SMRT sequencing of the *Campylobacter coli* BfR-CA-9557 genome sequence reveals unique methylation motifs

Andreas E. Zautner¹*, Anne-Marie Goldschmidt¹, Andrea Thürmer², Jörg Schuldes², Oliver Bader¹, Raimond Lugert¹, Uwe Groß¹, Kerstin Stingl³, Gabriela Salinas⁴ and Thomas Lingner⁴

**Abstract**

**Background:** *Campylobacter species* are the most prevalent bacterial pathogen causing acute enteritis worldwide. In contrast to *Campylobacter jejuni*, about 5% of *Campylobacter coli* strains exhibit susceptibility to restriction endonuclease digestion by DpnI cutting specifically 5'-GmATC-3' motifs. This indicates significant differences in DNA methylation between both microbial species. The goal of the study was to analyze the methylome of a *C. coli* strain susceptible to DpnI digestion, to identify its methylation motifs and restriction modification systems (RM-systems), and compare them to related organisms like *C. jejuni* and *Helicobacter pylori*.

**Results:** Using one SMRT cell and the PacBio RS sequencing technology followed by PacBio Modification and Motif Analysis the complete genome of the DpnI susceptible strain *C. coli* BFR-CA-9557 was sequenced to 500-fold coverage and assembled into a single contig of 1.7 Mbp. The genome contains a CJIE1-like element prophage and is phylogenetically closer to *C. coli* clade 1 isolates than clade 3. 45,881 6-methylated adenines (ca. 2.7% of genome positions) that are predominantly arranged in eight different methylation motifs and 1,788 4-methylated cytosines (ca. 0.1%) have been detected. Only two of these motifs correspond to known restriction modification motifs. Characteristic for this methylome was the very high fraction of methylation of motifs with mostly above 99%.

**Conclusions:** Only five dominant methylation motifs have been identified in *C. jejuni*, which have been associated with known RM-systems. *C. coli* BFR-CA-9557 shares one (RAATTY) of these, but four ORFs could be assigned to putative Type I RM-systems, seven ORFs to Type II RM-systems and three ORFs to Type IV RM-systems. In accordance with DpnI prescreening RM-system IIP, methylation of GATC motifs was detected in *C. coli* BFR-CA-9557. A homologous IIP RM-system has been described for *H. pylori*. The remaining methylation motifs are specific for *C. coli* BFR-CA-9557 and have been neither detected in *C. jejuni* nor in *H. pylori*.

The results of this study give us new insights into epigenetics of Campylobacteraceae and provide the groundwork to resolve the function of RM-systems in *C. coli*.

**Keywords:** *Campylobacter coli*, Genome, Methylation, Motifs, Methylome, Restriction modification systems, Isoschizomer digestion assay, SMRT sequencing, PacBio
Background

Campylobacteriosis is the most prevalent form of bacterial acute enteritis worldwide. In symptomatic cases it is characterized by a prodromal phase with fever, vomiting, and headaches followed by watery or bloody diarrhea and abdominal cramps [1, 2]. In consequence of acute enteritis, extraintestinal post-infectious sequelae, namely, the Guillain-Barré syndrome, inflammatory bowel disease, and reactive arthritis may occur [3, 4]. The average incidence reported in the European Union was 64.8 per 100,000 population in 2013 [5], in the USA 14.3 cases per 100,000 population in 2012, and in China 161 cases per 100,000 population in urban areas compared to 37 cases per 100,000 population in rural areas [6]. In Europe, 80.6 % were reported to have been caused by Campylobacter jejuni and 7.1 % by Campylobacter coli [5].

C. coli is phylogenetically subdivided into three clades [7, 8]: clade 1 isolates commonly colonize swine but can also be isolated from poultry and humans, although less frequently. Clades 2 and 3 are typically isolated from environmental waters [8, 9].

At the moment, seven completed C. coli chromosomal genome sequences [10–13], several scaffold genomes, and various contigs have been deposited in the NCBI Genome database [14–17]. The completed genome sequences, range from 1.685 to 1.872 Mb, have a G + C content of about 31 to 32 %, and contain 1715 – 1970 predicted genes including 1642 – 1861 protein coding ORFs [10–13].

One of the major epigenetic mechanisms in prokaryotes is DNA methylation [18]. DNA methylation patterns influence gene expression [19], through silencing of transcription [20, 21] as well as DNA replication initiation [22, 23] and mismatch repair [24]. DNA methylation also serves as a protection of the host genome against extraneous DNA [18] through restriction-modification systems (RM-systems). RM-systems consist of two components: (i) a restriction endonuclease that recognizes a specific DNA motif and (ii) a cognate DNA methyltransferase that methylates the same DNA, preventing its cleavage by the restriction endonuclease [25]. The majority of RM-systems can be categorized into four types [25–29]:

Type I RM-systems typically consist of three types of subunits: two restriction endonuclease subunits (R), which facilitate DNA cleavage, one specificity subunit (S) for recognition of specific DNA sequence motifs, and two DNA methylase subunits (M) that catalyse N⁶ adenine methylation [30, 31]. This composition enables Type I RM-systems to digest unmethylated DNA, whereas hemimethylated DNA is further methylated and fully methylated DNA is insusceptible to restriction [32].

Type II RM-systems are mostly composed of two homodimeric R subunits and a separated M subunit. The R and M subunits recognize the same DNA motif, which is typically a 4–8 bp palindrome [33].

Type III RM-systems are comprised of two modification (Mod) subunits and two restriction (R) subunits. Type III RM-systems must bind to two inversely oriented copies of its 5–6 bp asymmetric recognition motif. Cleavage of unmethylated DNA typically occurs 25–27 bp away from the binding sites [34].

Type IV RM-systems consist of two separate R subunits cleaving DNA that contains methylated, hydroxymethylated or glucosyl-hydroxymethylated cytosines. Cleavage typically occurs 30 bp away from one of the binding sites [35]. Furthermore, there exist many orphan DNA methylases that are not part of a RM-system e.g. DNA adenine methylase (Dam) and cell cycle-regulated DNA methyltransferase (CcrM) [29].

It must be considered that the genes encoding for particular components of a RM-system are diverse within microbial species i.e. C. jejuni and C. coli [36–38]. In C. jejuni the putative Type I RM-system locus (genes cjeI–cj1553; hsd locus) reveals significant diversity regarding gene order, chromosomal location, intervening ORFs and gene sequence leading to the classification of the C. jejuni hsd loci into at least three families, namely the IAB, IC, and IF family. Especially sequence variations in the hsdS gene suggest at least 30 different target sequence specificities and therewith differences in DNA methylation [39]. Holt and coworkers demonstrated that cjeI constitutes an active restriction-modification Type IIG enzyme in C. jejuni that significantly decreases transformation efficiency with plasmids [40]. Additionally it was shown that Type IIS restriction modification enzyme Cj0030/Cj0031 is subject to phase variable gene expression due to mutations in polyC/G tracts [41–43]. Additionally cjeI and cjeII encode a putative 5-methylcytosine-restriction system, cjeI and cjeII a nuclease that could be part of a RM-system [41, 42].

