Check for updates

RESEARCH ARTICLE

Campylobacter jejuni genomes exhibit notable GC variation within housekeeping genes [version 1; peer review: 2 not approved]

Vathsala Mohan¹, Mark Stevenson²¹AgResearch Ltd, Grasslands, Palmerston North, New Zealand²Epicentre, Institute of Veterinary and Biomedical Sciences, Massey University, Palmerston North, New Zealand**V1** First published: 14 Mar 2013, 2:89
<https://doi.org/10.12688/f1000research.2-89.v1>Latest published: 14 Mar 2013, 2:89
<https://doi.org/10.12688/f1000research.2-89.v1>**Abstract**

Campylobacter jejuni (*C. jejuni*) is a rapidly evolving bacterial species with massive genetic recombination potential to generate niche specific genotypes. Generally the housekeeping gene lineage has been evidenced to undergo lateral gene transfer and recombination quite frequently compared to the information processing gene lineage. During such exchanges, genetic amelioration takes place over time acquiring the host genomes' molecular characteristics. In this study, fifty genes that comprised twenty five metabolic housekeeping lineage genes and twenty five information processing lineage genes from nineteen *C. jejuni* genomes were studied. These nineteen genomes included seven *C. jejuni* isolates that belonged to the same genotype or multilocus sequence type ST-474 that were sequences in New Zealand. The genes from both lineages were tested for recombination and the guanine-cytosine (GC) variation. There was a positive correlation between the GC variance and the number of recombination events amongst the metabolic housekeeping genes. Genes that showed wider GC variance had a relatively high number of recombination events. In contrast, although recombination was evident in all of the informational genes, there was no correlation between the GC variance and recombination. The observation of a positive correlation between the GC variance and the recombination events in the metabolic housekeeping genes may reflect the recent events of exchange of DNA and the regions that are constantly dynamic to undergo recombination under certain circumstances. While in the case of informational genes, the demand of stringent homology between genes may be a limiting factor for the absence of such correlation, however, the sites that involved in recombination may also represent the hotspots of recombination in those genes.

Open Peer Review**Approval Status**

	1	2
version 1 14 Mar 2013	view	view

1. **David Ussery**, The Technical University of Denmark, Lyngby, Denmark
2. **Trudy M. Wassenaar**, Molecular Microbiology and Genomics Consultants, Zotzenheim, Germany

Any reports and responses or comments on the article can be found at the end of the article.

Associated Data Article

Mohan V and Stevenson M » Molecular data analysis of selected housekeeping and informational genes from nineteen *Campylobacter jejuni* genomes, F1000Research 2013, 2:89 (<https://doi.org/10.12688/f1000research.2-89.v1>)

Corresponding author: Vathsala Mohan (lingvathsala@gmail.com)

Competing interests: No competing interests were disclosed.

Grant information: This work was part of Vathsala Mohan's PhD work and this project was supported by: 1. Building Research Capability in Strategically Relevant Areas (BRCSRA) fund; 2. Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North (RM12767); and 3. The Royal Society of New Zealand Marsden Fund (MAU0802). This was an original research work and the funders had no role in any of the processes or in the preparation of the manuscript.

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

Copyright: © 2013 Mohan V and Stevenson M. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: Mohan V and Stevenson M. *Campylobacter jejuni* genomes exhibit notable GC variation within housekeeping genes [version 1; peer review: 2 not approved] F1000Research 2013, 2:89 <https://doi.org/10.12688/f1000research.2-89.v1>

First published: 14 Mar 2013, 2:89 <https://doi.org/10.12688/f1000research.2-89.v1>

Introduction

Molecular events such as mutation, deletion, recombination and gene transfer play paramount roles in shaping the evolution of prokaryotes (reviewed by^{1,2}). As a consequence the genomes are more prone for nucleotide base compositional fluctuations. Particularly, the evolutionary forces pose a major impact on the guanine-cytosine (GC) content of bacteria at the level of genes and genomes³. Amongst all the important evolutionary forces, the impact of recombination (homologous, non-homologous or illegitimate) on the evolution of bacteria has been evidenced as the major driving force or factor of microevolution^{4–8}. However, the rate of recombination may differ greatly amongst different bacterial species; while some species recombine more frequently to have multiple recombination events than mutations that render them weakly clonal, where as in other species it appears to be a rare incident leading to distinct clonal lineages^{5,6,8,9}. Studies of genetic diversity in the bacterial kingdom have shown that bacteria form clusters of genetically related strains and that extensive recombination among related clusters have been regarded as normal rather than exceptional events¹⁰. However, not every single gene is involved in recombination or horizontal gene transfer^{11,12}.

The lineages of the genes were broadly classified into informational and operational or metabolic housekeeping genes. The informational genes include genes of translation (T), transcription (S), and replication (R) and also the ATPases, GTPases (G) and tRNA synthetases whereas the operational genes are those involved in cell operations such as amino acid synthesis (A), biosynthesis of co-factors (B), cell envelope proteins (C), energy metabolism (E), intermediary metabolism (I), fatty acid and phospholipid biosynthesis (L), nucleotide biosynthesis (N), and regulatory genes (Z)¹². The operational genes are the most modular genes in the cells that are inclined to be horizontally transferred or recombined most often^{11,13}. As a result of this behaviour of the housekeeping genes it is therefore prudent to speculate that housekeeping genes may exhibit notable molecular differences compared to the informational genes.

Campylobacter jejuni (*C. jejuni*), is a zoonotic pathogen that colonises the gut of a wide variety of birds and mammals and has been attributed to the majority of bacterial gastroenteritis cases in developed countries¹⁴. Most predominantly the disease caused is by *C. jejuni* and often the disease is self limiting, however on rare occasions there can be serious sequelae such as Guillain-Barré syndrome and reactive arthritis¹⁵. The natural competency and the plasticity of *C. jejuni* were not investigated until Dingle *et al.* (2001)¹⁶ designed an MLST scheme for *C. jejuni* which has subsequently been exploited to structure and investigate the association of *C. jejuni* populations with different hosts and the environment from which human clinical infection originated^{17–21}. Further, Wilson *et al.*²² used the population genetics-phylogenetics approach to demonstrate the massive evolutionary potential of *C. jejuni* inferring that

recombination plays a major role in the generation of diversity at twice the rate of mutation *per se*.

