



Visualization and Significance of DNA Structural Motifs in the *Campylobacter jejuni* Genome

Lise Petersen,^{1,2} Stephen L. W. On,² and David W. Ussery^{1,*}

¹Center for Biological Sequence Analysis, Technical University of Denmark, DK-2800 Lyngby, Denmark

²Department of Bacteriology, Danish Veterinary Laboratory, DK-1790 Copenhagen, Denmark

(Received: 15 October 2001; revised/accepted: 31 January 2002)

ABSTRACT: The genome sequence of *Campylobacter jejuni* NCTC11168 was analyzed in terms of DNA structural properties (intrinsic curvature, base-stacking energy, and DNA flexibility) throughout the chromosome. In addition, we calculated the frequency of DNA repeats in *C. jejuni* and in chromosomes from 25 other species of the class *Proteobacteria*. Compared with the average, global repeats are under-represented in *C. jejuni*, whereas local repeats are overrepresented and the percentage of palindromes exceeds that of most other pathogenic bacteria. These palindromic regions are localized to regions of known hypervariability. Furthermore, the proportion of purine and pyrimidine/purine stretches observed in both *C. jejuni* and *H. pylori* (ϵ -subdivision *Proteobacteria*) is significantly greater than that in the other *Proteobacteria* subdivisions and correlates with the high potential for recombination observed for these two species. The first half of the *C. jejuni* genome (e.g., in the direction of the sense strand of the sequence file) contains three RNA operons. In contrast, the other half of the genome contains several regions coding for surface structure genes, including lipo-oligosaccharide (LOS), surface polysaccharide and capsular antigens, and flagellin and flagellar modification. The structural parameters of these well-characterized genes and gene clusters correlate with current knowledge of their expression, stability, and recombination potential. We conclude that visualization of DNA structural parameters in the genome sequence of *C. jejuni* is a useful tool for identifying important characteristics of genes that correlate with functionality.

Keywords: *Campylobacter jejuni*, Genome Sequence, DNA Structural Motifs, *Proteobacteria*, Chromosomes.

1. INTRODUCTION

Campylobacter jejuni subsp. *jejuni* (hereafter *C. jejuni*) is an important human intestinal pathogen, with an estimated incidence in several developed countries exceeding 800 reported cases per 100,000 inhabitants [1]. Despite intensive study, there have been few advances in the development of control and intervention strategies; the incidence of human infection has increased for a decade [1]. A better understanding of gene organization, function, and regulation in *C. jejuni* could provide new insights into its fundamental biology, which could subsequently be exploited as the basis for rational control strategies.

The complete genome sequence of one *C. jejuni* isolate from human diarrhea (NCTC11168) has been determined [2]. The substantial informational content of any entire genome clearly requires the application of bioinformatics tools to improve our understanding of its function. In this study, we analyze the genome sequence of *C. jejuni* strain NCTC11168 with bioinformatics tools that calculate different DNA structural parameters (such as intrinsic curvature, base-stacking energy, and position preference) and the occurrence of global and local repeats and palindromic sequences. These parameters are visualized for the entire chromosome as a circular plot known as the GenomeAtlas [3, 4] and correlated with experimentally determined properties in well-characterized genes or gene clusters. In addition, we compare repeat sequence data with corresponding properties in whole chromosomes of other

*Author to whom correspondence should be addressed.

Table 1. *Proteobacteria* genomes included in this study.

Organism	Strain	Division	Chromosome	Length	% AT	Accession no. ^a	Ref.
<i>Campylobacter jejuni</i>	NCTC11168	ϵ	Main	1,641,481	69.5	NC_002163	[2]
<i>Helicobacter pylori</i>	J99	ϵ	Main	1,643,831	60.8	AE001439	[36]
<i>Desulfovibrio vulgaris</i>	Hildenborough	δ	1	3,571,425	36.9	Sanger Centre	See legend
<i>Buchnera sp.</i>	APS	γ	Main	640,681	73.7	AP000398	[37]
<i>Escherichia coli</i> O157:H7	RIMD0509952	γ	Main	5,498,450	49.2	BA000007	[38]
<i>Haemophilus influenzae</i>	Rd	γ	Main	1,830,138	61.9	L4203	[39]
<i>Pasteurella multocida</i>	Pm70	γ	Main	2,257,487	59.6	AE004439	[40]
<i>Pseudomonas putida</i>	KT2440	γ	Main	6,184,039	38.5	TIGR	See legend
<i>Pseudomonas aeruginosa</i>	PAO1	γ	Main	6,264,403	33.4	AE004439	[41]
<i>Salmonella enterica</i>	Serovar Typhimurium LT2	γ	Main	4,857,432	47.8	AE006468	[42]
<i>Salmonella enterica</i>	Serovar Typhi CT18	γ	Main	4,806,104	47.9	NC_002305	[43]
<i>Vibrio cholera</i>	N16961	γ	1	4,653,728	52.3	AE003852	[44]
<i>Xylella fastidios</i>	9a5c	γ	Main	2,679,306	47.3	NC_002488	[45]
<i>Yersinia pestis</i>	CO-92BiovarOrientalis	γ	Main	2,961,149	52.4	AL590842	[46]
<i>Bordetella pertussis</i>	Tomaha I	β	Main	4,086,186	32.3	Sanger Centre	See legend
<i>Neisseria gonorrhoeae</i>	FA1090	β	Main	2,146,877	47.3	U. Oklahoma	See legend
<i>Neisseria meningitidis</i>	Serotype A, strain Z2491	β	Main	2,184,406	48.2	NC_002203	[47]
<i>Ralstonia solanacearum</i>	GMI1000	β	1	3,716,413	33.0	AL646052	See legend
<i>Agrobacterium tumefaciens</i>	C58	α	1	2,841,581	40.6	AE007869	[48]
<i>Brucella melitensis</i>	16M	α	1	2,117,144	42.8	AE008917	[49]
<i>Caulobacter crescentus</i>	CB15	α	Main	4,016,947	32.8	AE005673	[50]
<i>Mershobium loti</i>	MAFF303099	α	Main	7,036,074	37.3	BA000012	[51]
<i>Rhodopseudomonas palustris</i>	CGA009	α	Main	5,459,213	35.0	NC_002804	See legend
<i>Rickettsia conorii</i>	Malish 7	α	Main	1,268,755	67.6	AE006914	[52]
<i>Rickettsia prowazekii</i>	Madrid E	α	Main	1,111,523	71.0	AJ235269	[53]
<i>Sinorhizobium meliloti</i>	Rm1021	α	Main	3,654,135	37.3	AL591688	See legend
Average	—	—	—	3,482,129	47.3	—	—

An online version of this table can be found at www.cbs.dtu.dk/services/GenomeAtlas/Campy/index.html.

