

Explorative Multifactor Approach for Investigating Global Survival Mechanisms of *Campylobacter jejuni* under Environmental Conditions

Birgitte Moen,^{1,2*} Astrid Oust,^{1,2} Øyvind Langsrud,² Nick Dorrell,³ Gemma L. Marsden,^{4†} Jason Hinds,⁴ Achim Kohler,² Brendan W. Wren,³ and Knut Rudi^{2*}

Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences,¹ and Matforsk, Norwegian Food Research Institute,² Ås, Norway, and Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine,³ and Department of Cellular and Molecular Medicine, St. George's Hospital Medical School,⁴ London, United Kingdom

Received 15 July 2004/Accepted 19 October 2004

Explorative approaches such as DNA microarray experiments are becoming increasingly important in microbial research. Despite these major technical advancements, approaches to study multifactor experiments are still lacking. We have addressed this problem by using rotation testing and a novel multivariate analysis of variance (MANOVA) approach (50-50 MANOVA) to investigate interacting experimental factors in a complex experimental design. Furthermore, a new rotation testing based method was introduced to calculate false-discovery rates for each response. This novel analytical concept was used to investigate global survival mechanisms in the environment of the major food-borne pathogen *C. jejuni*. We simulated nongrowth environmental conditions by investigating combinations of the factors temperature (5 and 25°C) and oxygen tension (anaerobic, microaerobic, and aerobic). Data were generated with DNA microarrays for information about gene expression patterns and Fourier transform infrared (FT-IR) spectroscopy to study global macromolecular changes in the cell. Microarray analyses showed that most genes were either unchanged or down regulated compared to the reference (day 0) for the conditions tested and that the 25°C anaerobic condition gave the most distinct expression pattern with the fewest genes expressed. The few up-regulated genes were generally stress related and/or related to the cell envelope. We found, using FT-IR spectroscopy, that the amount of polysaccharides and oligosaccharides increased under the nongrowth survival conditions. Potential mechanisms for survival could be to down regulate most functions to save energy and to produce polysaccharides and oligosaccharides for protection against harsh environments. Basic knowledge about the survival mechanisms is of fundamental importance in preventing transmission of this bacterium through the food chain.

DNA microarrays and other explorative approaches are increasingly being used to investigate microbial systems (7, 33). However despite these technical developments, such studies have mainly been limited to single-factor experiments (7). The knowledge gained is quite narrow and does not sufficiently reflect the biology of the systems investigated. There is clearly a need for new general concepts in explorative analyses of biological systems, which account for interacting factors in multifactor experiments.

In this work we set up a complex experimental design in which DNA microarrays were used to investigate global gene expression patterns and Fourier transform infrared (FT-IR) spectroscopy allowed the analysis of the total biomolecule composition in the cell (11, 12). Interacting effects of experimental factors were investigated by using a novel multivariate analysis of variance (MANOVA) approach (50-50 MANOVA) (20) and rotation testing (21). 50-50 MANOVA was developed

to handle several colinear responses in combination with general linear modeling (analysis of variance and regression). This is not possible with traditional MANOVA analysis. Rotation testing is advantageous compared to the commonly used permutation testing, since permutation testing on multifactor experiments is technically difficult and exact permutation analysis is limited to relatively simple models. In combination with rotation testing, a novel approach for determining the false-discovery rate (FDR) was developed for better significance testing of interacting parameters (see Appendix). This new FDR method is more reliable than other FDR methods when there is high correlation between responses.

The aim of this work was to investigate the global survival mechanisms of *Campylobacter jejuni* in the environment by using the novel analytical concepts described above. *C. jejuni* is the leading cause of food-borne bacterial diarrheal disease throughout the world (2). The four major sources of infection are undercooked poultry meat, untreated water, raw milk, and pets (19). Although the vast majority of infections are self-limiting, severe sequelae can occur in the form of a neuromuscular paralysis such as Guillain-Barré syndrome (24). *C. jejuni* does not normally grow outside the host, unlike most other bacterial food-borne pathogens, but it appears to have the ability to survive for extended periods in the environment (25).

* Corresponding author: Mailing address: Matforsk, Norwegian Food Research Institute, Osloveien 1, N-1430 Ås, Norway. Phone: 47 64 97 01 00. Fax: 47 64 97 03 33. E-mail for Birgitte Moen: birgitte.moen@matforsk.no. E-mail for Knut Rudi: knut.rudi@matforsk.no.

† Present address: Department of Genetics, University of Leicester, Leicester LE1 7RH, United Kingdom.

Nevertheless, *C. jejuni* is very sensitive to environmental stress and appears to lack most of the stress response factors necessary for survival of other bacteria in adverse environments (25, 26). Additionally, it does not normally grow aerobically and is not readily transmitted between humans, yet it is the most frequent food-borne pathogen. This is often referred to as the *Campylobacter* conundrum (15) and underlines the fact that the multiple survival mechanisms in the environment of this important human pathogen are poorly characterized. Understanding these mechanisms is crucial to the design of intervention strategies to reduce *C. jejuni* in the food chain and to reduce the burden of *C. jejuni*-associated disease.

We focused our study on nongrowth survival conditions and emphasized the interaction between temperature and oxygen tension in our experimental design. *C. jejuni* encounters a wide range of temperature and oxygen concentrations in its contamination cycle and must therefore respond to these conditions. Our novel analytical approaches were used to develop a model for the potential mechanisms for survival of *C. jejuni* in the environment.

MATERIALS AND METHODS

Experimental design. We chose to study the survival mechanisms of *C. jejuni* in a rich medium (Mueller-Hinton [MH] medium) at two different temperatures (5 and 25°C) and under three atmospheric conditions (aerobic, microaerobic, and severely oxygen limited [anaerobic]). Both temperature and atmospheric conditions were chosen to mimic the environment that the bacterium may encounter during nongrowth survival conditions. The bacteria were cultivated at 42°C (to simulate poultry growth conditions) under microaerobic conditions before incubation under the different survival conditions. This starting culture was used as the day 0 (reference) sample.

We based our experimental design on an initial screening to get an overview of the survival range in *C. jejuni*. This screening was performed by using FT-IR spectroscopy, viable-dead staining (BacLight), and plate counts. Samples from all of the experimental combinations were analyzed after 0 (reference), 1, 2, 3, and 4 weeks. In addition plate counts and BacLight analyses were performed on samples stored for up to 11 weeks for the 5°C anaerobic condition.

In the final experimental design, DNA microarray analyses were performed in addition to FT-IR spectroscopy, BacLight analysis, and plate counts. Samples were analyzed after 0 (reference), 2, 4, and 7 days by using all of the analysis techniques described above. DNA microarray analyses were not performed on the 5°C aerobic day 7 samples or the 25°C aerobic day 4 and day 7 samples, as more than 10% of the bacterial cells were nonviable (determined by BacLight analysis). All analysis was based on three biological replicates, and in addition, microarray hybridizations were performed in duplicate for each total RNA sample, giving a total of six hybridizations per sample. RNA from day 0 was used as a reference in all hybridizations.