Furthermore it was shown for C. jejuni that gene product of cjeI is a N⁶-adenine-specific DNA methyltransferase that is not a Dam homologue and not part of a RM-system. Knockout of cjeI affects flagellar appearance, motility, adherence, and invasion indicating its role for epigenetic control of proteins involved in these processes [44]. Further non-ubiquitous modification/methyltransferase gene loci have been observed in C. jejuni ST-677 isolates, namely ftxl (cjeI/cj14950) that showed homology with DNA adenine/modification methylases in Campylobacter rectus, a homolog of icaA1/lnaII (cjeI/cj14940), cjeI/cj14910c predicted as ulcer associated adenine-specific DNA methyltransferase, which is an ortholog of the CATG-specific methyltransferase hpyIM of H. pylori, and the orphan DNA methyltransferase in ORF cjeI [45].
strain-specific RM-systems have been reported for C. jejuni ST403 complex: R. HinPI restriction endonuclease (cje135_03870), Modification methylase Hhal (cje135_03865), and R.Pab1 restriction endonuclease (cje135_02348) [46].

In contrast to C. coli, Type III RM-systems have been reported for C. jejuni subsp. jejuni, C. jejuni subsp. doylei, Campylobacter lari, and Campylobacter upsaliensis [47].

Until to date the methylomes of C. jejuni subsp. jejuni (3 isolates) and of the related microbial species Helicobacter pylori (2 isolates) have been analyzed [48–50], but so far no C. coli methylome has been examined.

Studies performing isoschizomer digestion assays indicated differences in methylation at GATC sites in genomic DNA of C. coli isolates, which suggested host-associated DNA modification systems [51, 52].

In this study we analyzed the first methylome of C. coli using SMRT DNA sequencing. This approach facilitates analysis of methylation motifs as well as RM-system gene loci in parallel. For SMRT sequencing a C. coli isolate was selected, which was tested positive for GATC site methylation isoschizomer digestion assay screening.

Results and discussion
SMRT sequencing and annotation
Screening of 50 C. coli isolates performing 5'-G(m)ATC-3'-specific isoschizomer digestion assays revealed six bacterial isolates showing Dam activity. Of these six isolates, the isolate BFR-CA-9557 was selected for SMRT sequencing, since it reproducibly tested positive in five biological independent analyses. Furthermore, its origin from broiler hearts sampled at a retail store in Berlin, Germany, indicates its relevance for food hygiene and thus for the infection of humans.

MLST typing revealed sequence type ST-1589 and clonal complex CC-828. Accordingly, it is a potentially human pathogenic isolate of clade 1.

Susceptibility testing indicated a quinolone resistant phenotype (zone diameter 0 mm in disc diffusion; minimal inhibition concentration (MIC) measured by broth microdilution: 16 μg/ml ciprofloxacin, 32 μg/ml nalidixic acid). In contrast, the isolate was tested susceptible for macrolides, tetracyclines and aminoglycosides (MIC values: 2 μg/ml erythromycin, <=0.5 μg/ml tetracycline, 1 μg/ml gentamicin, 2 μg/ml streptomycin).

Sequencing of the C. coli isolate using one Pacific Bioscience SMRT cell resulted in 74,742 continuous long reads (CLR) with an average (total) length of 14,514 (1.08*10^6) base pairs (Additional file 1: Figure S1). 142,135 subreads (i.e. individual fragments) of high quality and an average length of 7,602 bp could be extracted from the CLRs.

Utilizing the Pacific Bioscience high-quality de novo genome assembly algorithm (HGAP.2), we obtained a single polished contig of 1,720,506 base pairs with an average 500-fold coverage and a confidence score of >99.99 % (Additional file 2: Figure S2). The G+C content of the contig was 31.4 % and the most closely related genome sequence available in the NCBI genome database was that of C. coli 15–537360 (taxonomy ID 1358410, 87 % coverage). No plasmids were detected in this isolate.

Application of the NCBI/RAST/Prodigal annotation pipelines resulted in 1637/1769/1797 predicted coding sequences and 275/366/453 (16.8/20.7/25.2 %) predicted hypothetical proteins. (GenBank ID: CP011777, Additional files 3 and 4). Furthermore, 54 RNA genes (44 tRNA/9 rRNA/1 ncRNA) were predicted by NCBI, 52 RNAs (43 tRNA/9 rRNA) by RAST, and 54 RNAs (44 tRNA/9 rRNA/1 tmRNA) by Prodigal.

The RAST subsystem coverage was 57 % (1003 genes), of which amino acid metabolism (296 of 1557 associated terms, 19.0 %), protein metabolism (216 terms, 13.9 %), and cofactors, vitamins, prosthetic groups, and pigments (144, 9.2 %) represented the largest groups (Fig. 1).

Figure 2a shows a comparison of the BFR-CA-9557 genome sequence to other C. coli genomes (76339, clade 3 and RM4661, clade 1) using the Artemis tool. Here, homologous regions as identified by BLAST are indicated by red (aligned in the same direction) or blue line segments (opposite direction). It can be seen that C. coli 76339 mainly shows three stretches of inverse homologous regions to BFR-CA-9557 and lacks similarity in the region of the identified CJIIE1 (green box, see also below). RM4661 displays a long stretch of close homology with minor gaps, but also inverse homologous segments within the plasmid and the beginning and end of the chromosome. A similar pattern can be observed for the comparison of BFR-CA-9557 to C. jejuni NCTC 11168 (Fig. 2b). This indicates that C. coli BFR-CA-9557 is phylogenetically closer to clade 1 and therewith to C. jejuni isolates than to clade 3 C. coli isolates.

Poly-G/C tracts and the capsular polysaccharid gene locus
Analysis of poly-G/C tracts within the contig of BFR-CA-9557 revealed 17 locations with homopolymeric stretches of at least eight G or C nucleotides (Tables 1 & 2). Most of these tracts occur in the vicinity of hypothetical proteins, however, two motifs could be identified within the capsular polysaccharid gene locus (cps) close to ORFs identified as glycosyltransferases (EC 2.4.99.-) by RAST. Other ORFs with neighboring poly-G/C motifs comprise e.g. transferases and ligases.

The cps locus of BFR-CA-9557 ranges from 1,656,271 to 1,691,672 (1,660,702-1,686,508 excluding flanking kps regions). Spanning 35,401 bp and encoding 29 ORFs (25,806 bp; 21 ORFs; excluding flanking kps regions) it
**Fig. 1** Pie chart representing RAST subsystems identified in the BFR-CA-9557 genome. The 20 most abundant subsystems on the “category” level as identified by RAST are represented by a particular color indicated at the right-hand side of the figure.