^aSources of unpublished sequence data.

Bordetella pertussis Tomaha I: Data were produced by the *Bordetella* Sequencing Group at the Sanger Institute and can be obtained from their ftp site (ftp.sanger.ac.uk/pub/pathogens/bp/).

Desulfovibrio vulgaris and *Pseudomonas putida*: Preliminary sequence data for these genomes were obtained from the Institute for Genomic Research (www.tigr.org).

Neisseria gonorrhoeae: Sequence data were downloaded from the University of Oklahoma sequencing site (www.genome.ou.edu/gono.html).

Ralstonia solanacearum: sequence data were downloaded from GenBank (AL646057).

Rhodopseudomonas palustris: sequence data were downloaded from the U.S. Department of Energy's Joint Genome Institute's web page (http://www.jgi.doe.gov/JGI_microbial/html/rhodopseudomonas/rhodops_hom_epage.html).

species belonging to the class *Proteobacteria*, a phylogenetically distinct but extremely diverse group of bacteria [5], of which *Campylobacter* and *Helicobacter* represent the epsilon subdivision [6], to determine broad similarities or differences between taxa.

2. METHODS

Complete genome sequences of *C. jejuni* NCTC11168 and other organisms were downloaded from GenBank or from the appropriate web page of the Genome sequencing center (Table 1). A complete list of all *Proteobacteria* genomes used in this study, pertinent URLs, calculated repeats, and other information can be found on a supplemental web page (see <http://www.cbs.dtu.dk/services/GenomeAtlas/Campy/>).

2.1. Structural Parameters

Three different types of structural characteristics were predicted, with the use of previously described models: DNA

curvature [7, 8], DNA stacking energy [9], and DNA flexibility [10, 11]. Briefly, the structural values are assigned to every nucleotide in a DNA sequence by looking up the values for corresponding di- or trinucleotides in a table; in the case of curvature, the value for a 21-bp window was calculated. The level of each parameter was depicted on the circles of the DNA Atlases as differences in color and color intensity.

2.2. Purine and pyr/pur Stretches

The fraction of purine (R) or alternating pyrimidine/purine (YR) stretches was calculated for *C. jejuni* and other genomes, using a PERL script which searched for regular expressions in the sequence, as described in more detail elsewhere [12]. A minimum length of 10 bp was chosen because this reflects one turn of the DNA helix.

Table 2. Calculating different types of repeats.

Type of repeat	Similarity level	Repeat length R (bp)	Window size W (bp)
Local repeats: two copies of a sequence that can be on either the same (direct) or opposite strands (inverted), and pointing in either the same or opposite directions	80%	15	100
Simple repeats: a region consisting only of a repeat oligonucleotide	80%	12, 14, 15, 16, 18	$2R$
Perfect palindromes: two copies of a sequence located on opposite strands	90%	7	30
Quasi-palindromes: two less exact copies of a sequence located on opposite strands	80%	7	30
Global repeats: two copies of a sequence that can be located on the same strand (global direct repeats) or opposite strand (global inverted repeats)	80%	100	Entire chromosome

2.3. Repeat Sequences

Repeats are multiple copies of the same sequence at different locations on a piece of DNA. Direct and inverted repeats were both calculated with the same basic algorithm [3], which finds the highest degree of homology for an R bp repeat within a window of length W (see Table 2). Similarly, palindromic sequences likely to form cruciform structures [13] were considered in this study and were located by application of the computational parameters described in Table 2. Global repeats for the 26 chromosomes shown in Figures 1 and 2 and in Table 1 were calculated with an alternative and faster approach, where the chromosomal DNA sequence was compared against itself with a BLAST search, as described elsewhere [14].

3. RESULTS AND DISCUSSION

3.1. Comparison of Repeats in *C. jejuni* with other *Proteobacteria* Chromosomes

The characteristics of the *C. jejuni* chromosomes differ from those of *Proteobacteria* chromosomes in several respects, as shown in Figures 1 and 2. Global repeats are less frequent in *C. jejuni* compared with the average of the other genomes (note in particular the very low levels of global inverted repeats; see Fig. 1a). This is consistent with the finding that global repeats are uncommon in *C. jejuni* [2], although we believe our results are the first to quantify the level of repeats in this and other chromosomes of *Proteobacteria* in a well-defined manner.

In contrast to the low level of repeats at the global level, we find that local repeats are quite abundant in *C. jejuni* (Fig. 1b), with approximately 15% of the genome containing local repeats of 80% or greater match. The proportion of local repeats is higher in *C. jejuni* than in most other *Proteobacteria* genomes, including *H. pylori*. Only the insect symbiont *Buchnera* sp. APS shows a higher level of local repeats. Thus, the amount of repeats in *C. jejuni*, compared with other genomes, depends strongly on the window size used, and the level of global repeats is low, but local repeats are abundant. The abundance of local repeats in *C. jejuni* is, in part, a reflection of the high AT content; as

the AT (or GC) content approaches 100%, the DNA alphabet effectively becomes reduced from a 4-letter alphabet to a 2-letter alphabet, with the chance of finding the same base increasing from 25% to 50%. This is illustrated in Figure 2a, where the level of palindromes is about 5% of the chromosome for both *C. jejuni* (69% AT) and *Bordetella pertussis* (32% AT, or 68% GC), and 7.5% for *Buchnera* sp. APS (74% AT).

Palindromes are quite abundant in the *C. jejuni* genome compared with other *Proteobacteria* chromosomes, as can be seen in Figure 2a. As in the case of local repeats, *C. jejuni* contains the second largest fraction of palindromes. This may be significant since such sequences can fold back on themselves and form stem-loop or cruciform structures.