Strain and growth conditions. The genome-sequenced *C. jejuni* NCTC 11168 strain was kindly provided by Brendan Wren. This clone was used for the Sanger Centre *C. jejuni* sequencing project and was previously obtained from the National Collection of Type Cultures. The subculture history of this variant is unknown.

Bacteria were grown on selective blood agar plates (Oxoid Ltd.) in a microaerobic atmosphere for 48 h at 42°C. One colony was used to inoculate 50 ml of MH broth (Oxoid Ltd.), which was then incubated microaerobically for 48 h at 42°C. This culture was used to inoculate two 1-liter volumes of MH broth (prewarmed to 42°C under microaerophilic conditions) to a 1:100 dilution. The cultures were incubated microaerobically at 42°C with shaking at 75 rpm for 12 h to yield a cell count of approximately 10^8 CFU/ml. The reasons for using cells in exponential growth are that *C. jejuni* lacks a stationary-phase response (17) and that the responses are complex and difficult to define when the cultures reaches high cell densities. The two cultures were mixed, day 0 samples were collected, and the culture was then divided into six flasks, each containing 300 ml. The flasks were placed at their respective temperatures and atmospheres, with shaking at 75 rpm. Microaerobic and severely oxygen-limited (anaerobic) conditions were established by using the Oxoid CampyGen and AnaeroGen systems, respectively. All plate counts were done on MH agar plates.

BacLight live-dead fluorescence microscopy. One milliliter of bacterial culture was centrifuged at $10,000 \times g$ and 4°C in a microcentrifuge for 10 min. The supernatant was removed, and the cells were resuspended in 1 ml of filter-sterilized peptone water. This suspension was diluted to give approximately 10^7 CFU/ml, and 0.5 ml of this dilution was mixed with 0.5 ml of the BacLight (Molecular Probes) mixture (BacLight ampoules mixed in 5 ml of filter-sterilized water). The BacLight-cell mixture was incubated for 15 min in the dark at 4°C before being filtered on a 0.22- μ m-pore-size black polycarbonate filter (Osmonics Inc.). The filter was washed two times with 1 ml of filter-sterilized water and then viewed by fluorescence microscopy (Leica DMLB microscope) with an RT color spot camera and Spot Advanced software (Diagnostic Instruments, Inc.).

FT-IR sample preparation and measurements. The samples (30 ml) for FT-IR measurements were centrifuged at 4°C and $10,000 \times g$ for 30 min. The supernatant was removed before the cells were washed twice with 1 ml of saline solution with centrifugation at 4°C and $16,000 \times g$ for 15 min. After being washed, the cells were suspended in 40 μ l of distilled water before 35 μ l of each suspension was transferred to an IR-transparent crystal (ZnSe). Before FT-IR analysis, the samples were dried in a desiccator at 10^4 Pa with anhydrous silica gel (Prolabo, Paris, France). To acquire FT-IR spectra of the samples, a Bio-module (Bruker Optics, Ettingen, Germany), which is especially designed to measure microorganisms, coupled to an Equinox 55 spectrometer (Bruker Optics) was used. The spectra were recorded in the region between 4,000 and 500 cm^{-1} .

FT-IR spectra of microorganisms are usually divided into five regions that contain information from different cell components (24), as follows: 3,000 to 2,800 cm^{-1} , fatty acids in the bacterial cell membrane; 1,800 to 1,500 cm^{-1} , amide bands from proteins and peptides; 1,500 to 1,200 cm^{-1} , mixed region (proteins and fatty acids); 1,200 to 900 cm^{-1} , polysaccharides within the cell wall; and 900 to 700 cm^{-1} , "true" fingerprint region containing bands which cannot be assigned to specific functional groups. Before data analysis, the spectra were preprocessed. In the region from 2,300 to 720 cm^{-1} we used extended multiplicative signal correction (23). In the region from 3,000 to 2,800 cm^{-1} , the second derivative was calculated prior to extended multiplicative signal correction.

RNA extraction. Total RNA was extracted from *C. jejuni* by using the RNeasy Protect Bacteria Mini Prep kit (Qiagen). Briefly, 20 ml of RNA Protect Bacteria reagent (Qiagen) was added directly to 10 ml of bacterial culture, vortexed for 5 s, incubated at room temperature for 5 min, and then centrifuged at $10,000 \times g$ for 30 min at room temperature. The supernatant was discarded, and the pellet was frozen at -20°C for later RNA purification. The pellet was thawed, resuspended in 200 μ l of Tris-EDTA buffer containing 1 mg of lysozyme per ml, and incubated at room temperature for 10 min with vortexing every 2 min. The manufacturer's protocol was followed from this point, including the "on-column" DNase treatment. RNase-free water was added to the membrane, and after 1 min the spin column was centrifuged at $10,000 \times g$ for 1 min. The concentration and purity of the total RNA were analyzed with an Ultraspec 3000 spectrophotometer (Pharmacia Biotech) and the RNA 600 Nano LabChip system (Agilent Technologies).

Labeling of total RNA and microarray hybridization. Total RNA was reverse transcribed by using random hexamers in the presence of aminoallyl-dUTP, followed by labeling with Cy3 and Cy5 monoreactive dyes (The Institute for Genomic Research [TIGR] protocol, standard operating procedure no. M007). Briefly, 2 μ g of total RNA and 6 μ g of random hexamers (Invitrogen) in a reaction volume of 18.5 μ l were denatured at 70°C for 10 min, snap cooled on ice, and centrifuged briefly. Then, 6 μ l of First Strand buffer (Invitrogen), 3 μ l of 0.1 M dithiothreitol, 1.2 μ l of 12.5 mM deoxynucleoside triphosphate-aminoallyl-dUTP labeling mix (aa-dUTP [Ambion]-dTTP [Invitrogen], 2:3), and 400 U of SuperScript II reverse transcriptase (Invitrogen) were added. The labeling reaction mixture was incubated at 42°C overnight (~16 h). The first-strand synthesis reaction was stopped by adding 10 μ l of 0.5 M EDTA and 10 μ l of 1 M NaOH. The reaction mixture was incubated at 65°C for 15 min, and then 25 μ l of 1 M Tris-HCl (pH 7.0) was added. The removal of unincorporated aa-dUTP and free amines was performed according to the TIGR protocol with the Qiagen QIAquick PCR purification kit. After being dried in a Speed Vac, the samples were stored at -20°C . Coupling of aminoallyl-labeled cDNA to Cy Dye Ester (Amersham Biosciences) was done according to the TIGR protocol. The eluted Cy3- and Cy5-labeled samples were mixed and dried in a Speed Vac.

