



Unité de recherche : Dynamique,
évolution et expression de génomes de
microorganismes (FRE 2326)

Ecole Doctorale des Sciences de la Vie et de la Santé

En vue de l'obtention du grade de Docteur en Sciences de l' Université

Louis Pasteur

Discipline : biologie cellulaire et moléculaire

Etude de la diversité génétique de l'espèce *Campylobacter jejuni* par l'utilisation de puces à ADN

Présentée et soutenue publiquement par :

Frédéric Poly

Le 22 avril 2005

Jury :

Mario Keller

Agnès Labigne

Pierre Cornelis

Jean-Marie Meyer

Alain Stintzi

Président du Jury

Rapporteur externe

Rapporteur externe

Directeur de thèse

Co-Directeur de thèse

Remerciements

Je voudrais tout d'abord remercier Alain Stintzi pour m'avoir accueilli dans son laboratoire et permis de poursuivre mes travaux de recherche. Je lui exprime toute ma reconnaissance pour son initiation aux techniques de biologie moléculaire. Sa patience et ses connaissances seront toujours pour moi une source d'inspiration.

Je remercie Jean-Marie Meyer pour m'avoir convaincu de poursuivre mes recherches aux Etats-Unis. Je tiens à lui exprimer mon entière reconnaissance pour son aide et son support qu'il m'a apportés tout au long de ma thèse et de la rédaction de ce manuscrit.

Je remercie Madame Agnès Labigne, Messieurs Pierre Cornelis et Mario Keller, d'avoir bien voulu accepter d'évaluer ces travaux.

Je tiens également à remercier mes collègues de travail, Lisa Whitworth, Hemant Naikair et Kiran Palayda pour leur soutien et leur aide au laboratoire.

Je voudrais également remercier Isabelle Turcot, Corinne Snownoski, Nathalie Brignon, Nicolas Tuillon et Thierry Vincent pour leur soutien constant et leur amitié apportée lors de mon séjour aux Etats-Unis.

Je voudrais exprimer ma profonde gratitude à mes parents qui m'ont continuellement apporté un support indispensable à la réalisation de mes buts, à Emmanuel et Alexandra pour leur soutien moral et finalement à Ashli pour avoir partagé ma vie ces 2 dernières années.

Sommaire

Liste des figures	8
Liste des tableaux	10
Préambule	11
Abréviations	13
 Chapitre I : Introduction	 15
1 Description générale de l'espèce <i>Campylobacter jejuni</i>	16
2 Facteurs de pathogénicité chez <i>Campylobacter jejuni</i>	19
2-1 Définition	19
2-2 Glycosylation de protéines	20
2-3 Le système de O-glycosylation	21
2-4 Le système de N-glycosylation	23
2.5 Capsule	24
2.6 Lipooligosaccharide (LOS)	26
2.7 Toxines	30
 3 Vue d'ensemble des techniques de comparaison de génomes : cas des bactéries entéropathogènes.	 33
3-1 Introduction	33
3-2 Qu'ont besoin les bactéries pour devenir pathogènes?	34
3-2-1 Acquisition de pathogénicité via des plasmides	35
3-2-2 Acquisition de pathogénicité via un PAI	37
3-2-2-1 Définition d'un PAI	37
3-2-2-2 Système de sécrétion	38
3-2-2-3 Invasine, moduline et effecteurs	39
3-2-2-4 Toxines	40
3-2-2-5 Système d'acquisition de fer	40
3-3 Méthodes d'étude comparative des génomes bactériens	40