The occurrence of purine and pyr/pur tracts in *Proteobacteria* chromosomes averages 1.15% and 1.47%, respectively, which is higher than what is expected by random chance (1.0%) [12]. However, the corresponding values for the *C. jejuni* and *H. pylori* genomes (ϵ -subdivision *Proteobacteria*) are 2.69% and 2.76% (purine stretches), and just 0.2% and 0.3% (pyr/pur stretches) (Fig. 2b). Although the high AT content of *C. jejuni* (69%) may partly account for the higher occurrence of repeats and palindromes as compared with other organisms, this cannot explain the elevated level of purine stretches in *C. jejuni*, or the rather low level of pyr/pur stretches, since the chance of finding a purine is still one-half, whether a sequence is 50% AT or 100% AT.

3.2. Biological Implications of Repeats

The frequent occurrence of local repeat elements within the *C. jejuni* genome may be indicative of an elevated level of recombination on this genome compared with other *Proteobacteria* genomes. Although bias in AT/GC content could easily be the explanation for the abundance of the repeats, the fact remains that a significant fraction of the genome contains local repeats. The biological result is a much greater chance of mutation, in terms of replication-induced errors (such as slipped-strand mispairing), quasi-palindrome correction (whereby an imperfect palindrome is

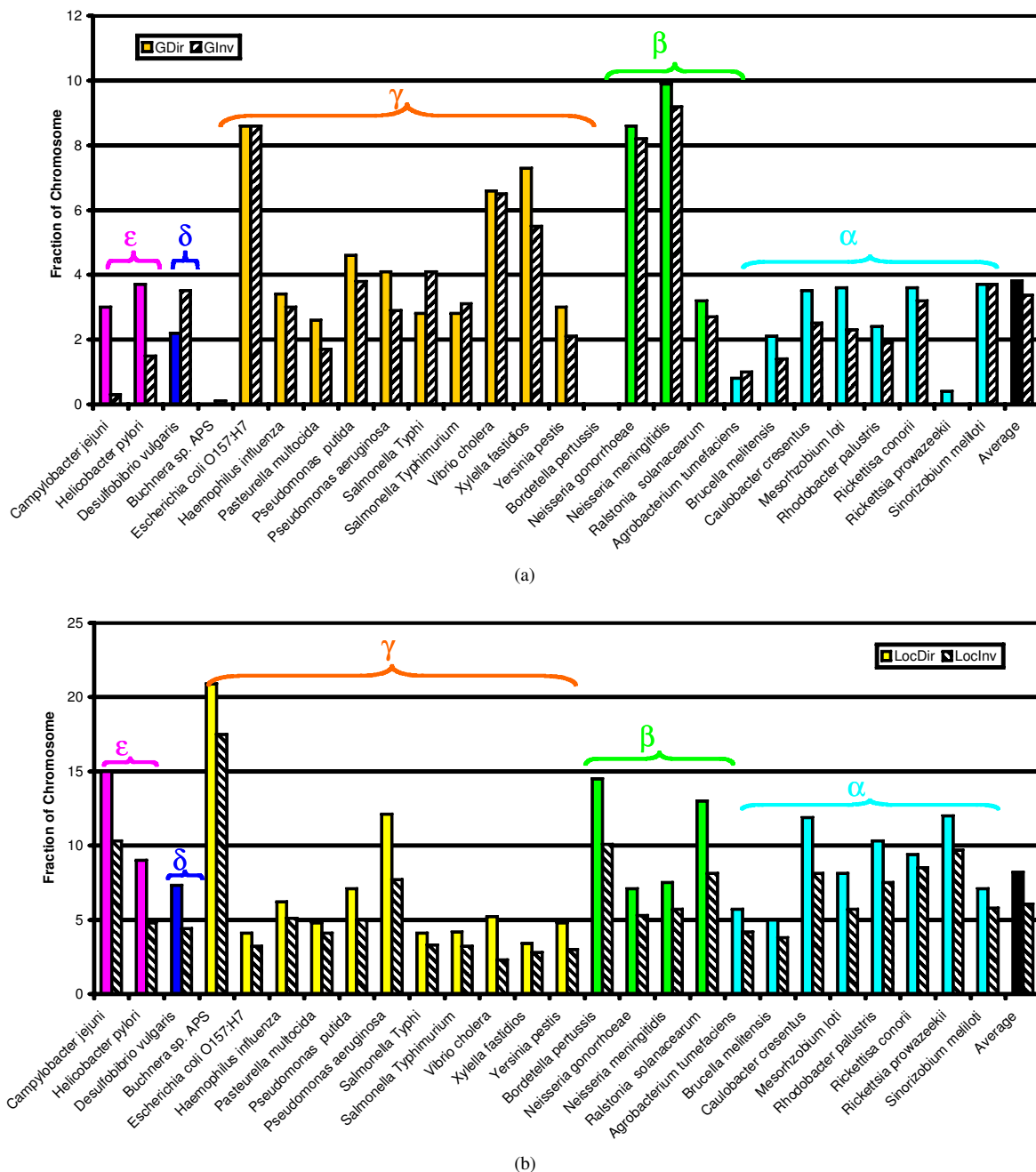


Figure 1. Repeat sequences in sequenced *Proteobacteria* genomes. (a) Global repeats (match > 80%). (b) Local repeats (match > 80%). The repeats are calculated as described in the text and in Table 2. Legends on the x axis refer to bacterial strains (Table 1). *Proteobacteria* subdivisions are shown with brackets on each panel.

altered to produce a perfect palindrome [15]), and homologous recombination. In addition, certain (often G-rich) purine stretches have the potential to form triplex DNA, and it has been hypothesized that triple-stranded DNA has an increased potential to undergo recombination, via a triple-stranded intermediate [16]. It is well established that *C. jejuni* and *H. pylori* are organisms that show substan-

tial variation in genomic organization [17, 18], with a high degree of geno- and phenotypic diversity [19–21]. Results from multilocus sequence typing analyses suggest that the uptake and integration of foreign DNA is an important factor driving genetic change and results in weakly clonal population structures [22–24], whereas slipped-strand mispairing has been observed in key regions of the *C. jejuni*