As *C. jejuni* is an actively recombining bacterial species, an attempt to investigate the GC variations in a subset of operational genes and informational genes was made in an effort to understand the difference between these two lineages of genes within *C. jejuni* genomes. The assumptions to conduct this analysis were: (1) amelioration or coalescence of GC content or the nucleotide base composition takes relatively long time²³ which means that an event of recombination in a given group of genes with lack of time for coalescence will exhibit notable GC variation compared to other genes in the population; (2) housekeeping genes are relatively more prone for recombination where interspecies recombination has been demonstrated between donor and recipient DNA molecules that differ by up to 25–30% of their nucleotide sites^{24–26}. Given the fact that *C. jejuni* is a competent bacterial species where generation of host adapted variants has been documented using MLST datasets by several studies^{27–31}, it may be hypothesised that the housekeeping genes may exhibit notable differences in their GC content²⁴; (3) Informational genes require highly stringent homology for an event of recombination to occur and recombination takes place as a part of DNA repair³² and hence these genes may not exhibit such notable GC variation even in the presence of recombination events. In order to test these hypotheses, nineteen *C. jejuni* genomes were analysed in this study.

Methods

Analyses of metabolic and informational housekeeping genes

Housekeeping genes were located on the *C. jejuni* NCTC 11168 genome (GenBank accession number NC002163) and a total of 50 genes were selected for the analyses^{33–35}. The genes were further classified into operational genes (metabolic housekeeping genes) and informational genes based on their function by referring to the KEGG pathway and the gene function websites (http://www.genome.jp/dbget-bin/www_bget?ko+K03495 and http://www.igs.cnrs-mrs.fr/mgdb/cgi/www_gene_catalog?rpr.ann). The positions of the categorised selected subset of genes (metabolic housekeeping and informational genes) are marked on the reference *C. jejuni* NCTC 11168 (GenBank accession number NC002163) circular genome and are shown in Figures 1A and 1D, respectively. Here after in this paper the genes are referred as housekeeping genes (metabolic housekeeping genes) and informational genes for the purpose of plain comparison and interpretation. Twelve fully sequenced *C. jejuni* genomes (Table 1) were used to compare 50 selected genes. The gene sequences were downloaded from the GenBank database (<http://www.ncbi.nlm.nih.gov/genbank/>). Seven *C. jejuni* MLST ST-474 isolates were sequenced at the Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand. The overall GC and GC3 contents of individual genes were compared using DnaSP v5³⁶. The frequency distribution graphs of GC contents from all

the 19 genomes compared were generated in R programming language³⁷. Inferences on recombination within each gene under investigation were drawn using Dual-Brothers within Geneious v5.3.4³⁸ and the number of recombination events, referred to as Rm was estimated using DnaSP v5. The relationship between the GC variance and the recombination events was analysed using linear models by having Rm as a dependent variable, and the log GC variance and the length as independent variables. Detailed description of methods used in the article are available from a related data article³⁹.

Results and Discussion

GC variance in housekeeping genes

Guanine-Cytosine (GC) content ranged between 31.9% and 36.4% across the metabolic housekeeping genes in general, where

both the high and low ranges of the GC contents was evident in the *C. jejuni* subsp. *doylei* (CJJD269.97269.97) genome. The GC content of the housekeeping alleles varied among genomes with the MLST scheme alleles the *tkr* and *gltA* alleles showing a relatively wider variation followed by *glyA*, *glmM*, *aspA*, *glnA* and *uncA* alleles. GC contents were identical amongst all genes across all the ST-474 genomes except for two genes, namely *fumC* and *trpC* however, the GC3 content of the *pycA* gene varied between the ST-474 genomes. Analysis of recombination events showed that the *infB* gene possessed the highest number of 27 sites (Rm [the number of recombination events] = 27) (Table 2a) while the *trpB* and *uncA* possessed the least number of sites (n = 1) to be involved in recombination. The linear model showed that the number of recombination events was positively correlated with the GC variance, where the genes that showed a wider GC variance

Figure 1A Represents the relationship between the GC variance and the number of recombination events.

Figure 1C Represents the relationship between the GC variance and the number of recombination events in informational genes.

Figure 1B Represents the relationship between the GC variance and the length of genes.

Figure 1D Represents the relationship between the length and the number of recombination events in informational genes.

showed a high number of recombination events. However, the *infB* and *gltB* genes showed relatively high numbers of recombination events ($n = 27$ and $n = 26$, respectively) and they did not reveal convincing GC variation between genomes as it was observed in other genes. While the GC variance and the number of recombination events was found to be positively correlated (p value = 0.009), the length was not found to influence the number of recombination events significantly (p value = 0.7). Figures 1A and 1B illustrate the relationship between the GC variance and the recombination events, and the relationship between the length and the recombination events, respectively. The frequency distribution graphs of the GC contents for all the 25 genes are provided in Mohan *et al.*³⁷ in figures 2-5.

GC variance in informational genes

The overall guanine-cytosine (GC) content varied greatly between informational genes in general, where genes such as *mfd*, *ogt*, *polA*, *recJ*, *recN* and *xseA* showed lower ranges of GC contents (from 0.265 to 0.298) whereas the remainder showed a relatively higher GC content with *rpL* showing the highest GC content of 0.392. The analysis of informational genes for recombination events showed that *ogt* gene and *ssb* possessed a single recombination site, while *mutS* was found to possess the highest number of recombination events ($n = 38$) (Table 1b describes the recombination events that occurred in the informational genes). The linear model showed that Rm was dependent on the length of the genes (p value = 0.05) rather than on the GC variance (p value = 0.9). Figures 1C and 1D illustrate the relationship between the GC variance and the number of recombination events, and the relationship between the length and recombination events, respectively. (Frequency distribution graphs of GC contents for the informational genes are provided in Mohan *et al.*³⁷ in figures 6-9).

Differences in the nucleotide base composition of a gene and/or genome is a fundamental element shaping the genomic evolution which in turn, directly influences the GC contents of genes and/or genomes⁴⁰. GC variation has been thought to be driven both by neutral mutational effects and adaptive selection pressures^{41,42}. In this study, GC contents of the housekeeping and informational genes were measured across 19 *C. jejuni* genomes. Variation in the GC contents amongst the housekeeping genes investigated in this study within the 19 genomes was clearly evident. However, it should be noted that the GC contents of the housekeeping genes within the seven ST-474 genomes were identical except for two genes, the *fumC* and *trpC*, with a variation in the GC3 content of the *pycA* gene (but showed an identical overall GC content across all the ST-474s). There was an association between the GC variance and the number of recombination sites that occurred within different housekeeping genes, where the majority of the genes (investigated in this study) that possessed a wider GC variance showed a higher number of recombination sites.