The *C. jejuni* slides were developed and printed by the B μ G@S Group (Bacterial Microarray Group at St. George's) in the United Kingdom. All procedures in the development of these arrays have been described previously (14). Microarray slides were incubated in a prehybridization buffer (3.5 \times SSC [1 \times SSC is 0.15 M NaCl plus 0.015 M sodium citrate], 0.1% sodium dodecyl sulfate [SDS], 10 mg of bovine serum albumin per ml) at 65°C for 20 min. After prehybridization, the slides were washed in distilled water for 1 min, followed by a 1-min wash in isopropanol. The dried Cy3- and Cy5-labeled cDNA was resuspended in 30- μ l

hybridization solution (4× SSC, 0.3% SDS), denatured at 95°C for 5 min, vortexed at low speed, and heated again at 95°C for 5 min. The sample was centrifuged briefly and applied to the prehybridized microarray underneath a 22-by 25-mm LifterSlip coverslip (Erie Scientific Company). The slides were placed in a waterproof hybridization chamber for hybridization in a 65°C water bath overnight (~16 h). After hybridization, the slides were washed in 1× SSC buffer with 0.05% SDS at 65°C for 2 min, followed by two washes in 0.06× SSC buffer, each for 2 min at room temperature. The slides were dried by centrifugation (90 × g for 12 min).

Data acquisition and analysis. Slides were scanned with the ScanArray Express 1.0 scanner (Packard BioScience), following the manufacturer's guidelines. The intensities of the fluorescent spots were quantified with ImaGene 5 (BioDiscovery Inc.) software. Background subtraction and normalization were performed in GeneSpring 6.1 (Silicon Genetics).

Statistical analyses were performed according to the approach implemented in the 50-50 MANOVA software (20; www.matforsk.no/ola). Global significance was tested by 50-50 MANOVA, and corresponding explained variances were calculated based on sums of squares summed over all responses. Significance analysis for each individual response was performed with *P* value adjustment calculated according to the rotation testing principles described by Langsrud (21). However, a modified procedure that adjusts *P* values according to an FDR criterion was used (see Appendix). The main effect of day (time) was illustrated by using adjusted (adjusted for unbalanced design caused by missing observations) means. These means were shown directly as curves (FT-IR spectroscopy). A principal-component score plot (22) was used to illustrate the microarray data.

The analyses were performed in two steps to handle the split-plot structure (4) of this experiment. The first step used a model that was saturated in the whole-plot factors (biological replicate, temperature, and atmosphere). In such a model the effects of day (time) and interactions with day could be tested the usual way. Only data for day 2 were used to analyze the whole-plot factors to simplify the analyses. Then, the biological replicate was considered as a blocking factor (i.e., no interactions with biological replicate). Ordinary second-order models were used, and the factors day and atmosphere were modeled as both categorical and continuous variables. Reformulated models that include the effects of day and atmosphere separately for the two temperatures were also utilized.

The input data for the 50-50 MANOVA analyses were the normalized log of ratio data (normalized to 16S and 23S rRNAs). The data did not include any missing values, resulting in the analysis of 1,169 genes out of a total of 1,654 *C. jejuni* NCTC 11168 genes.

The microarray data are available through BμG@Sbase and ArrayExpress. The accession number for BμG@Sbase is E-BUGS-19 (<http://bugs.sghms.ac.uk/E-BUGS-19>), and the accession number for ArrayExpress is E-BUGS-19.

RESULTS

Investigation of survival range. The survival range for *C. jejuni* was investigated by using a combination of plate counts, BacLight staining, and FT-IR spectroscopy. *C. jejuni* survived best at low temperature (5°C) under anaerobic conditions. There was only a 0.1-log-unit decline in CFU after 4 weeks under these conditions. After 11 weeks the cells were difficult to recover on plates, but approximately 90% of the cells still stained as viable by BacLight analysis. Surprisingly, cells survived better under 25°C anaerobic conditions than under either the microaerobic or aerobic conditions tested (as determined by plate counts). The general trends were that the cells lost the ability to be recovered on plates before the membrane integrity was disrupted. Most changes in the FT-IR spectra occurred before day 14. After that, only minimal spectral changes were observed (data not shown).

Based on this initial investigation, time points up to 7 days were chosen for further investigation of the survival mechanisms of *C. jejuni*. This was a balance between maintaining at least 90% viable cells as assessed by BacLight staining and ensuring that the experiment should cover the main changes in the FT-IR spectrum.

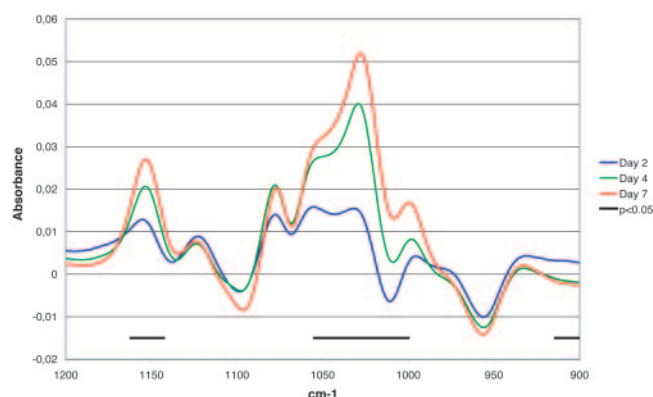


FIG. 1. Difference spectra (FT-IR) of the polysaccharide-oligosaccharide region. Difference spectra between the average of the day 0 samples and the average of the samples from days 2, 4, and 7 are shown. The horizontal bars indicate the areas of the spectra with significant changes ($P < 0.05$ by FDR).

Investigation of the survival mechanism. (i) Global macromolecular changes. FT-IR spectroscopy was used to investigate the global changes in the cell biomolecule composition. The FT-IR spectra showed an overall decrease in the protein region (data not shown) and an increase over time in the polysaccharide-oligosaccharide region (1,200 to 900 cm^{-1}) for the conditions tested (Fig. 1). Table 1 shows that both day (time) and temperature influenced the polysaccharide-oligosaccharide region significantly. The influence of the temperature is seen as slight pattern differences in some areas in the spectra between the two temperatures (based on day 2 only) (data not shown). The effect of atmosphere, however, did not influence the polysaccharide-oligosaccharide region. The lipid region of the spectrum (3,000 to 2,800 cm^{-1}) was influenced mainly by the incubation temperature (Table 2) and to some degree by time (lower explained variance than temperature). The effect of atmosphere was not significant (Table 2).

(ii) Global changes in gene expression. The choice of normalization plays an important role in the interpretation of the gene expression data. We evaluated both LOWESS (27, 32) and normalization to 16S and 23S rRNAs. Normalization to

TABLE 1. Explained variance and significance (by 50-50 MANOVA) of the polysaccharide-oligosaccharide region of the FT-IR spectra

Source	Explained variance	<i>P</i> value
Biological replicate ^c	0.136503	0.391103 ^a
Atmosphere	0.044705	0.231840 ^a
Temp	0.090576	0.000585 ^a
Day	0.092035	0.000000 ^b
Atmosphere-temp	0.013410	0.059553 ^a
Atmosphere-day	0.006753	0.533973 ^b
Temp-day	0.031119	0.002340 ^b
Atmosphere-atmosphere	0.006339	0.244451 ^a
Day-day	0.021056	0.000534 ^b
Error ^d	0.457255	

^a *P* value from day 2 only.