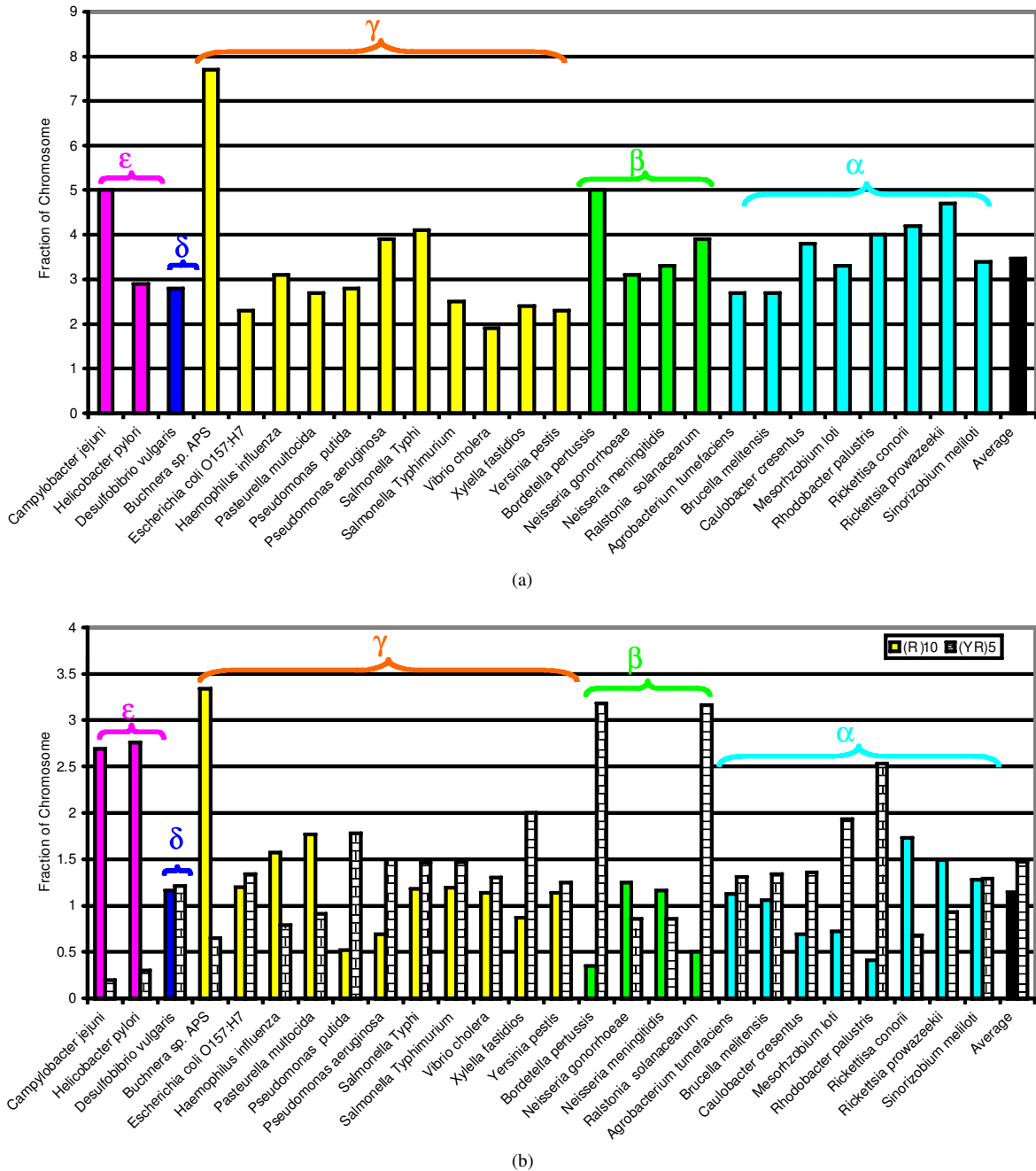


Figure 2. Structural properties in sequenced *Proteobacteria* genomes. (a) Palindromes (match > 90%); the fraction of palindromes is calculated as described in Table 2. (b) Purine or pyr/purine stretches, as defined in the text; the fraction of purine and pyr/pur stretches is as described in Experimental Details. Legends on the x axis refer to bacterial strains as indicated in Table 1. *Proteobacteria* subdivisions are shown with brackets in each panel.

genome [2]. The greater proportion of local repeats in *C. jejuni* compared with *H. pylori* is somewhat anomalous in the population structures of these bacteria since *H. pylori* seems to be panmictic [20], although weakly clonal groups have been observed in this species [22]. It is evident that mechanisms leading to recombination are complex, and even though the presence of repeat elements is important

in this context, other hitherto uncharacterized factors may play a significant role in some bacteria.

3.3. Visualization of Repeats and Structural Motifs in the *C. jejuni* Genome

Although determination and comparison of frequency of repeat sequences offer an insight into the mutational potential of bacteria, it is also important to localize these struc-

tures within the chromosome—are they homogeneously distributed throughout, or are they spatially localized into small regions of functional significance? Figure 3 shows a “GenomeAtlas” for the *C. jejuni* genome. The figure shows distinct structural differences between chromosomal regions. The innermost circle (purple/turquoise) shows the GC skew [25], which readily shows the origin of replication and two replicores. The next two inner circles in Figure 3 are the global repeats, with the solid blue circle representing the location of global Direct Repeats, and the solid red circle indicating global Inverted Repeats. There are many more direct repeats than inverted repeats (see also Fig. 1a), and the major regions (e.g., dark blue bands) correspond to the rRNA genes and a few other genes, including the *fla* locus. The next circle shows the location and direction of annotated genes, and it is obvious that there is a bias toward genes in the forward direction of the leading strand in the right half of the genome, and vice versa. The three out-

ermost circles show the position preference, base-stacking energy, and intrinsic curvature, respectively. Similar to the other circles, extreme values are indicated with a darker color, and it is evident that there are localized regions of extreme values.