The variation in base composition is a consequence of differences in the patterns of evolutionary events⁴³ where variation in the GC content is dependent on the mutational patterns and/or the evolutionary events that had occurred in a given nucleotide sequence^{43,44}. GC content has also been shown to be correlated with various biological factors⁴⁵ and is a potential research area where there has been a significant level of research carried out^{3,41,46-52} as well as is ongoing. Previous reports that investigated the causes for the differences in the GC content have evidenced that the GC content at the third codon position (GC3) and the conversion of GC to AT and AT to GC at this position to occur in favour of a selection for the GC content of any given genome^{47,53}. Further, this selection pressure possesses a great impact on evolution^{47,53}. Furthermore, the major changes in the GC content has been shown to predict the future direction of the evolutionary changes of the genomic GC content⁵⁴. The tRNA abundances in a genome has been shown to be yet another important selective pressure that determines the synonymous codon usage (change in the GC3 position) which in turn reflects the differential evolution in organisms^{50,55}. Hence, in the light of previous research reports it will be prudent to hypothesise that since the housekeeping genes are involved in metabolic processes in a genome this may be a reflection of the different metabolic evolutionary pressures that acted upon these genes as a measure of adaptation to the prevailing environmental conditions. Also, the base composition may have changed as a result of recombination between the same species and/or between bacteria with a similar base composition and tRNA pools. It is very intriguing to determine the driving forces for the change in the base composition which may help to better understand the biological reasons behind the frequent nucleotide changes within housekeeping genes in *C. jejuni*.

Further, there may be few additional explanations for the variation in the GC contents that might have arisen as a result of recombination such as (1) they may not have ameliorated or coalesced after an event of recombination; or (2) may not have had ample time for amelioration after recombination; or (3) those sites may be the hotspots or active spots on the genes that are dynamic and continuously engaging in recombination.

Reports using MLST datasets have revealed mosaic alleles within the seven housekeeping alleles and have also suggested and/or raised the possibilities of both host adaptation and convergence of *C. jejuni* and *Campylobacter coli* (*C. coli*) species^{27,30,31}. There may be an additional influence which may be relatively biological that triggers such recombination between *C. jejuni* isolates present in a host and/or convergence of *C. jejuni* with *C. coli* that may in turn enable better survival of the bacteria in certain hosts. Hybrid alleles of *tkt* and *gltA* have been frequently documented in previous reports^{30,31} where in our study,

Table 1 *Campylobacter jejuni* subsp. *jejuni* reference genomes.

Species	Strain	Size (Mb)	%GC	ORFs	Disease/source	Year and place of isolation	GenBank
<i>Campylobacter jejuni</i> subsp. <i>jejuni</i>	NCTC 11168	1.64	30.5	1643	Clinical Food poisoning	1977, UK	AL111168
<i>Campylobacter jejuni</i> subsp. <i>jejuni</i>	RM1221	1.78	30.3	1838	Chicken	2005*, USA	CP000025
<i>Campylobacter jejuni</i> subsp. <i>jejuni</i>	81-176	1.6	30.6	1653	Clinical Food poisoning	1981, USA	CP000538
<i>Campylobacter jejuni</i> subsp. <i>jejuni</i>	81116	1.63	30.5	1626	Clinical Food poisoning	2007, UK	CP000814
<i>Campylobacter jejuni</i> subsp. <i>jejuni</i>	CG8421	1.6	30.4	1512	Clinical Food poisoning	ns	ABGQ000000000
<i>Campylobacter jejuni</i> subsp. <i>jejuni</i>	HB93-13	1.7	30.6	1710	Clinical GBS	2006*, USA	AANQ000000000
<i>Campylobacter jejuni</i> subsp. <i>jejuni</i>	CG8486	1.65	30.4	1425	Clinical Food poisoning	ns	AASY000000000
<i>Campylobacter jejuni</i> subsp. <i>jejuni</i>	CF93-6	1.67	30.5	1757	Clinical MFS	2006*, Japan	AANJ000000000
<i>Campylobacter jejuni</i> subsp. <i>jejuni</i>	84-25	1.67	30.4	1748	Clinical Meningitis	2006*, USA	AANT000000000
<i>Campylobacter jejuni</i> subsp. <i>jejuni</i>	260.94	1.65	30.5	1716	Clinical GBS	ns	AANK000000000
<i>Campylobacter jejuni</i> subsp. <i>jejuni</i>	IA3902	1.64	30.5	1718	Sheep abortion	2010, USA	CP001876
<i>Campylobacter jejuni</i> subsp. <i>doylei</i>	269.97	1.85	30.6	2094	Human blood	2007, UK	CP000768
<i>Campylobacter jejuni</i> subsp. <i>Jejuni</i> -P179a	ST-474	-	-	-	Poultry	New Zealand	-
<i>Campylobacter jejuni</i> subsp. <i>Jejuni</i> -P110b	ST-474	-	-	-	Poultry	New Zealand	AEIO000000000.1
<i>Campylobacter jejuni</i> subsp. <i>Jejuni</i> -P694a	ST-474	-	-	-	Poultry	New Zealand	-
<i>Campylobacter jejuni</i> subsp. <i>Jejuni</i> -P569a	ST-474	-	-	-	Poultry	New Zealand	-
<i>Campylobacter jejuni</i> subsp. <i>Jejuni</i> -H22082	ST-474	-	-	-	Human	New Zealand	AEIP01000008.1
<i>Campylobacter jejuni</i> subsp. <i>Jejuni</i> -H704a	ST-474	-	-	-	Human	New Zealand	-
<i>Campylobacter jejuni</i> subsp. <i>Jejuni</i> -H73020	ST-474	-	-	-	Human	New Zealand	-

ST: Multilocus sequence type; GBS: Guillain-Barré syndrome; ORFs: Open reading frames; Mb: Mega bases; GC: Guanine; Cytosine; ns: not stated; *:Date of start of project.

Table 2a GC content range and recombination sites identified in the housekeeping genes.