3-3-1 Introduction	40
3-3-2 Apport du séquençage d'autres bactéries entériques pathogènes .	42
3-3-2-1 génome de <i>Yersinia pestis</i>	42
3-3-2-1-1 <i>Y. pestis</i> Biovar Orientalis souche CO92	42
3-3-2-1-2 <i>Y. pestis</i> biovar Mediaevalis souche KIM	44
3-3-2-2 Séquences génomiques d' <i>Helicobacter pylori</i>	45
3-3-2-3 Le génome de <i>Salmonella enterica</i> sérovar typhi	47
3-3-2-4 Génomes de <i>Shigella flexneri</i> .	49
3-3-2-5 Séquençage du génome de la souche <i>E. coli</i> O157:H7	50
3-3-2-6 Séquences des génomes de <i>Vibrio paraheamolyticus</i> et <i>Vibrio cholerae</i>	52
3-4-2-7 Comparaisons des génomes de bactéries du genre <i>Campylobacter</i>	53
3-3-3 Hybridation soustractive (SSH)	57
3-3-3-1 Introduction	57
3-3-3-2 Technique	57
3-3-3-3 Applications	59
3-3-4 Comparaison génomique par utilisation de puces à ADN	62
3-3-5 Puces à ADN de type « shotgun »	66
 Chapitre II : Matériels et Méthodes	 68
1 Souches bactériennes, lignées cellulaires et conditions de culture	69

2 Mesures de concentration d'ADN	69
3 Test d'invasion bactérienne	70
4 Tests préliminaires à la fabrication de la puce à ADN par «shotgun ».	71
4-1 Elimination de l'ADN chromosomique	71
4-2 Restriction du plasmide	71
4-3 Dénaturation de l'ADN par chauffage	71
4-4 Préparation, marquage et hybridation des sondes	72
5 Construction de puces à ADN des souches <i>C. jejuni</i> 81-176 et TGH 9011	72
5-1 Extraction de l'ADN chromosomique de <i>C. jejuni</i>	72
5-2 Nébulisation	73
5-3 Clonage dans un vecteur plasmidique	74
5-3-1 Réalisation d'extrémités franches	74
5-3-1-1 Principe	74
5-3-1-2 Réaction	75
5-3-2 Déphosphorylation du vecteur	75
5-3-3 Ligation	76
5-4 Transformation par électroporation	76
5-5 Conservation des clones	77
5-6 Préparation d'ADN plasmidique pour la construction de banque génomique	77
5-7 Dépôts des ADN plasmidiques	79
6 Hybridation des puces à ADN	80
6-1 Prétraitement des puces avant utilisation	80
6-2 Marquage de l'ADN	81
6-2-1 Principe	81
6-2-2 Amorçage aléatoire ou « Random priming »	82
6-2-3 Marquage fluorescent	82

7 Révélation de l'hybridation et sélection de spots uniques	83
7-1 Révélation de l'hybridation	83
7-2 Sélection des spots spécifiques à la souche testée	83
8 Sélection des ORFs de la souche <i>C. jejuni</i> NCTC 11168 absents de la souche <i>C. jejuni</i> TGH 9011	84
8-1 Sélection	84
8-2 Vérifications	85
9 Détermination des clones correspondants au plasmide pTet de <i>C. jejuni</i> 81-176	85
10 Séquençage des clones et annotation	86
11 Etudes des gènes régulés par la protéine Fur par utilisation de la puce à ADN de TGH 9011	87
11-1 Clonage du gène <i>fur</i> dans un vecteur d'expression.	87
11-1-1 Conception d'amorces	88
11-1-2 PCR	88
11-1-3 Clonage	89
11-1-4 Vérifications des clones	90
11-2 Obtention de la protéine	90
11-2-1 Culture	90
11-2-2 Lyse	91
11-2-3 Purification	91
11-2-4 Vérification	92
11-3 Synthèse d'anticorps anti-Fur	92
11-4 Immunoprécipitation de la protéine Fur liée à l'ADN chromosomique	93
11-4-1 Préparation des cellules, fixation, lyse bactérienne et ultrasonation	93