^b *P* value according to split-plot model.

^c Two degrees of freedom.

^d Thirty-two degrees of freedom.

TABLE 2. Explained variance and significance (by 50-50 MANOVA) of the lipid region of the FT-IR spectra

Source	Explained variance	P value
Biological replicate ^c	0.178863	0.002386 ^a
Atmosphere	0.008688	0.356250 ^a
Temp	0.252734	0.000000 ^a
Day	0.031989	0.000002 ^b
Atmosphere-temp	0.011888	0.394882 ^a
Atmosphere-day	0.018434	0.022058 ^b
Temp-day	0.079724	0.000003 ^b
Atmosphere-atmosphere	0.011588	0.516633 ^a
Day-day	0.022976	0.002153 ^b
Error ^d	0.338983	

^a P value from day 2 only.

^b P value according to split-plot model.

^c Two degrees of freedom.

^d Thirty-two degrees of freedom.

16S and 23S rRNAs gave the most biologically sound results, since most genes seemed to be down regulated for the conditions tested. To validate our choice of normalization, the total amount of RNA that was retrieved per cell was quantified. The total amount of RNA did not change during storage (results not shown). This means that the rRNA level in the cell is constant, since most of the RNA in the cell is ribosomal.

The overall biological information in the data was evaluated by using principal-component analyses prior to the in-depth statistical analyses. A principal-component score plot of all the data showed that the biological information in the data is high compared to the variation between the biological replicates (Fig. 2). The major biological trends were that most of the genes were either down regulated or unchanged under non-growth survival conditions and that the 25°C anaerobic condition was distinct from the other conditions tested (Fig. 3). In this condition, the cells had even more genes down regulated than in the other conditions. Although they were down regu-

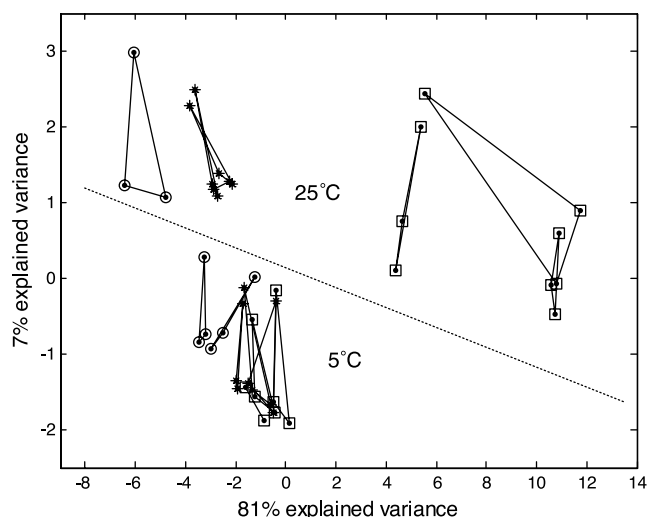


FIG. 2. Principal-component score plot of the microarray data. The different symbols represent the different atmospheric conditions (○, aerobic; *, microaerobic; □, anaerobic), and the lines between the symbols connect the different biological replicates.

lated compared to day 0, most genes increased their expression over time in this condition (Table 3). To identify up-regulated genes, significance (5% FDR) and corresponding *t* values were tested for all genes. The analysis was based on day 2 values only and identified 102 genes with an average log of expression ratio greater than zero over all experimental levels.

Atmosphere and the interaction between atmosphere and temperature were the factors with the largest influence on the gene expression patterns (Table 4). There were larger differences between samples from 25°C than between those from 5°C. This is illustrated in Fig. 2 and shows that the response to oxygen is larger at higher temperatures. The effect of day (time) was mainly dominated by linear trends (Table 4). Analyses also showed that the effect of day was dependent on the temperature (Table 4). We therefore investigated the effect of day at the two different temperatures, 5 and 25°C. Only 12 genes were significantly affected by day at 5°C, all of which showed a negative effect of day. This means that the ratio of the significant genes decreased over time. At 25°C almost all genes were significantly positively affected by day, meaning that the ratio of the significant genes increased over time (Fig. 4).

Analysis of the two temperatures showed that a large number of the significant genes were negatively affected by temperature, meaning that these genes had a higher ratio at 5°C than at 25°C (Table 3). Analyses of the effect of temperature and atmosphere were based only on samples from day 2. The effect of atmosphere at 5°C was positive for all except three significant genes (Cj0264c, encoding a molybdopterin-containing oxidoreductase; Cj0265c, encoding a putative cytochrome *c*-type heme-binding periplasmic protein; and Cj1356, encoding a putative integral membrane protein). At 25°C, the effect of atmosphere was positive for all significant genes. This means that the ratio of the significant genes increases with increasing oxygen concentration (Fig. 5).

(iii) Specific pathways. From the interaction analyses, gene lists were made and studied according to biological function and pathway interaction. The genes that were down regulated compared to day 0 were involved mainly in translation, transcription, and synthesis and modification of macromolecules. The genes that were up regulated compared to day 0 were mainly genes involved in energy metabolism, cell envelope, transport or binding proteins, chaperones, and heat shock proteins, in addition to some genes coding for conserved hypothetical proteins and genes with unknown function. Among the up-regulated conserved hypothetical protein-coding genes were genes Cj0041 and Cj1463, which have been predicted to be *fliK* and *flgJ*, respectively (C. Penn, unpublished observation). Interesting up-regulated unknown genes were genes Cj0428 and Cj1242, which have been predicted to be class II flagellar genes (5). Many of the up-regulated cell envelope genes are involved in flagellar assembly. In this assembly cascade, most class I genes (*fli* genes) were down regulated or unchanged, while most class II genes (*flg* genes) were up regulated. The proposed anti- σ^{28} factor gene *flgM* (6, 13, 31) was up regulated, while the alternative sigma factor genes *rpoN* and *fliA* were down regulated in all conditions except the 25°C aerobic condition. The flagellin genes *fliA* and *fliB* were down regulated. Among the genes that were expressed at a higher level at the lower temperature were genes involved in the