Genes	Sites	Rm	GC range
<i>argF</i>	(114, 163) (168, 195) (195, 207) (207, 231) (231, 282) (282, 291) (291, 303) (303, 308) (308, 339) (420, 465)	10	0.331-0.351
<i>aroE</i>	(381, 392) (392, 435) (514, 645) (645, 654) (654, 702) (702, 717)	6	0.288-0.303
<i>aspA</i>	(198, 336) (336, 414) (414, 615) (627, 897) (1203, 1251)	5	0.330-0.334
<i>atpD</i>	(348, 408) (408, 651) (924, 1047)	3	0.351-0.354
<i>dapE</i>	(171, 216) (216, 230) (231, 321) (321, 351) (351, 357) (357, 387) (387, 453) (474, 522) (531, 543) (543, 560) (609, 690) (701, 867) (867, 942) (942, 945) (990, 1035)	15	0.350-0.362
<i>ftsZ</i>	(90, 138) (348, 561) (645, 681) (840, 903) (903, 936)	5	0.351-0.358
<i>fumC</i>	(45, 108) (168, 303) (354, 357) (576, 591) (624, 687) (687, 711) (772, 909) (930, 987) (987, 1038) (1038, 1227) (1281, 1311) (1311, 1335) (1335, 1380)	13	0.358-0.366
<i>gapA</i>	(288, 381) (381, 501) (501, 567) (570, 594) (594, 732) (819, 921) (948, 961)	7	0.351-0.358
<i>glnA</i>	(180, 230) (231, 258) (258, 279) (358, 448) (448, 513) (711, 768) (768, 945) (945, 1188) (1188, 1260)	9	0.319-0.325
<i>gltA</i>	(312, 325) (325, 351) (351, 372) (372, 453) (453, 654) (654, 756) (780, 921) (921, 1089) (1089, 1140) (1203, 1305)	10	0.331-0.335
<i>gltB</i>	(159, 675) (741, 855) (1236, 1401) (1401, 1581) (1581, 1605) (1605, 1902) (1974, 2067) (2151, 2178) (2226, 2325) (2349, 2412) (2412, 2418) (2418, 2700) (2700, 2838) (2853, 2982) (3204, 3210) (3210, 3318) (3387, 3462) (3462, 3636) (3651, 3861) (3885, 4023) (4023, 4092) (4092, 4098) (4098, 4119) (4119, 4149) (4149, 4305) (4324, 4356)	26	0.352-0.364
<i>glyA</i>	(165, 219) (474, 490) (492, 562) (612, 618) (618, 657) (684, 744) (966, 1156)	7	0.347-0.358
<i>groEL</i>	(1008, 1023) (1023, 1029) (1029, 1035) (1200, 1236) (1285, 1290)	5	0.328-0.336
<i>hemN</i>	(82, 174) (174, 207) (210, 372) (544, 584) (585, 783) (828, 903) (903, 927) (930, 975) (975, 996) (996, 1050) (1050, 1065) (1104, 1140) (1140, 1163)	13	0.364-0.379
<i>ilvD</i>	(249, 300) (300, 399) (492, 576) (654, 777) (777, 816) (864, 891) (1098, 1137) (1170, 1263) (1263, 1296) (1323, 1347) (1389, 1416) (1416, 1530) (1530, 1665)	13	0.30-0.306
<i>InfB</i>	(1128, 1140) (1140, 1188) (1230, 1263) (1263, 1284) (1284, 1365) (1365, 1392) (1392, 1416) (1476, 1551) (1596, 1692) (1743, 1824) (1824, 1845) (1845, 1914) (1914, 1938) (1938, 1962) (1962, 2124) (2124, 2127) (2127, 2235) (2235, 2259) (2268, 2286) (2286, 2301) (2319, 2376) (2376, 2394) (2394, 2436) (2436, 2478) (2478, 2487) (2487, 2490) (2538, 2562)	27	0.367-0.381
<i>lysA</i>	(198, 222) (222, 672) (706, 790) (790, 825) (825, 944) (945, 1035) (1035, 1122) (1122, 1171)	8	0.349-0.385
<i>nuoD</i>	(78, 93) (105, 249) (249, 267) (393, 453) (453, 483) (543, 566) (660, 678) (780, 852) (852, 864) (875, 894) (894, 954) (1008, 1050)	12	0.32-0.326
<i>glmM</i>	(171, 291) (291, 366) (366, 450) (492, 573) (573, 636) (689, 798) (897, 972) (972, 1020) (1119, 1286)	9	0.355-0.363
<i>pycA</i>	(78, 194) (194, 546) (672, 768)	3	0.33-0.34
<i>sdhA</i>	(684, 686) (686, 693) (693, 700) (700, 701) (775, 787) (787, 848) (1051, 1066) (1066, 1141) (1219, 1291) (1291, 1329)	10	0.37-0.373
<i>tkt</i>	(54, 102) (159, 237) (237, 297) (297, 342) (342, 366) (366, 660) (966, 1029) (1092, 1107) (1125, 1138) (1138, 1254) (1254, 1287) (1320, 1486) (1525, 1749) (1749, 1761) (1761, 1791) (1791, 1827) (1827, 1876)	17	0.355-0.361
<i>trpB</i>	(1053, 1119)	1	0.345-351
<i>trpC</i>	(138, 195) (210, 294) (567, 750)	3	0.315-326
<i>uncA</i>	(678, 1065)	1	0.352-0.355

Rm: Number of recombination sites.

Table 2b GC content range and recombination sites identified in the informational genes.

Genes	Sites	Rm	GC range
<i>dnaE</i>	(210, 376) (376, 396) (396, 501) (714, 762) (828, 917) (963, 981) (981, 1023) (1131, 1347) (1347, 1449) (1890, 1923) (1992, 2028) (2037, 2048) (2048, 2049) (2049, 2217) (2226, 2409) (2409, 2679) (2688, 2778) (2778, 3016) (3177, 3240) (3240, 3285) (3285, 3303) (3303, 3423) (3426, 3535)	23	0.319-0.324
<i>gidA</i>	(345, 365) (365, 399) (419, 441) (444, 546) (624, 669) (669, 687) (687, 768) (768, 771) (771, 858) (912, 930) (939, 942) (942, 955) (955, 959) (959, 1020) (1020, 1127) (1127, 1143) (1158, 1233) (1233, 1245) (1245, 1281) (1281, 1347) (1347, 1389) (1446, 1577) (1578, 1585) (1602, 1611) (1635, 1746)	25	0.329-0.341
<i>guaA</i>	(132, 174) (210, 223) (226, 232) (232, 331) (331, 351) (351, 378) (402, 471) (489, 496) (496, 516) (570, 582) (663, 687) (687, 723) (723, 765) (765, 801) (825, 858) (879, 960) (960, 1116) (1116, 1140) (1140, 1179) (1179, 1230) (1230, 1254) (1332, 1392) (1398, 1410) (1410, 1428) (1482, 1521)	25	0.347-0.361
<i>gyrA</i>	(243, 257) (257, 357) (891, 1017) (1017, 1476) (1770, 1994) (1994, 2142) (2283, 2421) (2472, 2478) (2478, 2547)	9	0.333-0.335
<i>gyrB</i>	(165, 324) (324, 360) (360, 375) (375, 813) (852, 924) (952, 1053) (1053, 1107) (1113, 1143) (1167, 1209) (1209, 1360) (1449, 1551) (1570, 1662) (1662, 1728) (2169, 2277)	14	0.329-0.332
<i>ligA</i>	(125, 188) (188, 558) (558, 579) (681, 708) (708, 720) (734, 787) (819, 832) (832, 1032) (1050, 1104) (1104, 1245) (1332, 1371) (1371, 1431) (1449, 1500) (1500, 1539) (1561, 1587) (1587, 1649) (1665, 1717) (1743, 1752) (1752, 1815) (1815, 1878)	20	0.335-0.346
<i>mfd</i>	(25, 87) (87, 174) (195, 291) (465, 531) (693, 729) (729, 867) (867, 876) (888, 919) (936, 1071) (1071, 1185) (1350, 1371) (1372, 1410) (1464, 1506) (1506, 1512) (1536, 1554) (1554, 1767) (1767, 2070) (2080, 2713)	18	0.287-0.29
<i>mutS</i>	(120, 132) (198, 210) (210, 225) (225, 316) (339, 372) (393, 421) (421, 423) (435, 474) (474, 489) (489, 591) (599, 636) (636, 729) (774, 823) (843, 849) (852, 867) (867, 876) (1026, 1028) (1028, 1107) (1134, 1243) (1243, 1266) (1269, 1305) (1305, 1326) (1339, 1365) (1497, 1548) (1548, 1596) (1596, 1626) (1626, 1680) (1710, 1869) (1869, 1914) (1975, 1978) (1978, 1983) (1984, 1996) (1996, 2014) (2071, 2077) (2077, 2089) (2095, 2116) (2152, 2185) (2185, 2203)	38	0.324-0.33
<i>mutY</i>	(75, 141) (141, 261) (294, 327) (327, 336) (336, 351) (363, 393) (438, 501) (501, 553) (553, 564) (613, 670) (670, 715) (766, 847)	12	0.308-0.318
<i>ogt</i>	-	1	0.269-0.286
<i>polA</i>	(21, 142) (240, 274) (274, 327) (369, 540) (600, 649) (649, 672) (1028, 1083) (1083, 1121) (1253, 1278) (1278, 1305) (1305, 1347) (1347, 1368) (1380, 1464) (1578, 1611) (1683, 1720) (1721, 1743) (1747, 1860) (1893, 1911) (2256, 2262) (2472, 2487) (2487, 2523)	21	0.289-0.294
<i>pyrC</i>	(5, 246) (252, 312) (312, 396) (396, 402) (402, 723) (723, 747) (747, 975) (975, 987)	8	0.304-0.312
<i>pyrG</i>	(915, 945) (945, 1018) (1146, 1392) (1419, 1445) (1445, 1479)	5	0.343-0.35
<i>recA</i>	(226, 351) (447, 570) (570, 612) (612, 624) (807, 825) (861, 912)	6	0.359-0.382
<i>recJ</i>	(250, 300) (318, 355) (355, 375) (678, 688) (688, 972) (972, 1059) (1059, 1062) (1062, 1245) (1245, 1530)	9	0.294-0.303
<i>recN</i>	(87, 111) (111, 192) (243, 396) (399, 421) (421, 539) (540, 738) (759, 834) (834, 870) (870, 879) (900, 1068) (1257, 1272) (1272, 1311)	12	0.272-0.278
<i>recR</i>	(78, 123) (282, 432) (432, 447)	3	0.3-0.309
<i>rplB</i>	(123, 228) (240, 258) (384, 444) (468, 588) (675, 762)	5	0.384-0.394
<i>rpoB</i>	(606, 1263) (1791, 1812) (3363, 3366) (3366, 3417)	4	0.333-0.337
<i>rpoD</i>	(129, 156) (156, 174) (174, 435) (495, 735) (735, 822) (822, 858) (1272, 1326) (1326, 1758)	10	0.304-0.309