11-4-2 Immunoprécipitation	94
Chapitre III : Identification des gènes spécifiques à la souche <i>Campylobacter jejuni</i> TGH 9011	95
1 Introduction	96
2 Publication	97
3 Résultats additionnels	113
Chapitre IV : Identification des gènes spécifiques à la souche <i>Campylobacter jejuni</i> 81-176	116
1 Introduction	117
2 Publication	117
3 Résultats additionnels	158
Chapitre V : Utilisation de la puce à ADN de <i>C. jejuni</i> TGH 9011 pour la caractérisation des gènes régulés par Fur	166
1 Clonage du gène <i>fur</i> dans un vecteur d'expression	169
2 Analyses préliminaires d'immunoprécipitation	173
3 Immunoprécipitation de la protéine Fur liée à l'ADN chromosomique	174
3-1 Application	174
3-2 Calculs théoriques de la quantité d'anticorps et de billes MagnaBind à utiliser pour l'immunoprécipitation.	174
Chapitre VI: Conclusion et Perspectives	176
Références	180

Liste des figures

Chapitre I

Figure 1 : Photo de <i>C. jejuni</i> par microscopie électronique (x 10000)	16
Figure 2 : Nombre de cas de maladies à entéropathogènes, relevés en 2002 aux Etats-Unis (source FoodNet)	17
Figure 3 : Dissection de poulet mettant en évidence les ceca	18
Figure 4 : Représentation de différents locus de modification des flagelles de 2 souches de <i>C. jejuni</i>	22
Figure 5 : Voie hypothétique de N-glycosylation de protéines chez <i>C. jejuni</i>	24
Figure 6 : Schéma de biosynthèse d'un LOS de type GM 1 produit par la souche NCTC 11168	27
Figure 7 : Locus des gènes impliqués dans la biosynthèse de LOS de <i>C. jejuni</i>	29
Figure 8 : Modèle du mécanisme d'action de la toxine CDT	32
Figure 9 : schéma récapitulatif de l'action de CDT sur les cellules épithéliales humaines	33
Figure 10 : Schéma explicatif de la technique d'hybridation soustractive	58

Chapitre II

Figure 1 : Schéma de ligation de l'insert dans le vecteur pCR4-TOPO	74
Figure 2 : Elution des plasmides des débris cellulaires	78
Figure 3 : Concentration et lavage des plasmides	78
Figure 4 : Elution de plasmides	79
Figure 5 : puce à ADN après hybridation	81
Figure 6 : Vecteur pASK-IBA	87

Figure 7 : Représentation schématique de la liaison du marqueur Strep-Tag II sur la molécule de strep-tactin (phase solide)	88
--	----

Chapitre III

Figure 1: Gènes identifiés par analyse par puce à ADN des gènes conservés et divergents entre les souches <i>C. jejuni</i> NCTC 11168 et ATCC 43431	103
Figure 2: Gel PFGE de l'ADN de <i>C. jejuni</i> digéré par les enzymes <i>SmaI</i> et <i>SalI</i>	107
Figure 3: Comparaison des clusters des gènes LOS des souches <i>C. jejuni</i> NCTC 11828 et NCTC 81116	108

Chapitre IV

Figure 1: Amplification PCR et représentation schématique chromosomique de la région TraG-like identifiée chez <i>C. jejuni</i> ATCC 43431 et 81-176	149
Figure 2: Comparaisons des gènes des souches <i>C. jejuni</i> NCTC 11168, ATCC 43431 et 81-176	149
Figure 3 : Représentation schématique du plasmide pTet de la souche <i>C. jejuni</i> 81-176	159

Chapitre V

Figure 1: Contrôle des clones fur II et VI par digestion avec EcoRI	170
Figure 2 : Contrôle des clones <i>fur</i> II et VI par PCR	170
Figure 3 : Digestion contrôlée des clones fur II et VI par les enzymes de restriction BclI et Sau3I	171
Figure 4 : dépôt de l'éluât contenant la protéine Fur recombinante sur gel de polyacrylamide	172
Figure 5: Surnageant d'immunoprecipitation suivie de purification par MagnaBind billes déposés sur gel de SDS polyacrylamide 15%	173

Liste des tableaux

Chapitre I

Tableau 1 : tableau récapitulatif des données numériques des génomes bactériens 47

Tableau 2 : synthèse des données de séquençage des génomes du genre 55

Campylobacter

Chapitre III

Tableau 1: ORFs de la souche *C. jejuni* ATCC 43431 absents du génome de 108
C. jejuni NCTC 11168

Tableau 2: tests préliminaires de traitement des clones avant construction 115
de la puce à ADN.