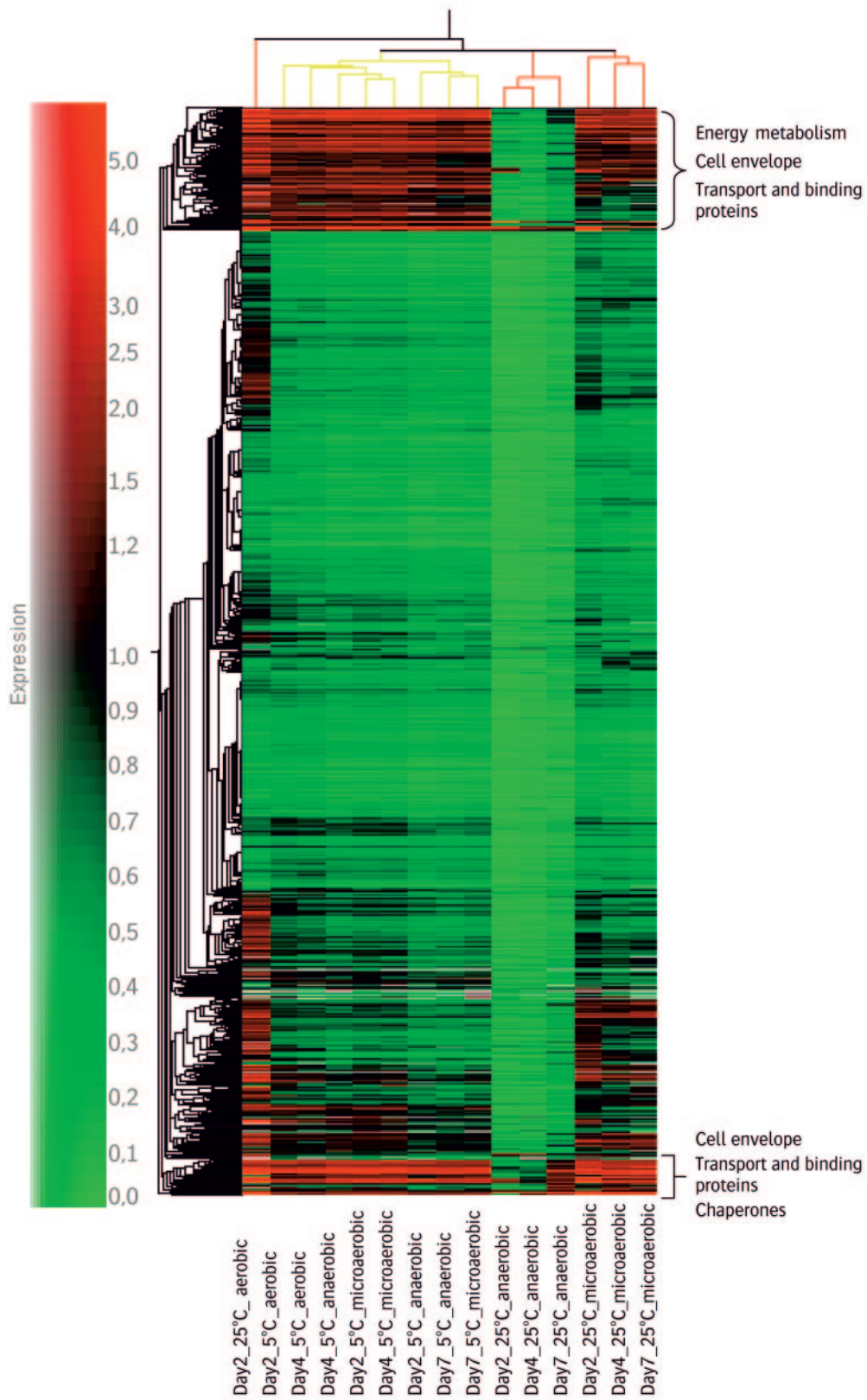


FIG. 3. Gene and condition cluster. The yellow lines represent samples from 5°C, and the red lines represent samples from 25°C. The brackets show areas with up-regulated genes and the corresponding pathways. Similarity measure: change correlation.

TABLE 3. Significance and corresponding *t* values of the gene expression data according to 5% FDR

Source	No. of significant genes (5% FDR)	No. of genes with:	
		Positive <i>t</i> value	Negative <i>t</i> value
Temp ^a	549	56	493
Day at 5°C	12	0	12
Day at 25°C	919	881	38
Atmosphere at 5°C ^a	332	329	3
Atmosphere at 25°C ^a	1,158	1,158	0

^a Based on day 2 values only.

tricarboxylic acid cycle, oxidative phosphorylation, and glycolysis and gluconeogenesis.

DISCUSSION

Significance testing of multifactor experiments. Rotation testing and 50-50 MANOVA enabled analysis of data generated by a relatively complex experimental design. These methods are exact under the assumption of the multivariate normal distribution, and they can be viewed as generalizations of the ordinary F test. This means that the classical single-response techniques (ANOVA and regression) for model building and analysis are directly extended to multiple responses. Once a model is chosen, the analysis according to 50-50 MANOVA or rotation testing is straightforward. The analysis in the present paper is an example of how the traditional tools of experimental design, analysis of variance, and regression modeling can be combined with new multivariate techniques.

Permutation testing (1, 8, 30) is an alternative approach, which avoids the need for data to fit the normal distribution. However, to detect differences, such methods need larger experiments than the corresponding normal-distribution-based methods. Permutation testing in multifactor experiments is also technically very difficult. To analyze each factor, the permutation process has to be done within level combinations of other factors. Models with interactions make such analysis even harder. Finally, exact permutation analysis is limited to relatively simple models. Nevertheless, complex models could be analyzed by certain permutation methods (e.g., permutation of residuals) and various bootstrap techniques (30). However,

TABLE 4. Explained variance and significance (by 50-50 MANOVA) of the gene expression data

Source	Explained variance	<i>P</i> value
Biological replicate ^c	0.039999	0.000000 ^a
Atmosphere	0.319991	0.000001 ^a
Temp	0.115880	0.000000 ^a
Day	0.014978	0.000003 ^b
Atmosphere-temp	0.227773	0.000003 ^a
Atmosphere-day	0.009907	0.062347 ^b
Temp-day	0.038288	0.000000 ^b
Atmosphere-atmosphere	0.023880	0.079563 ^a
Day-day	0.004401	0.033802 ^b
Error ^d	0.125063	

^a *P* value from day 2 only.

^b *P* values according to split-plot model.

^c Two degrees of freedom.

^d Thirty-two degrees of freedom.