Genes	Sites	Rm	GC range
<i>ruvA</i>	(96, 123) (147, 264) (264, 387)	3	0.33-0.341
<i>ssb</i>	(234, 448)	1	0.33-0.346
<i>uvrA</i>	(27, 105) (105, 132) (132, 177) (222, 243) (255, 291) (291, 294) (351, 372) (372, 384) (498, 540) (579, 621) (651, 681) (681, 804) (921, 1031) (1101, 1170) (1170, 1255) (1326, 1425) (1425, 1671) (2058, 2094) (2094, 2139) (2202, 2276) (2276, 2397) (2433, 2442) (2442, 2532) (2610, 2628) (2643, 2700) (2700, 2715)	26	0.342-0.348
<i>uvrB</i>	(36, 72) (120, 165) (165, 168) (327, 378) (420, 594) (696, 714) (714, 751) (757, 777) (777, 784) (816, 856) (876, 924) (924, 928) (936, 948) (948, 988) (988, 1005) (1042, 1068) (1068, 1074) (1083, 1086) (1086, 1122) (1122, 1182) (1182, 1221) (1221, 1248) (1248, 1266) (1272, 1284) (1284, 1356) (1362, 1383) (1437, 1530) (1530, 1533) (1533, 1534) (1536, 1593) (1701, 1776) (1779, 1785) (1785, 1791) (1791, 1866) (1881, 1890) (1914, 1938)	36	0.333-0.341
<i>xseA</i>	(90, 186) (402, 669) (832, 930) (930, 961)	4	0.281-0.288

Rm: Number of recombination sites.

these two genes showed wider GC variance within genomes. In contrast, *atpD* and *trpB* showed the least number of recombination sites where, the GC content did not vary as evidenced in the *tkt* and the *gltA* alleles across the investigated genomes. However, there was one exception, the *infB* gene that showed a high number of recombination sites showed relatively small GC variance.

In the case of informational genes, the overall GC content varied greatly amongst genes investigated within the genomes. Although there was variation in the GC contents within genes, it was not positively correlated with the number of recombination sites, whereas it was correlated with the length of genes which is not surprising. It may be speculated that since an event of recombination in informational genes demand a relatively high degree of homologies⁵⁶, recombination in this subset of genes may have occurred between sequences with high homologies. Moreover, according to the complexity theory^{11,12} the repair and ribosomal genes belong to the lineages of higher complex, which may not allow them to compromise or tolerate GC variation during recombination. Further, apart from the differences between individual genomes, the overall GC variation amongst the informational genes investigated in this study, (lower GC content in *mfd*, *ogt*, *polA*, *recJ*, *recN* and *xseA* genes and higher GC contents in the remainder) reflect the differences in the functional conservation and complexity. For example, *rpLB* showed the highest GC content of 39.2% which is relatively complex and conserved as it is a 50S ribosomal protein – L2, involved in several discrete steps of polypeptide synthesis such as peptidyl transferase activity, binding of aminoacyl-tRNA to A and P-sites⁵⁷.

Future research

Even though *C. jejuni* has been shown over the past decade through a significant amount of research to be a promiscu-

ous and a competent bacterial species, the biological triggers behind recombination which leads to host adaptation and emergence of new variants are very unclear. Further, most of the studies are based on the seven housekeeping alleles that cover the internal fragments (approximately 400–500 base pairs) of those genes used in the MLST scheme for typing the isolates^{16,58}. The nucleotide differences outside the typing region always get neglected which is important when a functional gene is to be evaluated for evolution. Housekeeping genes involved in various metabolic functions and cellular processes play a critical role in the overall integrity and survival of micro-organisms in general. The correlation between GC variance and recombination indicates the vulnerability of housekeeping genes to evolutionary forces and further it also shows how dynamic the regions on these genes are to continuously respond to such stimuli. Although base composition varies with mutation and recombination events (which is an expected biological plausibility) it is intriguing to speculate (1) how frequently the nucleotide base composition change in *C. jejuni*; (2) how much nucleotide changes can a *C. jejuni* gene and/or genome tolerate; (3) what are the biological triggers that generate new *C. jejuni* variants; (4) in the event of recombination how long does the exchanged gene portion take to coalesce; and (5) what are the functional genes mostly affected by evolutionary forces?