The replicore on the right half of the genome (e.g., in the direction of the sense strand of the sequence file) contains three rRNA operons that are readily identified by their low position preference and highly negative stacking energy (Fig. 3). It has been hypothesized that genes with low position preference have the potential to be highly expressed [4], and preliminary results from microarray-based expression data from *Escherichia coli* support this hypothesis (D. Ussery, unpublished data). A highly negative stacking energy is indicative of more thermodynamically stable (usually GC-rich) regions and likely a lower mutation rate [26]. The high expression potential and low mutation fre-

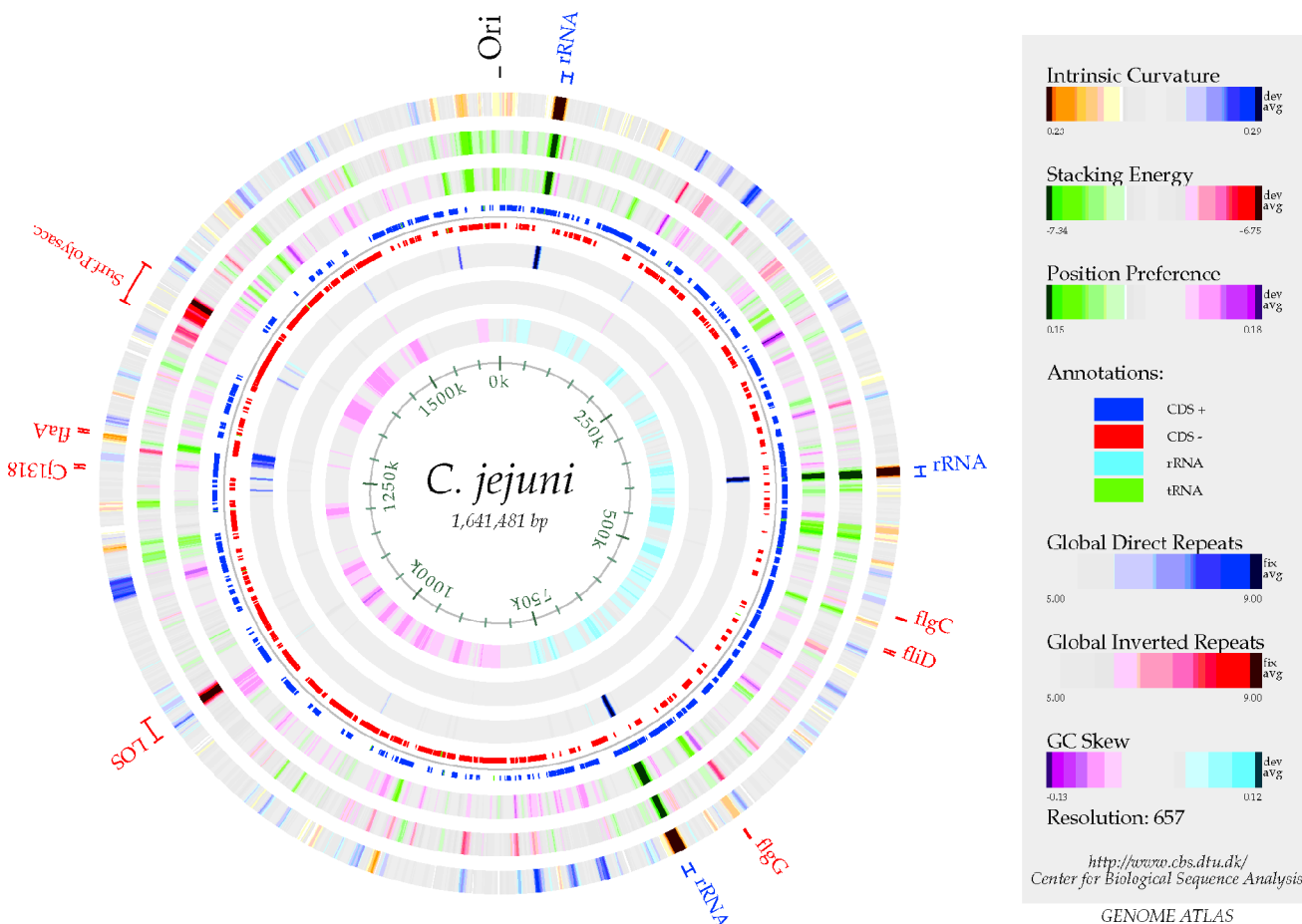


Figure 3. GenomeAtlas showing different structural parameters of the genome of *C. jejuni* NCTC11168. Legends: Ori, origin of replication; rRNA, regions containing genes encoding ribosomal RNA, located at the following positions in the sequence 39,249–44,860 nt, 394,130–399,763 nt, 696,424–702,058 nt; *flgC*, probable flagellar basal-body rod protein, 489,458–489,952 nt; *fliD*, probable flagellar hook-associated protein, 510,552–512,480 nt; *flgG*, probable flagellar basal-body rod protein, 656,043–656,834 nt; LOS, genes encoding lipo-oligosaccharide synthesis, 1,067,817–1,082,408 nt; Cj1318, 1,246,678–1,248,627 nt; *flaA*, flagellin A, 1,269,232–1,270,950 nt; Surf. Polysacc., genes encoding surface polysaccharide synthesis, 1,354,215–1,380,897 nt.

quency indicated by these structural parameters are fundamental characteristics of the ribosomal gene loci [27] and therefore are congruent with our observations.

In contrast, the replicore on the left half of the genome contains several regions coding for surface-exposed structures, including lipo-oligosaccharide (LOS), surface polysaccharide, and capsular antigens, and flagellin and flagellar modification (locations of selected genes are shown in Figs. 3, 4, and 5). The DNA structural characteristics of the former two gene clusters include high proportions of palindromic and repeat sequences (see Fig. 4) and a comparatively low stacking energy, indicated by distinct red bands in Figure 3. These properties suggest a potential for rapid genomic change by recombination or mutation, features consistent with the observed variability of surface structures in *C. jejuni*. Quasi-palindromes and perfect palindromes, as well as local inverted and direct repeats, are concentrated in those regions (Fig. 4). Quasi-palindrome correction, whereby an imperfect palindrome is altered to produce a perfect palindrome [15], and palindrome-stimulated deletions [28] are common mechanisms of mutagenesis. Likewise, simple tandemly repeated sequences can lead to deletions or insertions [29], and in general repeats

can be involved in homologous recombination events, thereby contributing to variability in surface structures. The LOS and surface polysaccharide regions, but not the *fla* gene region, are characterized by less negative stacking energies, and the AT content in those regions is higher than the genome average. Chromosomal regions with less negative stacking energies will melt more readily and therefore may have a higher tendency to mutate, since many mutagenic events occur as single-stranded DNA. In summary, we see several indications of a high concentration of mutational hot spots in the surface structure regions. Note in Figure 5 that often palindromes are found within intergenic regions. Palindromes occurring at the 3' end of genes might form stem loop structures involved in termination of transcription.

