Corcoran D, Dooley JS, et al. Campylobacter. Vet Res 2005;36:351-82.
6. Tam CC, O'Brien SJ, Adak GK, et al. Campylobacter coli - an important food-borne pathogen. J Infect 2003;47:28-32.
7. Fox GE, Stackebrandt E, Hespell RB, et al. The phylogeny of prokaryotes. Science 1980;209:457-63.
8. Gorkiewicz G, Feierl G, Schober C, et al. Species-specific identification of campylobacters by partial 16S rRNA gene sequencing. J Clin Microbiol 2003;41: 2537-46.
9. Yamamoto S, Harayama S. PCR Amplification and Direct Sequencing of gyrB Genes with Universal Primers and Their Application to the Detection and Taxonomic Analysis of Pseudomonas putida Strains. Appl Environ Microbiol 1995;61: 1104-9.
10. Kawasaki S, Fratamico PM, Wesley IV, Kawamoto S. Species-specific identification of Campylobacters by PCR-restriction fragment length polymorphism and PCR targeting of the gyrase B gene. Appl Environ Microbiol 2008;74:2529-33.
11. Cottrell MT, Kirchman DL. Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization. Appl Environ Microbiol 2000;66:5116-22.
12. Thompson JD, Gibson TJ, Plewniak F, et al. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997;25:4876-82.
13. Page RD. TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 1996;12:357-8.
14. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. J Mol Biol 1990; 215:403-10.
15. Drancourt M, Bollet C, Carlioz A, et al. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. J Clin Microbiol 2000;38:3623-30.
16. Stackebrandt E, Goebel BM. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. J Syst Bacteriol 1994;44:846-9.
17. Van Doorn LJ , Verschuuren-Van Haperen A, Van Belkum A, et al. Rapid identification of diverse Campylobacter lari strains isolated from mussels and oysters using a reverse hybridization line probe assay. J Appl Microbiol 1998;84:545-50.
18. Duim B, Wagenaar J. Amplified fragment-length polymorphism and protein profiling for identification of Campylobacter lari subgroups. Methods Mol Biol 2006;345: 119-30.
19. Van Doorn LJ, Verschuuren-van Haperen A, Burnens A, et al. Rapid identification of thermotolerant Campylobacter jejuni, Campylobacter coli, Campylobacter lari, and Campylobacter upsaliensis from various geographic locations by a GTPase-based PCR-reverse hybridization assay. J Clin Microbiol 1999;37:1790-6.
20. Klena JD, Parker CT, Knibb K, et al. Differentiation of Campylobacter coli, Campylobacter jejuni, Campylobacter lari, and Campylobacter upsaliensis by a multiplex PCR developed from the nucleotide sequence of the lipid A gene lpxA. J Clin Microbiol 2004;42:5549-57.