Even though our study has made an effort to differentiate between two lineages of genes in *C. jejuni* and to substantiate and speculate the possible triggers for GC variation and its relationship with recombination, the data we used is very small to make concrete statements. However, larger genomic datasets will be able to provide distinctive resolution on the hypotheses formulated in our study and will also provide answers to the questions that we are raising in this paper.

Author Contributions

VM did the data collection, analyses and wrote the paper in partial fulfillment of her PhD thesis. MS was VM'S supervisor and worked alongside her to get this manuscript out.

Grant Information

This work was part of Vathsala Mohan's PhD work and this project was supported by:

1. Building Research Capability in Strategically Relevant Areas (BRCSRA) fund;
2. Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North (RM12767); and
3. The Royal Society of New Zealand Marsden Fund (MAU0802).

This was an original research work and the funders had no role in any of the processes or in the preparation of the manuscript.

Competing Interests

No competing interests were disclosed.

Acknowledgements

I thank Prof. Nigel French and Dr. Patrick Biggs, mEpiLab, IVABS, Massey University, for giving their comments on this piece of work. I thank Dr. Patrick Biggs for kindly providing me with the gene predictions for my PhD work. I thank Massey University Doctoral Scholarship and New Zealand International Doctoral Scholarship for funding me. This work was carried out in mEpiLab, IVABS, Massey University, New Zealand.

References

1. Rocha EP: Evolutionary patterns in prokaryotic genomes. *Curr Opin Microbiol.* 2008; **11**: p. 454–460.
2. Feil EJ: Small change: keeping pace with microevolution. *Nat Rev Microbiol.* 2004; **2**: p. 483–495.
3. Genereux DP: Evolution of genomic GC variation. *Genome Biology.* 2002; **3**: p. 58.
4. Feil EJ, Holmes EC, Bessen DE, *et al*: Recombination within natural populations of pathogenic bacteria: Short-term empirical estimates and long-term phylogenetic consequences. *Proc Natl Acad Sci U S A.* 2001; **98**: p. 182–187.
5. Spratt BG, Hanage WP, Feil EJ: The relative contributions of recombination and point mutation to the diversification of bacterial clones. *Curr Opin Microbiol.* 2001; **4**: p. 602–606.
6. Hanage WP, Fraser C, Spratt BG: Fuzzy species among recombinogenic bacteria. *BMC Biol.* 2005; **3**: 6.
7. Fraser C, Alm EJ, Polz MF, *et al*: The Bacterial Species Challenge: making Sense of Genetic and Ecological Diversity. *Science.* 2009; **323**: p. 741–746.
8. Fraser C, Hanage WP, Spratt BG: Recombination and the nature of bacterial speciation. *Science.* 2007; **315**: p. 476–480.
9. Hanage WP, Fraser C, Spratt BG: Sequences, sequence clusters and bacterial species. *Philos Trans R Soc Lond B Biol Sci.* 2006; **361**: p. 1917–1927.
10. Giovannoni SJ, Britschgi TB, Moyer CL, *et al*: Genetic diversity in Sargasso Sea bacterioplankton. *Nature.* 2004; **345**: p. 60–63.
11. Jain R, Rivera MC, Lake JA: Horizontal gene transfer among genomes: The complexity hypothesis. *Proc Natl Acad Sci U S A.* 1999; **96**: p. 3801–3806.
12. Rivera MC, Jain R, Moore JE, *et al*: Genomic evidence for two functionally distinct gene classes. *Proc Natl Acad Sci U S A.* 1998; **95**: p. 6239–6244.
13. Ma HW, Zeng AP: Phylogenetic comparison of metabolic capacities of organisms at genome level. *Mol Phylogenet Evol.* 2004; **31**: p. 204–213.
14. Humphrey T, O'Brien S, Madsen M: Campylobacters as zoonotic pathogens: a food production perspective. *Int J Food Microbiol.* 2007; **117**: p. 237–257.
15. Zia S, Wareing D, Sutton C, *et al*: Health problems following Campylobacter jejuni enteritis in a Lancashire population. *Rheumatology.* 2003; **42**: p. 1083–1088.
16. Dingle KE, Colles FM, Wareing DR, *et al*: Multilocus sequence typing system for Campylobacter jejuni. *J Clin Microbiol.* 2001; **39**: p. 14–23.
17. Colles FM, Jones K, Harding RM, *et al*: Genetic diversity of Campylobacter jejuni isolates from farm animals and the farm environment. *Appl Environ Microbiol.* 2003; **69**: p. 7409–7413.
18. Miller WG, Englen MD, Kathariou S, *et al*: Identification of host-associated alleles by multilocus sequence typing of Campylobacter coli strains from food animals. *Microbiology.* 2006; **152**: p. 245–255.
19. Kwan PS, Barrigas M, Bolton FJ, *et al*: Molecular Epidemiology of Campylobacter jejuni Populations in Dairy Cattle, Wildlife, and the Environment in a Farmland Area. *Appl Environ Microbiol.* 2008; **74**: p. 5130–5138.
20. Carter PE, McTavish SM, Brooks HJ, *et al*: Novel Clonal Complexes with an Unknown Animal Reservoir Dominate Campylobacter jejuni Isolates from River Water in New Zealand. *Appl Environ Microbiol.* 2009; **75**: p. 6038–6046.
21. Wilson DJ, Gabriel E, Leatherbarrow AJ, *et al*: Tracing the Source of Campylobacteriosis. *PLoS Genet.* 2008; **4**: e1000203.
22. Wilson DJ, Gabriel E, Leatherbarrow AJ, *et al*: Rapid Evolution and the Importance of Recombination to the Gastroenteric Pathogen Campylobacter jejuni. *Mol Biol Evol.* 2009; **26**: p. 385–397.
23. Marri PR, Golding GB: Gene amelioration demonstrated: the journey of nascent genes in bacteria. *Genome.* 2008; **51**: p. 164–168.
24. Bowler LD, Zhang QY, Riou JY, *et al*: Interspecies recombination between the penA genes of Neisseria meningitidis and commensal Neisseria species during the emergence of penicillin resistance in N. meningitidis: natural events and laboratory simulation. *J Bacteriol.* 1994; **176**: p. 333–337.
25. Dowson CG, Hutchison A, Brannigan JA, *et al*: Horizontal transfer of penicillin binding protein genes in penicillin resistant clinical isolates of Streptococcus pneumoniae. *Proc Natl Acad Sci U S A.* 1989; **86**: p. 8842–8846.
26. Reeves P: Evolution of Salmonella O antigen variation by interspecific gene transfer on a large scale. *Trends Genet.* 1993; **9**: p. 17–22.
27. McCarthy ND, Colles FM, Dingle KE, *et al*: Host-associated genetic import in Campylobacter jejuni. *Emerg Infect Dis.* 2007; **13**: p. 267–272.
28. Meinersmann RJ, Patton CM, Evins GM, *et al*: Genetic diversity and relationships of Campylobacter species and subspecies. *Int J Syst Evol Microbiol.* 2002; **52**: p. 1789–1797.
29. Sheppard SK, Colles F, Richardson J, *et al*: Host association of Campylobacter genotypes transcends geographic variation. *Appl Environ Microbiol.* 2010; **76**: p. 5269–5277.