**Fig. 2** Comparison of *C. coli* BFR-CA-9557 genome to those of *C. coli* RM4661 and 76339 *a* as well as *C. jejuni* NCTC 11168 *b* using the Artemis Comparison Tool (ACT). Comparison of the BFR-CA-9577 genome to those of *C. coli* clade 1 strain RM4661 and *C. coli* clade 3 strain 76339 as well as to the genome of *C. jejuni* NCTC 11168 indicates that the clade 1 strain BFR-CA-9577 genome organization is more related to *C. coli* clade 1 and *C. jejuni* strains than to clade 3 strains. The CJIE1-homologue prophage in the BFR-CA-9557 genome is indicated in green.
is of comparable size to the largest known other *Campylobacter* cps cluster, which measures 38 kb excluding flanking kps regions (*C. jejuni* strain X) [53]. The cps loci of different *C. coli* and *C. jejuni* strains are very variable in size and gene content and there are only three genes that are comparatively conserved: the capsular polysaccharide export system periplasmic protein gene *KpsD*, the GDP-mannose 4,6-dehydratase gene and the capsular polysaccharide biosynthesis/export periplasmic protein *wcbA*/*kpsC*. The highest sequence similarities exist to the cps clusters of *C. coli* strains RM4661 (query coverage 71 %, identity 99 %) and FB1 (query coverage 54 %, identity 99 %). Compared to the cps locus of strain X, query coverage of 28 % (identity 87 %) was observed.

The genes for *kpsC*, *hddC*, *gmhA2*, *hddA*, *dmhA*, *fcl-1*, *fcl-2*, hypothetical protein *x.25* and *kpsF* are present in both cps clusters.

### N-linked flagellar glycosylation locus

The N-linked flagellar glycosylation locus extends from 1,319,815-1,332,144 (12,329 bp). It consist of 10 ORFs and is 99 % identical to the corresponding locus in other *C. coli* genomes e.g. RM4661 (clade 1) or 76339 (clade 3). In comparison to other *C. jejuni* strains (e.g. M1 and 81116), insertion of the lipid carrier UDP-N-acetylgalactosaminyltransferase gene and an α-1,4-N-acetylgalactosamine transferase PglH gene was observed.

Directly upstream of the N-linked flagellar glycosylation locus the lipooligosaccharide biosynthesis locus (LOS locus) is located. The LOS locus extends from the UDP-glucose 4-epimerase gene *galE* to the D-glycero-D-manno-heptose 1,7-bisphosphate phosphatase gene *rfaD*.

### Table 1 Positions of poly-G tracts in the BFR-CA-9557 genome

| Position | Length | RAST annotation of closest ORF | Pos. rel. to ORF location | ORF location |
|----------|--------|-------------------------------|---------------------------|--------------|
| 46933    | 9      | hypothetical protein          | −52                       | 46985        |
| 258726   | 10     | FIG 00470070: hypothetical protein | −15                       | 258741       |
| 689956   | 10     | UDP-N-acetyl:muramoylalanine--D-glutamate ligase (EC 6.3.2.9) | −6                       | 691179       |
| 851085   | 10     | FIG 00470965: hypothetical protein | 556                       | 850529       |
| 1562460  | 10     | Putative transerase            | −48                       | 1562508      |
| 1565336  | 9      | FIG 00469667: hypothetical protein | 251                       | 1565508      |
| 1575332  | 10     | Motility accessory factor      | 167                       | 1575165      |
| 1645063  | 9      | Ferrous iron transport protein B | 1576                      | 1643487      |

**Table 2 Positions of poly-C tracts in the BFR-CA-9557 genome**

| Position | Length | RAST annotation of closest ORF | pos. rel. to ORF location | ORF location |
|----------|--------|-------------------------------|---------------------------|--------------|
| 442261   | 11     | Putative lipoprotein of feric iron transporter system | 139                       | 442254       |
| 1099189  | 10     | hypothetical protein          | 739                       | 1099187      |
| 1257487  | 9      | Phosphoglycerol transferase    | 1728                      | 1257735      |
| 1427910  | 9      | Filamentous haemagglutinin domain protein | 1554                      | 1427911      |
| 1547848  | 9      | FIG 00469527: hypothetical protein | 637                       | 1548436      |
| 1587049  | 8      | hypothetical protein          | 679                       | 1587617      |
| 1607917  | 9      | FIG 00470049: hypothetical protein | 2305                      | 1607875      |
| 1660758  | 9      | CMP-N-acetylneuraminic-beta-galactosamido-alpha-2,3-sialyltransferase (EC 2.4.99.-) | 64                        | 1661370      |
| 1668862  | 8      | CMP-N-acetylneuraminic-beta-galactosamido-alpha-2,3-sialyltransferase (EC 2.4.99.-) | 40                        | 1670442      |
depending on gene content and organization 19 different LOS classes (A-S) have been defined [54, 55]. The LOS locus of BFR-CA-9557 ranges from position 1,332,138 to 1,349,328 (17,190 bp) and contains 17 ORFs. No sialyltransferase stll/III homologue genes and no N-acetylgalactosaminyltransferase neolABC homologue genes are present in this cluster and therefore BFR-CA-9557 expresses an unsialylated LOS. It does not exactly belong to one of the LOS classes described for C. jejuni but it is closer related to the LOS classes E, P, O, and H. Remarkably, the waaM and waaV homologue genes are immediate neighbours, therefore the established sequencing approach would have been difficult due to problems resolving repetitive genomic regions [54].

**Virulence-associated genes**

Neither dimethylsulfoxide (DMSO) reductase systems nor gamma glutamyltranspeptidase (ggg) gene could be detected in the BFR-CA-9557 genome, which is typical for a clade 1 C. coli isolates, but three genes homologous to the iron transport protein TonB were present in the BFR-CA-9557 genome, which has been reported as typical for clade 2 and 3 C. coli isolates [13]. The cytolethal distending toxin (cdt) operon contains all three subunits and is therewith complete.