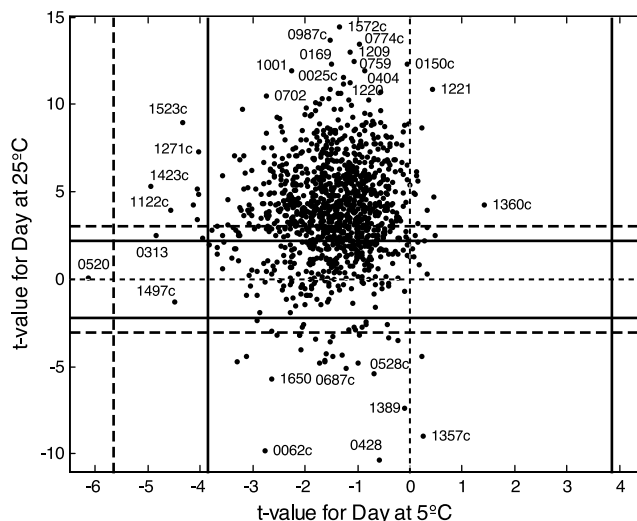


FIG. 4. Plot of interaction between day and temperature (5 and 25°C). The solid line represents the 5% FDR, and the dotted line represents the 1% FDR. The *t* value indicates whether the genes have increasing an ratio over time (positive *t* value) or a decreasing ratio over time (negative *t* value). Genes: Cj0025c, putative transmembrane symporter gene; Cj0062c, putative integral membrane protein gene; Cj0150c, aminotransferase gene; Cj0169, *sodB*; Cj0313, putative integral membrane protein gene; Cj0404, putative transmembrane protein gene; Cj0428, hypothetical protein Cj0428 gene; Cj0520, putative membrane protein gene; Cj0528, *flgB*; Cj0687c, *flgH*; Cj0702, *purE*; Cj0759, *dnaK*; Cj0774c, ABC transport system ATP-binding protein gene; Cj0987c, putative integral membrane protein gene, Cj1001, *rpoD*; Cj1122c, *wlaJ*; Cj1209, hypothetical protein Cj1209 gene; Cj1220, *groES*; Cj1221, *groEL*; Cj1271, *tyrS*; Cj1357c, putative periplasmic cytochrome *c* gene; Cj1360c, putative proteolysis tag for 10Sa RNA; Cj1389, pseudogene (transmembrane transport protein); Cj1423c, putative sugar-phosphate nucleotidyltransferase gene; Cj1497c, hypothetical protein Cj1497c gene; Cj1523c, hypothetical protein Cj1523c gene; Cj1572c, *nuoH*; Cj1650, hypothetical protein Cj1650 gene.

these methods assume that the data follow the normal distribution, and they are only approximately correct. Therefore, rotation testing is preferable, since this method is known to be correct when a normal distribution is assumed.

Adjusted *P* values according to false discovery-rates, or so-called *q* values, could alternatively be calculated by using the methodology described by Storey and Tibshirani (28). This means that the *q* values are calculated purely from the unadjusted *P* values, without allowing for dependence among the responses. It has, however, been shown that such *q* values are conservative under so-called weak dependence (28). It is furthermore hypothesized that it is likely that the weak dependence requirement is met in a genome-wide data set. Whether the weak dependence requirement is met in a microarray data set is dependent not only on the nature of the actual genes. The external sources of variation present in a specific experiment will also influence the dependence structure. Note that the dependence among the wavelengths of the described FT-IR data is definitely not “weak.”

The suggested false-discovery rate method makes use of the actual dependence between the responses, and it can safely be applied to data with any kind of dependence. The results are more conservative than the *q* values described above. This

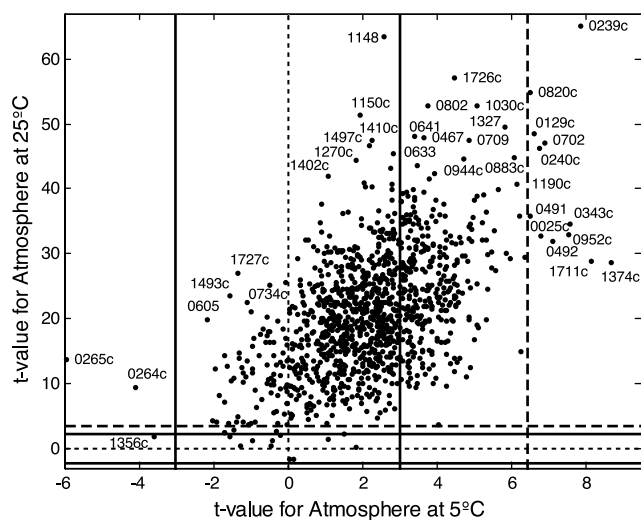


FIG. 5. Plot of interaction between atmosphere and temperature (5 and 25°C). The solid line represents the 5% FDR, and the dotted line represents the 1% FDR. The t value indicates whether the genes have an increasing ratio with increasing oxygen concentration (positive t value) or a decreasing ratio (negative t value). Genes: Cj0025c, putative transmembrane symporter gene; Cj0129c, outer membrane protein gene; Cj0239c, NifU protein homolog gene; Cj0240c, putative aminotransferase (NifS protein homolog gene); Cj0264c, molybdopterin-containing oxidoreductase gene; Cj0265c, putative cytochrome c -type heme-binding periplasmic protein gene; Cj0343c, putative integral membrane protein gene; Cj0467, amino acid ABC transporter integral membrane protein gene; Cj0491, *rpsL*; Cj0492, *rpsG*; Cj0605, putative amidohydrolase gene; Cj0633, putative periplasmic protein gene; Cj0641, hypothetical protein Cj0641 gene; Cj0702, *purE*; Cj0709, *ffh*; Cj0734c, *hisJ*; Cj0802, *cysS*; Cj0820c, *fliP*; Cj0883c, hypothetical protein Cj0883c gene; Cj0940c, *glnP*; Cj0952c, putative membrane protein gene; Cj1030c, *lepA*; Cj1148, *waaF*; Cj1150c, *waaE*; Cj1190c, putative methyl-accepting chemotaxis protein domain signal transduction protein gene; Cj1270c, hypothetical protein Cj1270c gene; Cj1327, *neuB2*; Cj1356c, putative integral membrane protein gene; Cj1374c, hypothetical protein Cj1374c gene; Cj1402c, *pgk*; Cj1410c, putative membrane protein gene; Cj1493c, putative integral membrane protein gene; Cj1497c, hypothetical protein Cj1497c gene; Cj1711c, *ksgA*; Cj1726c, *metA*; Cj1727c, *metY*.

false-discovery rate approach can be combined with both rotation testing and permutation testing.

Survival model. It is not yet established in the literature whether the survival of *C. jejuni* in the environment is an active or passive process (18, 19). We have shown that it is likely that *C. jejuni* has active mechanisms for survival in the environment under nongrowth conditions. We have identified global changes in both the lipid and polysaccharide-oligosaccharide composition. Furthermore, several genes are also up regulated under these conditions.

The general trend for the gene expression is that there is a massive down regulation under the different survival conditions compared to the reference (day 0). This reflects the fact that the cells were still actively metabolizing in the reference condition. Nevertheless, many genes are up regulated compared to the reference, and analyses indicate that most of the expression is stress related. More genes are expressed under high oxygen tension and temperature than under low oxygen and temperature. These are also the conditions with the lowest survival frequency. A possible survival mechanism may be to

down regulate as many genes as possible to save energy and to up regulate genes involved in energy metabolism and modification of the cell wall components. Even though *C. jejuni* does not grow at below 30°C, it has previously been shown that *C. jejuni* has considerable electron transfer chain activity even at temperatures below the minimal growth temperature (10). Interestingly, a significant number of genes involved in energy metabolism had higher ratio at 5°C than at 25°C. This could indicate that *C. jejuni* actually has a greater need for energy at lower temperatures. The 25°C anaerobic condition is distinct from the other conditions tested in that more genes are down regulated than in the other conditions. Anaerobic conditions have previously been suggested to be important in the pathogenesis of *C. jejuni*. Gaynor et al. showed that exposing *C. jejuni* NCTC 11168 to anaerobic conditions at 37°C for 24 h prior to inoculation increased its colonization potential (9).