30. Sheppard SK, McCarthy ND, Falush D, *et al*: Convergence of *Campylobacter* Species: implications for Bacterial Evolution. *Science*. 2008; **320**: p. 237–239.
31. Sheppard SK, McCarthy ND, Jolly KA, *et al*: Introgression in the genus *Campylobacter*: Generation and spread of mosaic alleles. *Microbiology*. 2011; **157**: p. 1066–1074.
32. Gaasbeek EJ, Van der Wal FJ, Van Putten JPM, *et al*: Functional Characterization of Excision Repair and RecA-Dependent Recombinational DNA Repair in *Campylobacter jejuni*. *J Bacteriol*. 2009; **191**: p. 3785–3793.
33. Wertz JE, Goldstone C, Gordon DM, *et al*: A molecular phylogeny of enteric bacteria and implications for a bacterial species concept. *J Evol Biol*. 2003; **16**: p. 1236–1248.
34. Christensen H, Kuhnert P, Olseon JE, *et al*: Comparative phylogenies of the housekeeping genes *atpD*, *infB* and *rpoB* and the 16S rRNA gene within the Pasteurellaceae. *Int J Syst Evol Microbiol*. 2004; **54**: p. 1601–1609.
35. Margos G, Gatewood AG, Aanensen DM, *et al*: MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*. *Proc Natl Acad Sci U S A*. 2008; **24**: p. 8730–8735.
36. Librado P, Rozas J: DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009; **25**: p. 1451–1452.
37. Zhou BB, Elledge SJ: The DNA damage response: putting checkpoints in perspective. *Nature*. 2000; **408**: p. 433–439.
38. Minin VN, Dorman KS, Fang F, *et al*: Dual multiple change-point model leads to more accurate recombination detection. *Bioinformatics*. 2005; **21**: p. 3034–3042.
39. Mohan V, Stevenson M: Molecular data analysis of selected housekeeping and informational genes from nineteen *Campylobacter jejuni* genomes. [v1; ref status: not approved 1, <http://f1000r.es/um>]. *F1000 Research*. 2013; **2**: 87.
40. Hurst LD, Merchant AR: High guanine-cytosine content is not an adaptation to high temperature: a comparative analysis amongst prokaryotes. *Proc Biol Sci*. 2001; **268**: p. 493–497.
41. Musto H, Naya H, Zavala A, *et al*: Genomic GC level, optimal growth temperature, and genome size in prokaryotes. *Biochem Biophys Res Commun*. 2006; **347**: p. 1–3.
42. Foerstner KU, von Mering C, Hooper SD, *et al*: Environments shape the nucleotide composition of genomes. *EMBO Rep*. 2005; **6**: p. 1208–1213.
43. Freese E: On the evolution of base composition of DNA. *J Theor Biol*. 1962; **3**: p. 82–101.
44. Sueoka N: Correlation between base composition of deoxyribonucleic acid and amino acid composition of protein. *Proc Natl Acad Sci U S A*. 1961; **47**: p. 1141–1149.
45. Bentley SD, Parkhill J: Comparative genomic structure of prokaryotes. *Ann Rev Genet*. 2004; **38**: p. 771–792.
46. Bernardi G: Codon usage and genome composition. *J Mol Evol*. 1985; **22**: p. 363–365.
47. Hildebrand F, Meyer A, Eyre Walker A: Evidence of Selection upon Genomic GC-Content in Bacteria. *PLoS Genet*. 2010; **6**: p. e1001107.
48. McEwan CE, Gatherer D, McEwan NR: Nitrogen-fixing aerobic bacteria have higher genomic GC content than non-fixing species within the same genus. *Hereditas*. 1998; **128**: p. 173–178.
49. Naya H, Romero H, Zavala A, *et al*: Aerobiosis increases the genomic guanine plus cytosine content (GC%) in prokaryotes. *J Mol Evol*. 2002; **55**: p. 260–264.
50. Lynn DJ, Singer GA, Hickey DA: Synonymous codon usage is subject to selection in thermophilic bacteria. *Nucleic Acids Res*. 2002; **30**: p. 4272–4277.
51. Grosjean H, de Crecy-Lagard V, Marck C: Deciphering synonymous codons in the three domains of life: Co-evolution with specific tRNA modification enzymes. *FEBS Lett*. 2010; **584**: p. 252–264.
52. Ikemura T: Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol*. 1985; **2**: p. 13–34.
53. Akashi H: Inferring weak selection from patterns of polymorphism and divergence at 'silent' sites in *Drosophila* DNA. *Genetics*. 1995; **139**: p. 1067–1076.
54. Bellgard MI, Gojobori T: Significant differences between the G+C content of synonymous codons in orthologous genes and the genomic G+C content. *Gene*. 1999; **238**: p. 33–37.
55. Tuller T, Grishovich Y, Sella Y, *et al*: Association between translation efficiency and horizontal gene transfer within microbial communities. *Nucleic Acids Res*. 2011; **39**: p. 4743–4755.
56. Lovett ST, Hurley RL, Sutter VA, *et al*: Crossing over between regions of limited homology in *Escherichia coli*. RecA dependent and RecA-independent pathways. *Genetics*. 2002; **160**: p. 851–859.
57. Mikulik K, Man P, Halada P: Characterization of the *rplB* Gene from *Streptomyces collinus* and Its Protein Product by Mass Spectrometry. *Biochem Biophys Res Commun*. 2001; **285**: p. 1344–1349.
58. Maiden MC: Multilocus Sequence Typing of Bacteria. *Ann Rev Genet*. 2006; **60**: p. 561–588.

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 07 May 2013

<https://doi.org/10.5256/f1000research.1103.r932>

© 2013 Wassenaar T. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Trudy M. Wassenaar

Molecular Microbiology and Genomics Consultants, Zotzenheim, Germany

This paper compared 50 genes, obtained from 12 *Campylobacter* genomes, to test two hypotheses. I rephrase them for clarity here as: (1) Does GC content provide evidence of recombination events, and can this shed light on the history of recombinational events in the *Campylobacter* genome? (2) Are housekeeping genes more prone for recombination than informational genes in *Campylobacter*? From the 'Results and discussion' section it is not clear whether the authors have confirmed or rejected these hypotheses.