**Campylobacter jejuni integrated element 1-like element & CRISPR elements**

At position 1,427,993-1,467,476 we could identify a 39,483 bp sized *Campylobacter jejuni* integrated element 1 (CJIE1)-homologue prophage that shows 96 % identity (at 86 % query coverage) to CJIE1 described in *C. jejuni* RM1221 [45]. No CRISPR elements were found by any method.

| Table 3 Methylation motifs of C. coli BFR-CA-9557 |
|-----------------------------------------------|
| No.   | Motif   | Modified Position | Modification Type | % Motifs Detected | # of Motifs Detected | # of Motifs in Genome | Mean Modification Quality QV | Mean Motif Coverage | Partner Motif |
|-------|---------|--------------------|-------------------|-------------------|---------------------|----------------------|--------------------------|--------------------|------------|
| A     | RAATTY  | 3                  | m6A               | 98.67             | 27795               | 28170                | 309.40                   | 235.73             | RAATTY     |
| B     | GATC    | 2                  | m6A               | 99.66             | 7512                | 7538                 | 344.25                   | 257.15             | GATC       |
| C     | RCATC   | 3                  | m6A               | 99.43             | 4381                | 4406                 | 274.48                   | 256.54             |            |
| D     | CAAGAA  | 6                  | m6A               | 99.71             | 2069                | 2075                 | 290.35                   | 249.45             |            |
| E     | GGGTDA  | 6                  | m6A               | 99.44             | 1607                | 1616                 | 327.70                   | 254.15             |            |
| F     | DACATTGB| 4                  | m6A               | 69.69             | 223                 | 320                  | 80.60                    | 255.05             |            |
| G1    | TAAANNNTTGT| 3            | m6A               | 99.75             | 392                 | 393                  | 271.96                   | 259.91             | CACNNNNNTTGT |
| G2    | CACNNNNNTTAA | 2       | m6A               | 99.75             | 392                 | 393                  | 331.81                   | 263.02             | TAAANNNNNTTAA |
| H1    | CAAYNNNNNNNTTGY | 3       | m6A               | 99.58             | 237                 | 238                  | 316.89                   | 246.49             | CRAANNNNNNNTTGY |
| H2    | CRAANNNNNNNNNTTGG | 4        | m6A               | 99.58             | 237                 | 238                  | 276.97                   | 247.15             | CAAYNNNNNNNTTGG |

The % QV = quality value

Methylation motifs were identified using the PacBio SMRT Analysis software (see Methods). Column 2 shows the sequence consensus of the motif, whereby non-uniform positions are represented by IUPAC ambiguity codes. Column 3 and 4 denote the position of the modified base within the motif and the type of methylation. Column 5 represents the fraction of a motif's occurrences in the genome (column 7) for which a methylation has been detected (column 6). Column 8 and 9 denote the average modification quality (in Phred Q-scores) and average coverage of motifs detected as modified. The last column shows the partner motif, i.e. the reverse complement of the motif.

**Methylation motifs and RM-systems in the genome of C. coli BFR-CA-9557**

In addition to reconstruction of the genome sequence SMRT sequencing allows determination of base modification by analysis of the sequencing kinetics. Using the SMRT Analysis Modification and Motif detection, we could identify 45,881 putatively N-6-methylated adenines (m6A, ~2.7 % of genome), 1788 4-methylated cytosines (m4C, ~0.1 %) and further 53,350 rather unspecified “modified bases” where the type of modification was not recognized by the software (Additional file 5: Figure S3).

Methylated bases were arranged within eight different dominant methylation motifs (Table 3, Fig. 3). All motifs are recognized by N-6 adenine-specific methyltransferases. A small fraction (2.8 %) of m6A bases were not clustered into any of the motifs, and no consensus motif could be identified for either m4C-methylated bases or the majority of other unspecified modified bases.

The motifs G1 & G2 and H1 & H2 (Table 3) are partner motifs and motifs A and B represent palindromic sequences, all of these partners containing methylated bases on both strands. In contrast, bases of motifs C, D, E, F and W were methylated on only one strand. The major methylation motifs A, B, C, E, G, and H were almost completely methylated across their occurrences within the genome. The percentage of methylation ranges from 98.67 % to 99.75 %. In contrast, only 69.69 % of the m6A methylation motif F sequences were methylated.

Using the recognition sequence search feature within the Restriction Enzyme Database – REBASE [56], only two out of the eight motifs (GATC, RAATTY) could be matched to existing recognition sequences of restriction systems (both Type I; http://rebase.neb.com/cgi-bin/pacbioget?17032). The remaining motifs represent yet
unknown recognition sequences with the longer motifs (G and H) most likely being specific to C. coli or to this particular strain.

Using the REBASE sequence search feature and the NCBI, G2L and Prodigal/Prokka annotations four ORFs encoding subunits of Type I, seven ORFs encoding subunits of Type II, and three ORFs encoding subunits of Type IV RM-systems have been identified (Table 4). Only two of the REBASE predicted recognition sequences correspond to a specific motif detected by SMRT sequencing, namely: GAATTC/RAATTY (motif A) recognized by the DNA modification methylase (Adenine-specific methyltransferase) FokIM_2 (ORF #02605) and GATC (motif B) recognized by the DNA modification methyltransferase DpnA (ORF #2895). Additionally REBASE predicts a second candidate for a GATC-specific DNA modification methyltransferase in ORF #8910. In contrast, RAST and Prodigal predict a 16S rRNA (guanine(966)-N(2))-methyltransferase for this ORF.

The two recognition sequences GAGNNNNNGT and GGATG predicted for the Type I restriction-modification system2C DNA-methyltransferase subunit M (ORF #6540) and modification methylase FokI homologue (ORF #720), respectively, do not correspond to any motif detected by SMRT sequencing in the genome of C. coli BfR-CA-9557.

REBASE search hits to the remaining ten RM-system subunit genes have not been associated with a specific recognition sequence. Therewith, REBASE was not able to predict a corresponding RM-system subunit for the motifs C, D, E, F, G, and H. This is most likely due to the fact that besides motifs A (RAATTY) and B (GATC) none of the published C. jejuni and H. pylori methylomes contains one of these motifs [49, 50, 57].

Motif B (GATC) is present in the methylomes of H. pylori 26695 and J99-R3 [57] but not in any of the publicly available C. jejuni methylomes. In the two H. pylori genomes ORFs hp0092 and jhp0085 encoding a RM-system IIM subunit have been assigned to this motif [58, 59]. The C. coli BfR-CA-9557 homologue of hp0092 (H. pylori 26695) demonstrates a sequence identity of 76 % (553/730; Query coverage: 86 %).

![Sequence logos of eight methylation motifs. The consensus sequences of all eight motifs are depicted as sequence logo as obtained by the WebLogo 3 server (weblogo.threeplusone.com/create.cgi). The height of each stack indicates the degree of conservation (bits). The height of the letters represents the relative frequency of the base. The asterisk under a particular letter indicates the modified/methylated base. The two motifs in G and in H are partner motifs that are methylated at both strands. All motifs are recognized by N-6 adenine-specific methyltransferases.](image-url)
In contrast, the RAATTY motif (A) has not been detected in one of the two *H. pylori* methylomes, but is present in *C. jejuni* F38011, NCTC 11168 as well as 81–176 and in all three isolates a corresponding RM-system homologous to *H. pylori* hrgA has been described as major component of the *H. pylori* transformation barrier [60]. This *Mbol* homologous RM-system named HpyIII is highly conserved among *H. pylori* strains and mediates protection against *Mbol* digestion.