3.4. DNA Structures in the Flagellar Loci

The region containing genes coding for flagellar production and modification is characterized by (in addition to an elevated level of local repeats) a remarkable concentration of global repeats, where the two *fla* genes (*flaA* and *flaB*) are

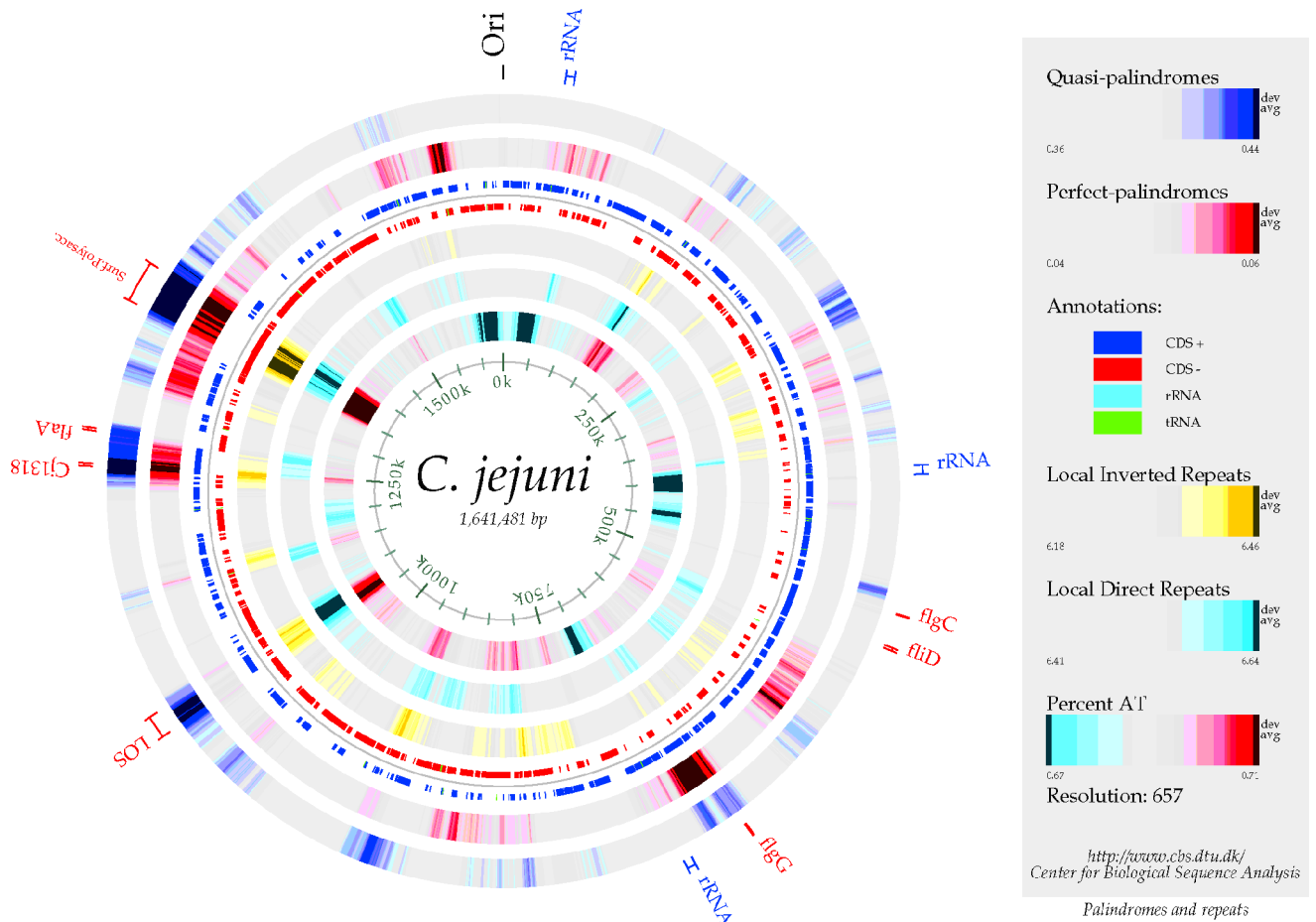


Figure 4. Atlas showing the distributions of repeats and palindromes, and AT content of the genome of *C. jejuni* NCTC11168. Legends: see Figure 3.