Many of the up-regulated surface structure genes are involved in flagellar assembly. The flagellum is important for both motility and pathogenicity (29). In this pathway most *flg* genes (rod-ring-hook) were up regulated, while most *fli* genes were either unchanged or down regulated. The flagellin genes *flaA* and *flaB* were also down regulated. The fact that most flagellar assembly genes are up regulated, except for the flagellin genes and the *fli* genes, could mean that the cells have most of the flagellar assembly ready but that specific environmental signals may be required for the production of the flagellum.

It should, however, be emphasized that the genome-sequenced strain used in this study has undergone a series of phenotypic changes over time and may not be completely representative of the wild type. Although this strain and the original clonal clinical isolate appear to be clonal, it has been shown that these two isolates differ in colonization, virulence, and gene expression associated with respiration and metabolism genes (9). However, Jones et al. (16) have shown that NCTC 11168 is adaptable to chicken colonization and concluded that this strain is a suitable model strain. Also, due to the diversity among *C. jejuni* strains and the fact that potential survival mechanisms of *C. jejuni* still are poorly characterized, the hypotheses presented represent a novel insight into survival mechanisms of *C. jejuni*. Further investigation using several different *C. jejuni* strains is needed to test the generality of these hypotheses. The same will also hold true for other *C. jejuni* model strains, since the genomic stability and phenotypic diversity within this species are not well defined.

The observation that the amount of polysaccharides and oligosaccharides increases over time in all nongrowth conditions may indicate that these are important for survival. However, since none of the known polysaccharide genes are up regulated, we do not know if these polysaccharides correspond to changes in capsular polysaccharide. Recently, a new glycan that is susceptible to changes in growth conditions has been identified (*C. Szymanski, L. Fiori, and H. Jarrell, personal communication*). This glycan may correspond to osmoregulated periplasmic glucans (OPGs), which are important for survival in extreme conditions (3). The increase of polysaccharides and oligosaccharides observed in the FT-IR spectrum may be related to the expression of these OPGs. Future research is needed to confirm that the increase in the FT-IR spectrum is caused by OPGs and potentially to elucidate the genes involved. There is also a need to gain more general

knowledge about polysaccharides and oligosaccharides and their potential role in the survival mechanisms of *C. jejuni*.

APPENDIX

A new method of adjusting P values according to false discovery rates is developed and given below. This is done by generalizing the methodology described by Langsrud (21). We start with a brief introduction to this topic.

When an ordinary significance test is performed in a situation with no real effect, the probability of incorrectly obtaining a significant result (type I error) equals the significance level (usually 5%). When several responses (genes) are analyzed by individual significance tests, we will expect many type I errors, and it is questionable whether significant results can be interpreted as real effects. To avoid this problem, an alternative is to adjust the P values according to the family-wise error rate (FWE) criterion, which means that the probability of at least one type I error among all responses does not exceed the significance level. A classical method for such P value adjustment is Bonferroni's correction; that is, the P values are multiplied by the number of responses. But this method is extremely conservative and becomes useless in cases with a large number of responses. This problem can, however, be handled by using modern FWE methods based on permutation testing (30) or rotation testing (21). These methods take the correlations among responses into account.

A common viewpoint is that such P value adjustment is too strict. Instead of considering the probability of at least one type I error, an alternative is to estimate the false-discovery rate, which is the (expected) proportion of type I errors among all responses reported as significant. When using adjusted P values according to FDR, one accepts that 5% of the responses reported as significant at the 5% level are type I errors. The common FDR methods do not make use of the correlation structure among the responses and cannot be trusted if the weak dependence requirement is not met (28). Below we describe a new method that handles this problem. This method is obtained by modifying (or generalizing) the permutation and rotation methods for FWE adjustment.

Assuming a data set of N responses, we consider a particular model term and the corresponding univariate F statistics, F_1, F_2, \dots, F_N . We want to compare these observed statistics to statistics that are simulated under the complete null distribution. As discussed by Langsrud (21), such simulations can be performed by permutations (assuming independent observations) or by rotations (assuming multivariate normality). Then, a set of simulated statistics, $F_1^{(m)}, F_2^{(m)}, \dots, F_N^{(m)}$, is generated M times (where $m = 1, 2, \dots, M$).

We denote the ordered observed F statistics by $F_{(1)} \geq F_{(2)} \geq \dots \geq F_{(N)}$, and we compute their corresponding adjusted P values, $P_{(1)}, P_{(2)}, \dots, P_{(N)}$. Similarly, the simulated statistics are ordered as $F_{(1)}^{(m)} \geq F_{(2)}^{(m)} \geq \dots \geq F_{(N)}^{(m)}$. For $n = 1, 2, \dots, N$, a preliminary P value is computed as

$$\tilde{P} = \frac{1 + \sum_{m=1}^M \sum_{i=1}^n I(F_{(i)}^{(m)} \geq F_{(n)})}{1 + M}$$

where the function I transforms *true* and *false* to 1 and 0. This means that $\tilde{P}_{(n)}$ is the estimated expected proportion of the n largest F statistics that exceed $F_{(n)}$. Accordingly, this is one way to estimate the false-discovery rate. Note that this definition makes active use of the dependence structure. To make the expression in accordance with exact Monte Carlo P values, "1 +" is added in the numerator and denominator.

We can see that $\tilde{P}_{(1)}$ is identical to an adjusted P value according to the FWE criterion. For $n > 1$, $\tilde{P}_{(n)}$ will be less than (or equal to) the FWE-adjusted P value. On the other hand, when simulations are performed by rotations, $\tilde{P}_{(N)}$ is an estimate of the ordinary raw P value. For $n < N$, $\tilde{P}_{(n)}$ is greater than (or equal to) the raw P value. This means that our new P values are something between FWE-adjusted P values and raw P values. When the method is based on permutations, the relation to raw P values is more complicated, since the test statistics and the raw-permutation P values can be differentially ordered.

Our P values are not necessarily increasing, and we enforce mono-

tonicity by $P_{(n)} = \min_{i \geq n} (\tilde{P}_{(i)})$ for $n = 1, 2, \dots, N$. In other words, we apply a step-up procedure.

Unlike for several other FDR methods, the above description does not involve an estimate of the amount of responses with true null hypotheses. Instead, the calculations are directly based on the complete null distribution, and the results are therefore conservative. This choice is made to allow any kind of dependence among the responses. In fact, the procedure is valid in the extreme case where the correlation between all responses is 1. In that case, the raw P value (equal for all responses) is unchanged.

ACKNOWLEDGMENTS

This work was supported by grant 142656/140 from the Norwegian Research Council, the Fund for Research Levy on Agricultural Products, the United Kingdom Biotechnology and Biological Sciences Research Council, and the Leverhulme Trust.

The microarrays were kindly provided by the B_uG@S facility. We thank the Wellcome Trust for support of the B_uG@S facility.