In *Mbol* sensitive strains the *hpyIII* gene was found absent but a homologue to *C. jejuni* cjo1602, namely hrgA, was detected. *HpyIII* negative but hrgA positive *M. pylori* strains have been associated with gastric cancer in Asian patients, while the pathogenic mechanism remains unclear [61].

Besides this role for DNA uptake the high intra-species variety in RM-system configuration and thereby the highly diverse methylation status of *H. pylori* chromosomal DNA was suggested to serve as a new

### Table 4 Putative *C. coli* BFR-CA-9557 restriction modification systems

| ORF # | Strand | Position in genome | Description | Type/ subunit | Predicted rec. seq. |
|-------|--------|-------------------|-------------|--------------|---------------------|
| 46S   | +      | 86828-89146       | *hsdR*, Type I restriction-modification system2C restriction subunit R (EC 3.1.21.3) | I/R          | -                   |
| 48S   | +      | 92205-93437       | *hsdS*, Type I restriction-modification system2C specificity subunit S (EC 3.1.21.3), EcoKI specificity protein | I/S          | -                   |
| 49S   | +      | 94709-96196       | Type I restriction-modification system2C DNA-methyltransferase subunit M (EC 2.1.1.72) | I/M          | -                   |
| 6540  | -      | 1246404-1250480    | Type I restriction-modification system2C DNA-methyltransferase subunit M (EC 2.1.1.72) / Type I restriction-modification system2C specificity subunit S (EC 3.1.21.3) | I/MS         | (GAGNNNNNGT)³⁴     |
| 720   | +      | 139123-140043     | *fokIM*₁, Modification methylase FokI, (EC 2.1.1.72), homologue to ulcer associated adenine specific DNA methyltransferase | I/M          | (GGATG)³     |
| 1470  | -      | 275768-271959     | Type IIS restriction enzyme EcoS7I | I/S          | -                   |
| 2605  | +      | 488272-489375     | DNA modification methylase (Adenine-specific methyltransferase), *fokIM*₂, Modification methylase FokI, (EC 2.1.1.72) | I/M          | GAATTC           |
| 2895  | +      | 541751-542602     | DNA modification methyltransferase, *dpmA*, Modification methylase DpnIIIB (EC 2.1.1.72) | I/M          | GATC              |
| 6520  | -      | 1240865-1243951   | N-6 adenine specific DNA methyltransferase | I/M          | -                   |
| 7730  | +      | 1464470-1465285   | DNA adenine methylase, *dpmM*, Modification methylase DpnIIIA, (EC 2.1.1.72) | I/M          | -                   |
| 8910  | +      | 1701685-1702374   | 16S rRNA (guanine(966)-N(2))-methyltransferase (EC 2.1.1.171) | I/M          | GATC⁵             |
| 2225  | +      | 409683-409970     | McrBC 5-methylcytosine-specific restriction endonuclease system2C McrB subunit2C putative | I/R          | -                   |
| 2293  | +      | 409982-411148     | McrBC 5-methylcytosine-specific restriction endonuclease system2C McrB₁ subunit2C putative | I/R          | -                   |
| 2293  | +      | 411208-412158     | McrBC 5-methylcytosine-specific restriction endonuclease system2C McrB₂ subunit2C putative | I/R          | -                   |
| 2225  | +      | 412124-413494     | McrBC 5-methylcytosine-specific restriction endonuclease system2C McrB subunit2C putative | I/R          | -                   |

1. ORF # according to the NCBI annotation pipeline for RAST and Prodigal
2. ORF # see Additional files 3 and 4
3. Disrupted ORF
4. Predicted sequence does not correspond to any motif detected by SMRT sequencing
5. According to REBASE ORF #6540 encodes a Type II RM-system
6. In contrast to RAST and Prodigal REBASE predicts a second candidate for a GATC-specific DNA modification methyltransferase for this ORF

**Methods**

The biological implications of this particular subset of RM-systems are difficult to predict, especially since the majority of methylation motifs and methylases in the genome of *C. coli* BFR-CA-9557 are quite novel. From the closely related bacterial species *H. pylori* it has been described in detail that RM-systems play a crucial role in forming strong barriers to prevent uptake of both plasmid and chromosomal DNA [60]. *H. pylori* bacterial cells are inherently very competent for DNA uptake, but this competence for DNA uptake varies significantly between specific strains. It has been observed that strains showing a very low endogenous RM-system activity demonstrate the highest transformation efficiency. Especially the presence of an RM-system homologous to *Mbol*, which has been used for isoschizomer digestion assays in our study, has been described as major component of the *H. pylori* transformation barrier [60]. This *Mbol* homologous RM-system named HpyIII is highly conserved among *H. pylori* strains and mediates protection against *Mbol* digestion.
typing system [62]. This DNA methylation based typing system may help to discriminate *H. pylori* isolates and as well isolates of related bacterial species like *C. jejuni* or *C. coli* for epidemiological and clinical purposes.

Conclusions
The objective of the study was to analyze the methylome of a clade 1 *C. coli* isolate using SMRT sequencing. The *C. coli* isolate BfR-9557 was tested positive in 5'-G**m**ATC-3' specific isoschizomer digestion assays and hence chosen for SMRT sequencing. Using one SMRT cell the complete genome of *C. coli* BfR-CA-9557 was sequenced and assembled into a single contig of 1.7 Mbp. The genome contains a CJIE1-like element prophage. SMRT Analysis Modification and Motif detection have identified eleven different dominant methylation motifs, while 14 RM-system subunits have been identified using REBASE and three different annotation pipelines. Only two of the REBASE predicted recognition sequences correspond to a specific motif detected by SMRT sequencing. The remaining 9 motifs did not correspond to any database record and are therefore characteristic for *C. coli* or at least for *C. coli* BfR-CA-9557. However, no comparable motifs have been described in the related microbial species *C. jejuni* or *H. pylori*.

Methods

**Bacterial isolates, culture conditions, DNA extraction and MLST typing**

50 *C. coli* isolates of different origin were obtained from the Federal Institute for Risk Assessment (BfR) in Berlin, Germany. The bacterial isolates were cultured on Columbia agar base (Merck) supplemented with 5 % sheep blood (BA) and incubated at 42 °C under microaerophilic conditions (5 % O₂, 10 % CO₂, 85 % N₂) for 18 hours prior to genomic DNA extraction. Genomic DNA of all *C. coli* isolates was extracted using the QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's instructions.

Species confirmation was performed using MALDI Biotyper system (Bruker Daltonics, Bremen, Germany). Results with MALDI Biotyper identification score values ≥ 2.000 were considered correct. Additionally multiplex PCR was used to discriminate between *C. jejuni* and *C. coli* [63, 64].