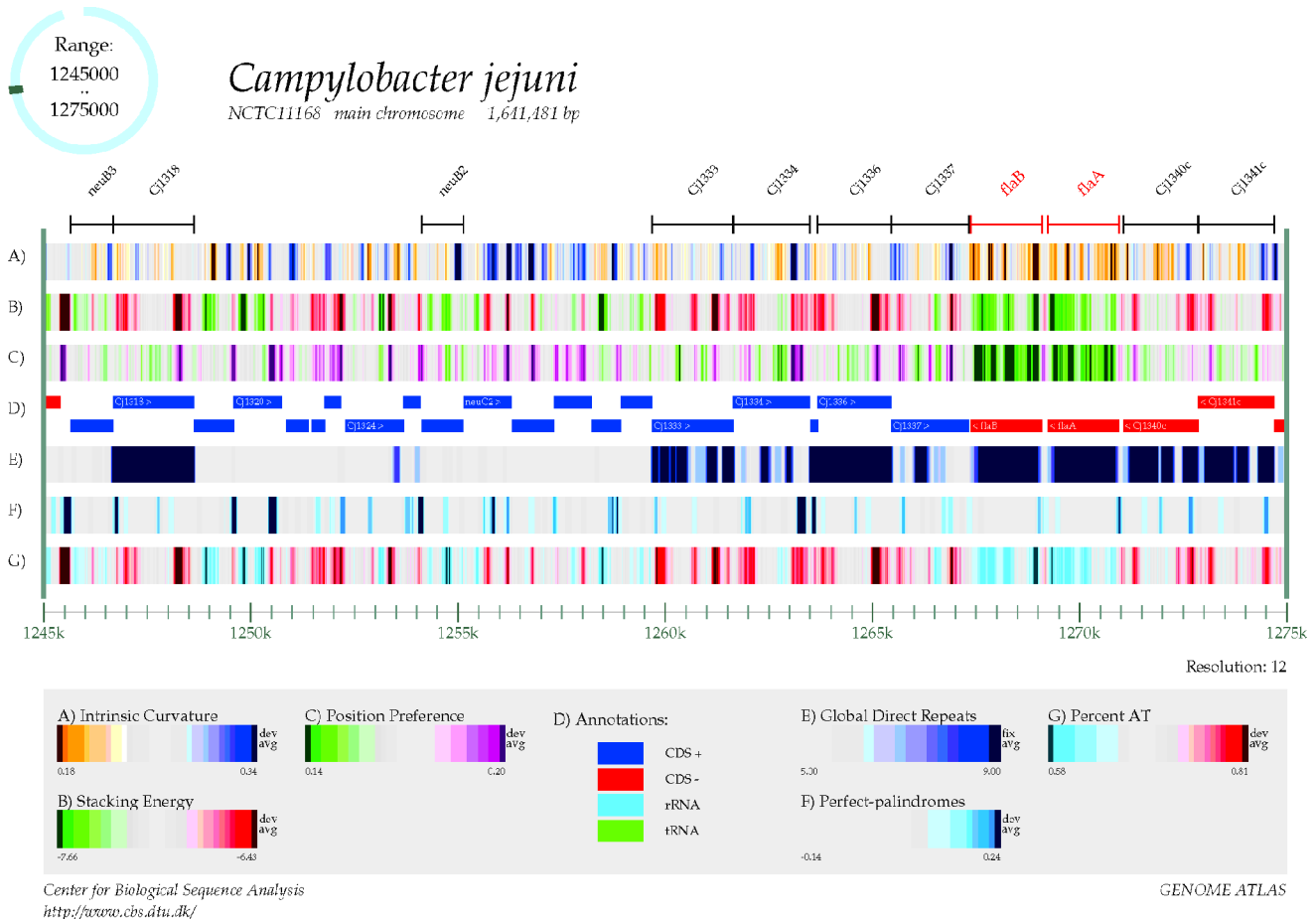


Figure 5. Atlas zoom showing structural parameters of the *flagellin* loci.

tandemly repeated, and the genes belonging to the 1318 gene family contain a conserved domain (Fig. 5) [30]. This indicates amenability to spontaneous genetic change either by recombination or by mutations. The diversity among *fla* genes of different *C. jejuni* strains is well known and is exploited in the *fla*-typing method [31], and the high potential for recombination of the *fla* genes is also well established [32]. Thus, a good correlation between experimental evidence of recombination potential and that inferred by DNA structural parameters as determined here is seen. Moreover, both *flaA* and *flaB* show low values for the position preference measure, and the propensity for high expression (discussed above) has been well documented in experiments showing that flagellar proteins of *C. jejuni* constitute a major component of the whole-cell protein content [33]. *FlaA* is the most well-characterized virulence factor of *C. jejuni* and is necessary for chemotactic behavior and colonization of host intestine [34], whereas the remaining genes marked in Figure 5 are involved in post-translational modification of flagellin [2, 35] and phase variation of flagella-mediated motility [30].

It is obvious that some of the parameters described are correlated to some degree. For instance, there is a clear

correlation between AT percentage and stacking energy (Fig. 5). There is, however, a less obvious relationship between the other parameters. Intrinsic curvature is not as strongly correlated with AT content, and although palindromes often can occur in AT-rich regions, this is not always the case.

4. SUMMARY AND CONCLUSIONS

There are many ways in which DNA structural conformations at the genomic level can affect biological processes. Examples of two important aspects that can be readily visualized in the *C. jejuni* GenomeAtlas (Fig. 3) are the localization of mutational hot spots (e.g., LOS, *flaA*, and surface polysaccharides) and the relationship between flexible regions and highly expressed genes (e.g., rRNA operons). There are significant correlations between DNA structures visualized in the GenomeAtlas and experimental data regarding the expression, stability, and recombination potential of genes. The concordance with well-characterized genes suggest a potential for determining key properties in less extensively studied loci and for investigating the possible significance of base substitutions in specific genes detected between different strains. We suggest

that the GenomeAtlas is one useful bioinformatics tool that can assist in unlocking the hitherto unrealized information potential contained in an entire genomic sequence.

Acknowledgments: The authors thank the people at the Center for Biological Sequence Analysis for their helpful discussions, particularly Carsten Friis, Lars Juhl Jensen, Marie Skovgaard, Hans Henrik Stærfeldt, and Peder Worning. This work is sponsored by a grant from the Danish Agricultural and Veterinary Research Council (grant SJVF-23-00-0285), and a grant to CBS (Søren Brunak) from the Danish National Research Foundation.