REFERENCES

- Anderson, M. J. 2001. Permutation tests for univariate or multivariate analysis of variance and regression. *Can. J. Fish. Aquat. Sci.* **58**:626–639.
- Blaser, M. J. 1997. Epidemiologic and clinical features of *Campylobacter jejuni* infections. *J. Infect. Dis.* **176**(Suppl. 2):S103–S105.
- Bohin, J. P. 2000. Osmoregulated periplasmic glucans in *Proteobacteria*. *FEMS Microbiol. Lett.* **186**:11–19.
- Box, G., and S. Jones. 1992. Split-plot designs for robust product experimentation. *J. Appl. Stat.* **19**:3–26.
- Carrillo, C. D., E. Taboada, J. H. Nash, P. Lanthier, J. Kelly, P. C. Lau, R. Verhulp, O. Myktyczuk, J. Sy, W. A. Findlay, K. Amoako, S. Gomis, P. Willson, J. W. Austin, A. Potter, L. Babiuk, B. Allan, and C. M. Szymanski. 2004. Genome-wide expression analyses of *Campylobacter jejuni* NCTC11168 reveals coordinate regulation of motility and virulence by *flhA*. *J. Biol. Chem.* **279**:20327–20338.
- Colland, F., J. C. Rain, P. Gounon, A. Labigne, P. Legrain, and H. De Reuse. 2001. Identification of the *Helicobacter pylori* anti-sigma28 factor. *Mol. Microbiol.* **41**:477–487.
- Conway, T., and G. K. Schoolnik. 2003. Microarray expression profiling: capturing a genome-wide portrait of the transcriptome. *Mol. Microbiol.* **47**:879–889.
- Dudoit, S., J. P. Shaffer, and J. C. Boldrick. 2003. Multiple hypothesis testing in microarray experiments. *Stat. Sci.* **18**:71–103.
- Gaynor, E. C., S. Cawthraw, G. Manning, J. K. MacKichan, S. Falkow, and D. G. Newell. 2004. The genome-sequenced variant of *Campylobacter jejuni* NCTC 11168 and the original clonal clinical isolate differ markedly in colonization, gene expression, and virulence-associated phenotypes. *J. Bacteriol.* **186**:503–517.
- Hazeleger, W. C., J. A. Wouters, F. M. Rombouts, and T. Abee. 1998. Physiological activity of *Campylobacter jejuni* far below the minimal growth temperature. *Appl. Environ. Microbiol.* **64**:3917–3922.
- Helm, D., H. Labischinski, G. Schallehn, and D. Naumann. 1991. Classification and identification of bacteria by Fourier-transform infrared spectroscopy. *J. Gen. Microbiol.* **137**:69–79.
- Helm, D., and D. Naumann. 1995. Identification of some bacterial cell components by FT-IR spectroscopy. *FEMS Microbiol. Lett.* **126**:75–80.
- Hendrixson, D. R., and V. J. DiRita. 2003. Transcription of sigma54-dependent but not sigma28-dependent flagellar genes in *Campylobacter jejuni* is associated with formation of the flagellar secretory apparatus. *Mol. Microbiol.* **50**:687–702.
- Hinds, J., K. G. Laing, J. A. Mangan, and P. D. Butcher. 2002. Glass slide microarrays for bacterial genomes. *Funct. Microb. Genomics* **33**:83–99.
- Jones, K. 2001. The *Campylobacter* conundrum. *Trends Microbiol.* **9**:365–366.
- Jones, M. A., K. L. Marston, C. A. Woodall, D. J. Maskell, D. Linton, A. V. Karlyshev, N. Dorrell, B. W. Wren, and P. A. Barrow. 2004. Adaptation of *Campylobacter jejuni* NCTC11168 to high-level colonization of the avian gastrointestinal tract. *Infect. Immun.* **72**:3769–3776.
- Kelly, A. F., S. F. Park, R. Bovill, and B. M. Mackey. 2001. Survival of *Campylobacter jejuni* during stationary phase: evidence for the absence of a phenotypic stationary-phase response. *Appl. Environ. Microbiol.* **67**:2248–2254.
- Kelly, D. J. 2001. The physiology and metabolism of *Campylobacter jejuni* and *Helicobacter pylori*. *Symp. Ser. Soc. Appl. Microbiol.* **90**:16S–24S.
- Kist, M., and S. Bereswill. 2001. *Campylobacter jejuni*. *Contrib. Microbiol.* **8**:150–165.
- Langsrud, Ø. 2002. 50-50 multivariate analysis of variance for collinear responses. *J. R. Stat. Soc. Ser. D* **51**:305–317.
- Langsrud, Ø. 2005. Rotation tests. *Stat. Comput.* **15**:53–60.
- Martens, H., and T. Naes. 1989. *Multivariate calibration*. Wiley, Chichester, United Kingdom.
- Martens, H., and E. Stark. 1991. Extended multiplicative signal correction

- and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. *J. Pharm. Biomed. Anal.* **9**:625–635.
24. **Nachamkin, I., B. M. Allos, and T. Ho.** 1998. *Campylobacter* species and Guillain-Barre syndrome. *Clin. Microbiol. Rev.* **11**:555–567.
 25. **Park, S. F.** 2002. The physiology of *Campylobacter* species and its relevance to their role as foodborne pathogens. *Int. J. Food Microbiol.* **74**:177–188.
 26. **Parkhill, J., B. W. Wren, K. Mungall, J. M. Ketley, C. Churcher, D. Basham, T. Chillingworth, R. M. Davies, T. Feltwell, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Moule, M. J. Pallen, C. W. Penn, M. A. Quail, M. A. Rajandream, K. M. Rutherford, A. H. van Vliet, S. Whitehead, and B. G. Barrell.** 2000. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* **403**:665–668.
 27. **Quackenbush, J.** 2002. Microarray data normalization and transformation. *Nat. Genet.* **32**(Suppl.):496–501.
 28. **Storey, J. D., and R. Tibshirani.** 2003. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**:9440–9445.
 29. **Wassenaar, T. M., B. A. van der Zeijst, R. Ayling, and D. G. Newell.** 1993. Colonization of chicks by motility mutants of *Campylobacter jejuni* demonstrates the importance of flagellin A expression. *J. Gen. Microbiol.* **139**:1171–1175.
 30. **Westfall, P. H., and S. S. Young.** 1993. Resampling-based multiple testing: examples and methods for P-value adjustment. Wiley, New York, N.Y.
 31. **Wosten, M. M., J. A. Wagenaar, and J. P. van Putten.** 2004. The FlgS/FlgR two-component signal transduction system regulates the fla regulon in *Campylobacter jejuni*. *J. Biol. Chem.* **279**:16214–16222.
 32. **Yang, Y. H., S. Dudoit, P. Luu, D. M. Lin, V. Peng, J. Ngai, and T. P. Speed.** 2002. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30**:e15.
 33. **Ye, R. W., T. Wang, L. Bedzyk, and K. M. Croker.** 2001. Applications of DNA microarrays in microbial systems. *J. Microbiol. Methods* **47**:257–272.