The MLS-type was established using amplification and sequencing primers reported before [65]. The cycling conditions were 94 °C for 1 min, followed by 35 cycles of 94 °C for 120 s, 50 °C for 60 s, 72 °C for 60 s, followed by a final elongation step of 72 °C for 5 min [65]. Amplicons of the seven genes included in the *C. jejuni/C. coli* MLST scheme were sent for sequencing to Seqlab Sequence Laboratories GmbH (Göttingen, Germany) using 10 pmol of the respective sequencing primer.

5’-G**m**ATC-3’ specific Isoschizomer digestion assay
Screening for a *C. coli* strain that methylates adenines in G**m**ATC sequences was performed using isoschizomer digestion assays in order to identify a *C. coli* strain expressing a Dam homologue. Genomic/chromosomal bacterial DNA was digested by the three restriction enzymes *MboI*, *DpnII* and *Sau3AI*. *MboI* is responsive to Dam methylation. Therefore adenosyl-methylated GATC sequences become refractory to cleavage by *MboI* but turn susceptible to *DpnII* cleavage. Additionally the restriction endonuclease *Sau3AI* that is insensitive to adenosyl-methylation by the Dam system was used as digestion control [51].

Antimicrobial susceptibility testing
*C. jejuni* BfR-CA-9557 antimicrobial susceptibility to quinolones (ciprofloxacin) macrolides (erythromycin), aminoglycosides (gentamicin, streptomycin) and tetracyclines (tetracycline) was tested using the disc diffusion method according to the standards of the European Committee on Antimicrobial Susceptibility Testing (EUCARD) [66] the microdilution assay according to the standards of the Clinical and Laboratory Standards Institute (CLSI) [67], using cut-off values set by EUCAST (www.eucast.org). Antimicrobial test discs (ciprofloxacin, erythromycin, gentamicin, tetracycline) and EUCAMP2 microtiter plates were obtained from Oxoid/Thermo-Fisher Scientific (Wesel, Germany). For MIC analysis the Sensititre semiautomated system (Trek Diagnostic Systems, Inc, Cleveland, Ohio) was used.

Library Preparation and Sequencing
Genomic DNA of *C. coli* BfR-CA-9557 (DSM 100395) was extracted using the QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's instructions. The DNA libraries have been prepared following the PacBio guidelines and sequenced on a SMRT cell using Pacific Biosciences RS sequencing technology (Pacific Biosciences, Menlo Park, USA) at Functional Genomics Center Zürich (FGCZ, Switzerland). Input genomic DNA concentration was measured using a Qubit Fluorometer dsDNA Broad Range assay (Life Technologies, Carlsbad, USA; p/n 32850). The SMRT bell was produced using the DNA Template Prep Kit 1.0 (Pacific Biosciences; p/n 100-259-100). 10 μg of gDNA were mechanically sheared to an average size distribution of 10Kbp, using a Covaris gTube (Kbiosciences, Hoddesdon, UK; p/n 520079). A Bioanalyzer 2100 12 K DNA Chip assay (Agilent Technologies, Santa Clara, USA; p/n 5067–1508) was used to assess the fragment size distribution. 5 μg of sheared gDNA were DNA damage repaired and end-repaired using polishing enzymes. A blunt end ligation reaction followed by exonuclease treatment was performed to create the SMRT bell template. A Blue Pippin device (Sage Science, Beverly, USA) was used to size select
the SMRT bell template and enrich the big fragments >8kbp. The sized selected library was quality inspected and quantified on the Agilent Bioanalyzer 12Kb DNA Chip and on a Qubit Fluorimeter.

A ready-to-sequence SMRT bell-polymerase Complex was created using the P6 DNA/Polymerase binding kit 2.0 (Pacific Biosciences, Menlo Park, USA; p/n 100-236-500) according to the manufacturer instructions.

The Pacific Biosciences RS2 instrument was programmed to load and sequence the sample on a single SMRT cell v3.0 (Pacific Biosciences p/n100-171-800), taking one movie of 120 minutes.

The MagBead loading method (PacBio, Menlo Park, USA; p/n 100-133-600) was chosen in order to improve the enrichment the longer fragments.

At the end of the run, a sequencing report was generated for every cell, via the SMRT portal. Thereby, the adapter dimer contamination, the sample loading efficiency, the obtained average read-length and the number of filtered sub-reads have been assessed.

Sequence analysis
Processing of the raw SMRT cell data was performed using the Pacific Biosciences SMRT Analysis System (version 2.3, January 2015; PacBio, Menlo Park, USA).

For de novo assembly of the *C. coli* BFR-CA-9557 genome the high-quality Hierarchical Genome Assembly Process (RS_HGAP_Assembly.2) was used with standard parameters (for details on algorithms please read https://github.com/PacificBiosciences/Bioinformatics-Training/wiki/HGAP-in-SMRT-Analysis).

The assembled genome was annotated using the rapid annotation using subsystem technology platform (RAST, http://rast.nmpdr.org) [68–70] and the Prodigal/Prokka annotation pipeline [71] implemented at Göttingen Genome Laboratory (G2L).

For identification of methylated bases and modification motifs the RS_Modification_and_Motif_Analysis.1 protocol within the SMRT Analysis System was used with standard parameters on the basis of the previous assembled genome. Putative restriction modification systems have been identified using the Restriction-ModificationFinder-1.0 server (https://cge.cbs.dtu.dk/services/Restriction-Modi-
ficationFinder-1.0/) based on the Restriction Enzyme database (REBASE, http://rebase.neb.com/rebase/rebase.html) [56].

Additionally, homologues of published *Campylobacter* and *Helicobacter* restriction modification [46–50] systems have been identified by BLAST search.

Additional checking for clustered regularly interspaced short palindromic repeats (CRISPRs) and CRISPR-associated (cas) genes was performed using CRISPR-finder (http://crispr.u-psud.fr/Server/Advanced.CRISPR-finder.php) [72].

**Ethics statement**
Ethical clearance for the analysis was obtained from Ethics Committee of the University Medical Center Göttingen, Germany. As the bacterial isolates from human donors were already part of an anonymized strain collection and no evaluation including personal patient data has been performed the Ethics Committee of the University Medical Center Göttingen waived the need for written informed consent from the donor or the next of kin.

**Availability of supporting data**
The genome is available at NCBI as *Campylobacter coli* BFR-CA-9557 with AB430 locus tags (BioProject Accession: PRJNA285481 ID: 285481; BioSample: SAMN03754337; GenBank ID: CP011777). Additionally RM-system and methylation motifs can be accessed via the index of the REBASE database (http://tools.neb.com/genomes/ or directly via this link: http://tools.neb.com/genomes/view.php?view_id=35944.

The bacterial isolate *Campylobacter coli* BFR-CA-9557 was deposited in the strain collection of the Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganis-
men und Zellkulturen GmbH, Braunschweig, Germany (German Collection of Microorganisms and cell cultures); isolate ID: DSM 100395.