References and Notes

- C. R. Friedman, J. Neimann, H. C. Wegener, and R. V. Tauxe, Epidemiology of *Campylobacter jejuni* infections in the United States and other industrialized nations, in *Campylobacter*, edited by I. Nachamkin, and M. J. Blaser, ASM Press, Washington, DC (2000), p. 121.
- J. Parkhill et al., The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences, *Nature (London)* 403, 665 (2000).
- L. J. Jensen, C. Friis, and D. W. Ussery, Three views of microbial genomes, *Res. Microbiol.* 150, 773 (1999).
- A. G. Pedersen et al., A DNA structural atlas for *Escherichia coli*, *J. Mol. Biol.* 299, 907 (2000).
- E. Stackebrandt, R. G. E. Murray, and H. G. Trüper, *Proteobacteria classis nov.*, a name for the phylogenetic taxon that includes the “purple bacteria and their relatives,” *Int. J. Syst. Bacteriol.* 38, 321 (1988).
- S. L. W. On, Taxonomy of *Campylobacter*, *Arcobacter*, *Helicobacter* and related bacteria: current status, future prospects and immediate concerns, *Symp. Ser. Soc. Appl. Microbiol.* 30, 1S (2001).
- A. Bolshoy et al., Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles, *Proc. Natl. Acad. Sci. U.S.A.* 88, 2312 (1991).
- E. S. Shpigelman, E. N. Trifonov, and A. Bolshoy, CURVATURE: software for the analysis of curved DNA, *Comput. Appl. Biosci.* 9, 435 (1993).
- R. L. Ornstein and R. Rein, An optimized potential function for the calculation of nucleic acid interaction energies. I. Base stacking, *Biopolymers* 17, 2341 (1978).
- I. Brukner et al., Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides, *EMBO J.* 14, 1812 (1995).
- S. C. Satchwell, H. R. Drew, and A. A. Travers, Sequence periodicities in chicken nucleosome core DNA, *J. Mol. Biol.* 191, 659 (1986).
- D. Ussery et al., Bias of purine stretches in sequenced chromosomes, *Comput. Chem.* 26 (2002).
- L. S. Shlyakhtenko et al., Structure and dynamics of supercoil-stabilized DNA cruciforms, *J. Mol. Biol.* 280, 61 (1998).
- M. Skovgaard et al., The Atlas visualization of genome-wide information, *Methods Microbiol.* 33 (2002).
- M. Visnawathan et al., A novel mutational hotspot in a natural quasi-palindrome in *Escherichia coli*, *J. Mol. Biol.* 302, 553 (2000).
- H. J. Datta et al., Triplex-induced recombination in human cell-free extracts, *J. Biol. Chem.* 276, 18018 (2001).
- Q. Jiang, K. Hiratsuka, and D. E. Taylor, Variability of gene order in different *Helicobacter pylori* strains contributes to genome diversity, *Mol. Microbiol.* 20, 833 (1996).
- T. M. Wassenaar, S. L. W. On, and R. J. Meinersmann, Genotyping and the consequences of genetic instability, in *Campylobacter*, edited by I. Nachamkin and M. J. Blaser, ASM Press, Washington, D.C. (2000), p. 369.
- S. L. W. On et al., Validity of SmaI-defined genotypes of *Campylobacter jejuni* examined by Sall, KpnI, and BamHI polymorphisms: evidence of identical clones infecting humans, poultry, and cattle, *Epidemiol. Infect.* 120, 231 (1998).
- R. A. Alm and T. J. Trust, Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes, *J. Mol. Med.* 77, 834 (1999).
- L. Petersen et al., Comparison of genotypes and serotypes of *Campylobacter jejuni* isolated from Danish wild mammals and birds and from broiler flocks and humans, *Appl. Environ. Microbiol.* 67, 3115 (2001).
- M. Achtman et al., Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions, *Mol. Microbiol.* 32, 459 (1999).
- K. E. Dingle et al., Multilocus sequence typing system for *Campylobacter jejuni*, *J. Clin. Microbiol.* 39, 14 (2001).
- S. Suerbaum et al., Allelic diversity and recombination in *Campylobacter jejuni*, *J. Bacteriol.* 183, 2553 (2001).
- M. J. McLean, K. H. Wolfe, and K. M. Devine, Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes, *J. Mol. Evol.* 47, 691 (1998).
- R. R. Sinden, V. I. Hashem, and W. A. Rosche, DNA-directed mutations. Leading and lagging strand specificity, *Ann. N.Y. Acad. Sci.* 870, 173 (1999).
- C. R. Woese, Bacterial evolution, *Microbiol. Rev.* 51, 221 (1987).
- M. Bzimek and S. T. Lovett, Evidence for two mechanisms of palindrome-stimulated deletion in *Escherichia coli*. Single-strand annealing and replication slipped mispairing, *Genetics* 158, 527 (2001).
- M. Bzimek and S. T. Lovett, Instability of repetitive DNA sequences: the role of replication in multiple mechanisms, *Proc. Natl. Acad. Sci. USA.* 98, 8319 (2001).
- A. V. Karlyshev, D. Linton, and B. W. Wren, *Campylobacter jejuni* proteins of Cj1318 family (Maf) are involved in flagella-mediated motility and phase variation, *Int. J. Med. Microbiol.* 291, 78 (2001).
- R. D. Ayling et al., Restriction fragment length polymorphism of polymerase chain reaction products applied to the differentiation of poultry campylobacters for epidemiological investigations, *Res. Veterinary Sci.* 60, 168 (1996).
- C. S. Harrington, F. M. Thomson-Carter, and P. E. Carter, Evidence for recombination in the flagellin locus of *Campylobacter jejuni*: implication for the flagellin gene typing scheme, *J. Clin. Microbiol.* 35, 2386 (1997).
- D. G. Newell, H. McBride, and A. D. Pearson, The identification of outer membrane proteins and flagella of *Campylobacter jejuni*, *J. Gen. Microbiol.* 130, 1201 (1984).
- T. M. Wassenaar et al., Colonization of chicks by motility mutants of *Campylobacter jejuni* demonstrates the importance of flagellin A expression, *J. Gen. Microbiol.* 139, 1171 (1993).
- P. Guerry, R. A. Alm, C. Szymansky, and T. J. Trust, Structure, function and antigenicity of *Campylobacter flagella*, in *Campylobacter*, edited by I. Nachamkin, and M. J. Blaser, ASM Press, Washington, D.C. (2000) p. 405.
- J. F. Tomb et al., The complete genome sequence of the gastric pathogen *Helicobacter pylori*, *Nature (London)* 388, 539 (1997).
- S. Shigenobu et al., Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS, *Nature (London)* 407, 81 (2000).
- T. Hayashi et al., Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12, *DNA Res.* 8, 11 (2001).

39. R. D. Fleischmann et al., Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd, *Science* 269, 496 (1995).
40. B. J. May et al., Complete genomic sequence of *Pasteurella multocida*, Pm70, *Proc. Natl. Acad. Sci. U.S.A.* 98, 3460 (2001).
41. C. K. Stover et al., Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen, *Nature (London)* 406, 959 (2000).
42. M. McClelland et al., Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2, *Nature (London)* 413, 852 (2001).
43. J. Parkhill et al., Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18, *Nature (London)* 413, 848 (2001).
44. J. F. Heidelberg et al., DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*, *Nature (London)* 406, 477 (2000).
45. A. J. Simpson et al., The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis, *Nature (London)* 406, 151 (2000).
46. J. Parkhill et al., Genome sequence of *Yersinia pestis*, the causative agent of plague, *Nature (London)* 413, 523 (2001).
47. J. Parkhill et al., Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491, *Nature (London)* 404, 502 (2000).
48. D. W. Wood et al., The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58, *Science* 294, 2317 (2001).
49. V. G. DelVecchio et al., The genome sequence of the facultative intracellular pathogen *Brucella melitensis*, *Proc. Natl. Acad. Sci. U.S.A.* 99, 443 (2002).
50. W. C. Nierman et al., Complete genome sequence of *Caulobacter crescentus*, *Proc. Natl. Acad. Sci. U.S.A.* 98, 4136 (2001).
51. T. Kaneko et al., Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*, *DNA Res.* 7, 331 (2000).
52. H. Ogata et al., Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*, *Science* 293, 2093 (2001).
53. S. G. Andersson et al., The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria, *Nature (London)* 396, 133 (1998).