I have difficulties with both of these hypotheses. My concerns with the first hypothesis are the following: The authors consider only recombination between donor DNA with a higher GC-content into an AT-rich *Campylobacter* genome. They ignore that the vast majority of recombination can occur within a *Campylobacter* species, or, to a lesser extent, between closely related species (*C. jejuni* and *C. coli*), in which case there is no base composition difference between donor and acceptor and you wouldn't see amelioration of GC content. The 'recombination' that the authors have identified and use in their analysis (they even mention recombination sites although it is unclear how these were identified) have more likely taken place between alleles of different *Campylobacter* clones than between genes of different species.

My concern with the second hypothesis is that there is no fundamental genetic or physiological difference between 'housekeeping genes' and 'informational genes'. It is a man-made division only, while the cell just maintains its physiology by means of all these proteins. The only relevant distinction here is whether a gene product is active as a sole contributor to a process (an enzyme, say, that works on a single substrate-to-product conversion without interaction with other factors) or whether a gene product acts in close contact with many other gene products (a ribosomal protein, say). The latter are constrained in mutation, but only for non-synonymous mutations.

The authors ignore codon usage effects, expression levels (highly expressed genes employ different codons), and compositional constraints of proteins that are reflected by their codons and thus affect the gene's GC content.

A recent paper studied evolution in the complete core genome of 27 *Campylobacter* genomes. It covered more than 1100 genes and employed various statistical methods. That paper reported that there was NO division in evolutionary signature between informational versus housekeeping genes (Snipen *et al.*, (2012) *Analysis of evolutionary patterns of genes in Campylobacter jejuni and C. coli*. *Microb. Inform. Expt.* 2:8).

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 02 May 2013

<https://doi.org/10.5256/f1000research.1103.r922>

© 2013 Ussery D. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

David Ussery

Center for Biological Sequence Analysis, Department of Systems Biology, The Technical University of Denmark, Lyngby, Denmark

This short article examines variation in GC content and recombination in two different sets of genes, across nineteen *Campylobacter jejuni* genomes. After reading the abstract several times, I'm still not sure exactly what hypothesis is being tested here. I am confused by the first sentence in the abstract, which states that *C. jejuni* is 'rapidly evolving' and has 'massive genetic recombination potential'. Compared to what? Certainly compared to a virus, *C. jejuni* is quite slowly evolving. Further, it seems from the larger picture of whole genome comparison, there is not THAT much difference within the *C. jejuni* genomes, compared to say for example *E. coli*, which can be three to four times the size of *C. jejuni*, and has a very large pan-genome - about ten times the size of any individual *E. coli* genome. Several years ago, based on a smaller set of genomes, we found that the *C. jejuni* genome was much less 'open' than the *E. coli* genome. (see PMID 19691844).

I had to read and re-read the first paragraph of the Introduction section many times. Going through just that first paragraph:

"Molecular events such as mutation, deletion, recombination and gene transfer play paramount roles in shaping the evolution of prokaryotes (reviewed by^{1,2})."

I'm curious as to what exactly a 'molecular event' is? I would think that most molecular biologists are 'atomists' that is, they think in terms of biochemistry. Is there an alternative, perhaps vitalism or some supernatural event that is an alternative to a 'molecular event'. What is the difference

between 'mutation' and 'deletion'? Would a 'deletion' not be considered a subset of 'mutation'? And what is meant by 'recombination'? Is this different to 'mutation'?

"As a consequence the genomes are more prone for nucleotide base compositional fluctuations."

The [prokaryotes] are MORE prone than what? Viruses? Eukaryotes? Not sure what is being referred to here. Are the authors saying that mutations, deletions, recombination, gene transfer happen more often in bacteria than in eukaryotes? Or PERHAPS the fact that, because bacterial genomes tend to be more coding-rich (80% or more of the genome codes for proteins), then variations in the third codon position might allow genomes to become more AT rich or GC rich??

"Particularly, the evolutionary forces pose a major impact on the guanine-cytosine (GC) content of bacteria at the level of genes and genomes³.

Amongst all the important evolutionary forces, the impact of recombination (homologous, non-homologous or illegitimate) on the evolution of bacteria has been evidenced as the major driving force or factor of microevolution⁴⁻⁸."

So recombination is driving the changes in GC content? How does that work exactly? Or perhaps the authors here are talking about something other than differences in GC content? If so, this could be elaborated.

"However, the rate of recombination may differ greatly amongst different bacterial species; while some species recombine more frequently to have multiple recombination events than mutations that render them weakly clonal, whereas in other species it appears to be a rare incident leading to distinct clonal lineages^{5,6,8,9}."

And where does *Campylobacter* fit on this scale? Is there more variation, more recombination, or less? And, once again, perhaps it would be nice for the authors to mention that *Campylobacter* is very AT rich.

"Studies of genetic diversity in the bacterial kingdom have shown that bacteria form clusters of genetically related strains and that extensive recombination among related clusters have been regarded as normal rather than exceptional events¹⁰.

However, not every single gene is involved in recombination or horizontal gene transfer^{11,12}."

This is comforting perhaps, but an indication of the fraction of genes undergoing recombination in *Campylobacter* would be good...

In the Methods section, the NCBI refseq number is given, rather than a true 'GenBank accession number', which for the *C. jejuni* genome is AL111168. The methods section states that the genomes were downloaded from GenBank - but it appears that perhaps NCBI RefSeq was used instead. RefSeq is a somewhat curated database, with different gene annotations (and hence there could be different gene lengths which might give different %GC contents) - so it is very important to clearly state WHICH database was actually used here. The INSDC accession numbers are shared between GenBank, EMBL, and the DNA DataBase of Japan (DDBJ). To be honest, I really do not see that large of a difference between the variation in %GC in the housekeeping genes vs. information genes. I think a simple box and whiskers plot, showing the genome distribution, compared to the distribution for the housekeeping vs. information genes is the best way to visualise and compare the three distributions. I suspect that there's not that much of a difference here. (Which by the way, I looked through the manuscript several times - nowhere could I find a mention or even brief discussion that *C. jejuni* is quite AT rich, compared to other bacterial genomes).

The conclusion, that the information genes have less recombination than the housekeeping genes is perhaps not surprising, but I'm not sure how 'recombination' is measured here - the authors refer to a computer program which was used to measure recombination in HIV sequences, with high rates of changes. However, applying this model to something like *Campylobacter*, where the variation is extremely low (compared to HIV certainly). So if there is a single nucleotide difference, is this called a 'recombination event'? Are SNPs recombination events? Maybe sometimes?

In summary, I find this short paper would reflect a nice student project, done as part of a course. It is a good exercise to write up what has been done, but I'm still missing the hypothesis that is being tested, the anticipated results, and what sorts of results would falsify the hypothesis. Why not come up with a very clear hypothesis that can be tested, and run it across the roughly 80 *C. jejuni* genomes available from NCBI? <http://www.ncbi.nlm.nih.gov/genome/149>

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research