**Additional files**

**Additional file 1:** Figure S1. SMRT sequencing of *Campylobacter coli* BFR-CA-9557. (A) Read length distribution of 74,742 continuous long reads (CLR) obtained from a single SMRT cell after filtering for low quality. The black line depicts the cumulated amount of bases covered by reads of a minimum size as shown on the x-axis. (B) Distribution of read quality values (1 = 100 %) for 74,742 CLRs after filtering. The black line denotes the average length of reads with a quality at least as good as indicated on the x-axis. (C) Subread length distribution of 142,135 subreads (i.e. individual fragments of CLRs). (DOC 74 kb)

**Additional file 2:** Figure S2. Reference Coverage. (A) Reference coverage (number of read base pairs per position) of the polished assembly across the contig comprising 1,720,506 bp. Average reference coverage is 50.08-fold. (B) Histogram of reference coverage across the assembled contig. (DOC 73 kb)

**Additional file 3:** RAST annotation of *C. coli* BFR-CA-9557. Annotation of the *C. coli* BFR-CA-9557 genome using the rapid annotation using subsystem technology platform (RAST, http://rast.nmpdr.org). (XLS 3605 kb)

**Additional file 4:** Prodigal annotation of *C. coli* BFR-CA-9557. Annotation of the *C. coli* BFR-CA-9557 genome using the Prodigal/Prokka annotation pipeline. (XLS 727 kb)

**Additional file 5** Figure S3. Methylated bases in *C. coli* BFR-CA-9557. (A) Scatter plot of modification quality values and per-strand coverage of 101,019 bases detected as methylated in the *C. coli* BFR-CA-9557 genome. (B) Histogram of modification quality values for all bases. (DOC 115 kb)

**Abbreviations**
CC: Clonal complex; CJE1: *Campylobacter jejuni* integrated element 1; MLST: Multi-locus sequence typing; RM-system: Restriction modification system; ST: Sequence type.
Competing interests
All authors declare no competing interests.

Authors’ contributions
AEZ and KS conceived the study idea, organized the bacterial isolate collection, and drafted the manuscript. AMG, RL, and OB performed bacterial culture, DNA isolation, isoschizomer digestion assays and MLST-PCR. TL, GSL, AT, JS and coordinated DNA sequencing and conducted bioinformatics analysis. TL, AT and JS performed annotation of the genome. UG participated in the study design and helped in drafting the manuscript. All authors assisted in drafting the manuscript and all authors read and approved the final manuscript.

Acknowledgements
The authors’ work was supported by the Deutsche Forschungsgemeinschaft (DFG GR906/13-1) and the Forschungsförderungsprogramm of the Universitätsgesellschaft Göttingen (UMG), Germany. This publication was funded by the Open Access support program of the Deutsche Forschungsgemeinschaft and the publication fund of the Georg August Universität Göttingen. We would like to thank the Functional Genomics Center in Zurich, Switzerland (FGCZ) for processing and sequencing the samples.

Author details
1Institute for Medical Microbiology, University Medical Center Göttingen, Kreuzbergstr 57, D-37075 Göttingen, Germany. 2Institute for Microbiology and Genetics, Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Georg-August-Universität Göttingen, Grisebachstr. 8, D-37077 Göttingen, Germany. 3Federal Institute for Risk Assessment (BfR), Department of Biological Safety – National Reference Laboratory for Campylobacter, D-12277 Berlin, Germany. 4Microarray and Deep-Sequencing Core Facility, University Medical Center Göttingen, Justus-von-Liebig-Weg 11, D-37077 Göttingen, Germany.

Received: 6 October 2015 Accepted: 15 December 2015
Published online: 21 December 2015

References
1. Dasti J, Tareen AM, Lugert R, Zautner AE, Gross U. Campylobacter jejuni: a brief overview on pathogenicity-associated factors and disease-mediating mechanisms. Int J Med Microbiol. 2010;300(4):205–11.
2. Zautner AE, Herrmann S, Groß U. Campylobacter jejuni - The Search for virulence-associated factors. Arch Lebensmittelhyg. 2010;61(3):91–101.
3. Allos BM. Association between Campylobacter infection and Guillain-Barre syndrome. J Infect Dis. 1997;176 Suppl 25125–28.
4. Zautner AE, Johann C, Strubel A, Busse C, Tareen AM, Masanta WO, et al. Seroreivalence of campylobacteriosis and relevant post-infectious sequelae. Eur J Clin Microbiol Infect Dis. 2014;33(6):1019–27.
5. Authority EFS. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2013. EFSA J. 2015;13(1):3991.
6. KR LL, AR S. Campylobacteriosis. http://www.cfsph.iastate.edu/FactSheets/pdfs/campylobacteriosis.pdf 2013.
7. Sheppard SK, McCarthy ND, Falsush D, Maidon MC. Convergence of Campylobacter species: implications for bacterial evolution. Science. 2008;320(5873):237–9.
8. Sheppard SK, Dallas JF, Wilson DJ, Strachan NJ, McCarthy ND, Jolley KA, et al. Evolution of an agriculture-associated disease causing Campylobacter coli clade evidence from national surveillance data in Scotland. PloS One. 2010;5(2):e9708.
9. Litup E, Tordahl M, Nielsen EM. Multilocus sequence typing performed on Campylobacter coli isolates from humans, broilers, pigs and cattle originating in Denmark. J Appl Microbiol. 2007;103(1):210–8.
10. Chen Y, Mukherjee S, Hoffmann M, Kotewicz ML, Young S, Abbott J, et al. Whole-genome sequencing of gentamicin-resistant Campylobacter coli isolated from U.S. retail meats reveals novel plasmid-mediated aminoglycoside resistance genes. Antimicrob Agents Chemother. 2013;57(11):5308–405.
11. Pearson BM, Rolney A, Crossman LC, Miller WG, Wain J, van Vliet AH. Complete genome sequence of the Campylobacter coli clinical isolate 15–537360. Genome Announc. 2013;1:16.
12. Tatusova T, Ciuffo S, Fedorov B, O’Neill K, Tolstoy I. RefSeq microbial genomes database: new representation and annotation strategy. Nucleic Acids Res. 2014;42(Database issue):D553–559.
13. Skarp-de Haan CP, Culebro A, Schott T, Rever J, Schweda BK, Hunninen ML, et al. Comparative genomics of unprogressed Campylobacter coli clades 2 and 3. BMC Genomics. 2014;15:129.
14. Fouts DE, Morgon EF, Mandrell RE, Miller WG, Rasko DA, Ravel J, et al. Major structural differences and novel potential virulence mechanisms from the genomes of multiple Campylobacter species. Plos Biol. 2005;3(1):e15.
15. Richards VP, Lefebvre T, Pavilski B, Rasalu M, et al. Comparative characterization of the virulence gene clusters (lipopolysaccharide [LOS] and capsular polysaccharide [CPS]) for Campylobacter coli, Campylobacter jejuni subsp. jejuni and related Campylobacter species. Infect Genet Evol. 2013;14:200–13.
16. Sheppard SK, Dodelot X, Kolley DA, Darling AE, Pascoe B, Meric G, et al. Progressive genome-wide introgression in agricultural Campylobacter coli. Mol Ecol. 2013;22(4):1051–64.
17. Criscuolo A, de la Blanchardere A, Coeuret S, Passet V, Sague-Rysanek V, Vernaua M, et al. Draft genome sequence of campylobacter coli strain 15PSD1-1 isolated from a patient with immunoproliferative small intestinal disease. Genome Announc. 2014;2:22.
18. Kumar R, Rao DN. Role of DNA methyltransferases in epigenetic regulation in bacteria. Subcell Biochem. 2013;61:81–102.
19. Srikhanta YN, Maguire TL, Stacey K, Grimmond SM, Jennings MP. The phasevariation: a genetic system controlling coordinated, random switching of expression of multiple genes. Proc Natl Acad Sci U S A. 2005;102(15):5047–51.
20. Reisenauer A, Kahng LS, McCollum S, Shapiro L. Bacterial DNA methylation: a cell cycle regulator? J Bacteriol. 1999;181(7):1535–9.
21. Reisenauer A, Shapiro L. DNA methylation affects the cell cycle transcription of the CtrA global regulator in Caulobacter. EMBO J. 2002;21(18):4969–77.
22. Braun RE, O’Day K, Wright A. Autoregulation of the DNA replication gene dnaA in E. coli K-12. Cell. 1985;40(1):159–60.
23. Yamaki H, Ohtsubo E, Nagai K, Maeda Y. The oriC unwinding by dam methylation in Escherichia coli. Nucleic Acids Res. 1988;16(11):5067–73.
24. Marinus MG, Casadesus J. Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. FEMS Microbiol Rev. 2009;33(3):488–503.
25. Tock MR, Dryden DT. The biology of restriction and anti-restriction. Curr Opin Microbiol. 2005;8(4):466–72.
26. Wilson GG, Murray NE. Restriction and modification systems. Annu Rev Genet. 1991;25:585–627.
27. Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, et al. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res. 2003;31(7):1805–12.
28. Roberts RJ, Mince T, Postla J, Marcelli D. REBASE-enzymes and genes for DNA restriction and modification. Nucleic Acids Res. 2007;35(Database issue):D286–290.
29. Casadesus J, Low D. Epigenic gene regulation in the bacterial world. Microbiol Mol Biol Rev. 2006;70(3):830–56.
30. Murray NE. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). Microbiol Mol Biol Rev. 2000;64(2):314–42.
31. Dryden DT, Murray NE, Rao DN. Nucleoside triphosphate-dependent restriction enzymes. Nucleic Acids Res. 2001;29(18):3728–41.
32. Vovis GF, Horuchi K, Zinder ND. Kinetics of methylation of DNA by a restriction endonuclease from Escherichia coli B. Proc Natl Acad Sci U S A. 1974;71(10):3810–3.
33. Pingoud A, Jeltsch A. Structure and function of Type II restriction endonucleases. Nucleic Acids Res. 2001;29(18):3705–27.
34. Meisel A, Bickle TA, Kruger DH, Schroeder C. Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. Nucleic Acids Res. 1992;20(16):3659–65.
35. Sutherland E, Coe L, Raleigh EA. Mrc8C: a multisubunit GTP-dependent restriction endonuclease. J Mol Biol. 1992;225(2):327–48.
36. Ahmed IH, Mannings G, Wassenaar TM, Cawthraw S, Newell DG. Identification of genetic differences between two Campylobacter jejuni strains with different colonization potentials. Microbiol. 2002;148( Pt 4):1203–12.
37. Poly F, Threadgill D, Stintzi A. Identification of Campylobacter jejuni ATCC 43431-specific genes by whole microbial genome comparisons. J Bacteriol. 2004;186(4):4781–9.
38. Poly F, Threadgill D, Stintzi A. Genomic diversity in Campylobacter jejuni: identification of C. jejuni -specific genes. J Clin Microbiol. 2005;43(5):2330–8.
Type I restriction-modification loci. Proc Natl Acad Sci U S A 2009;106(33):13692–7.

60. Ando T, Xu Q, Torres M, Kusugami K, Israel DA, Blaser MJ. Restriction-modification system differences in Helicobacter pylori are a barrier to interstrain plasmid transfer. Mol Microbiol. 2000;37(5):1052–65.

61. Ando T, Wassenaar TM, Peek JR RM, Aras RA, Tschumi AI, van Doorn LJ, et al. Helicobacter pylori restriction endonuclease-replacing gene, hrgA, is associated with gastric cancer in Asian strains. Cancer Res. 2002;62(18):4286–9.

62. Xu Q, Morgan RD, Roberts RJ, Blaser MJ. Identification of type II restriction and modification systems in Helicobacter pylori reveals their substantial diversity among strains. Proc Natl Acad Sci U S A 2000;97(17):9671–6.

63. Vandamme P, Van Doorn LJ, Al Rashid ST, Quint WG, van der Plas J, Chan VL, et al. Campylobacter pylori–Alderton et al. 1995 and sequence type 403 C. pylori reveals new evolutionary relationships. J Clin Microbiol. 2003;41(1):142–5.

64. Wright S, Wilson S, Miller WG, Mandrell RE, Siletzky RM, Kathariou S. Differences in methylation at GATC sites in genomic DNA of Campylobacter upsaliensis, Helicobacter pylori, and C. lari reveals new evolutionary relationships. J Bacteriol. 1992;174(24):8156–7.

65. Dingle KE, Colles FM, Wareing DR, Ure R, Fox AJ, Bolton FE, et al. Multilocus sequence typing system for Campylobacter jejuni. J Clin Microbiol. 2001;39(1):14–23.

66. Ge B, Wang F, Sjolund-Karlsson M, McDermott PF. Antimicrobial resistance in Campylobacter: susceptibility testing methods and resistance trends. J Microbiol Methods. 2013;95(1):57–67.

67. Jorgensen JH, Hindler JF. New consensus guidelines from the clinical and laboratory Standards Institute for antimicrobial susceptibility testing of infrequently isolated or fastidious bacteria. Clin Infect Dis. 2007;44(2):280–6.

68. Aziz RK, Bartels D, Best AA, DeLongh M, Dizd T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.

69. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Dizd T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 2014;42(Database issue):D206–214.

70. Brettin T, Davis JJ, Dizd T, Edwards RA, Gerdes S, Olsen GJ, et al. RASTk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep. 2015;5:8365.