Chapitre IV

Tableau 1: Distribution des tailles des contigs identifiés comme unique 150
des régions chromosomiques et des plasmides pTet et pVir

Tableau 2: fragments chromosomiques présents chez *C. jejuni* 81-176 151
mais absent du génome de *C. jejuni* NCTC 11168

Tableau 3 : Liste des ORFs identifiées au cours de l'annotation de notre 163
séquence du plasmide pTet.

Préambule

L'espèce bactérienne *Campylobacter jejuni* est à l'origine de nombreuses affections entériques aiguës, de gravité variable chez l'homme: à l'échelle mondiale, elle provoque plus de cas de diarrhées que ne le font les salmonelles (28). La plupart des mécanismes de pathogénèse restent cependant peu documentés. En février 2000, la séquence complète de l'ADN chromosomique de la souche *C. jejuni* NCTC 11168 a été publiée (123). Elle a permis la construction de puces à ADN interrogeant la totalité des phases ouvertes de lecture (ORFs) du génome (46,99,125). Ce type de puces s'est avéré la pièce maîtresse du génotypage de différentes souches appartenant à l'espèce *C. jejuni*. L'analyse de telles études a permis d'identifier sur la carte physique du chromosome de la souche NCTC 11168 les zones de diversité dans l'espèce *C. jejuni*. Cependant si une telle analyse comparative permet d'identifier les gènes divergents caractérisant la souche testée par rapport à la souche de référence, elle ne donne, par contre, aucune indication sur de possibles complémentations géniques ou sur les gènes spécifiques à la souche étudiée. Peu de techniques permettant l'identification de séquences uniques à une souche donnée sont actuellement disponibles. Au cours de ce travail de thèse, nous avons mis au point un nouveau type de puce à ADN générée par dépôt des plasmides représentant une banque génomique d'une souche non séquencée. Cette approche a été conduite sur deux souches de laboratoire, TGH 9011 (également référencée sous ATCC 43431) et *C. jejuni* 81-176, présentant des phénotypes d'invasion cellulaire *in vitro* plus marqués que la souche de référence, *C. jejuni*. L'hybridation compétitive ADN/ADN entre la souche expérimentale utilisée pour la construction de la puce et la souche séquencée NCTC 11168 nous a permis d'isoler et de caractériser les séquences spécifiques à la souche testée.

Le chapitre I correspond à une introduction décrivant nos connaissances sur les facteurs de pathogénicité et de colonisation connus chez l'espèce *C. jejuni*, ainsi qu'une vue d'ensemble des techniques utilisées couramment pour l'étude des comparaisons génomiques de bactéries proches phylogénétiquement telles que les souches séquencées de *Yersinia pestis* ou *Helicobacter pylori*.

Le Chapitre II décrit les matériels et méthodes utilisés au cours de nos recherches.

Le chapitre III correspond à la publication numéro 1 relative aux résultats de fabrication et d'analyse de la puce à ADN de la souche *C. jejuni* TGH 9011.

Le chapitre IV correspond à la publication numéro 2 concernant les résultats de fabrication et d'analyse de la puce à ADN de la souche *C. jejuni* 81-176. Le chapitre V correspond à l'utilisation de la puce à ADN de TGH 9011 pour la recherche de gènes régulés par le facteur de transcription Fur (pour Ferric Uptake Regulator) par utilisation d'une immunoprecipitation de la protéine Fur fixée sur l'ADN.

Enfin, un chapitre de conclusions inclue les perspectives engendrées par ces travaux.

Abréviations

aa : acide aminé
ASN : asparagine
ADN : acide désoxyribonucléique
Amp : ampicilline
ATCC : American Type Culture Collection
BSA : Sérum albumine de bovin
Coll. : collaborateur
DO : densité optique
EDTA : acide éthylènediamine tétraacétique
EHEC : *E. coli* entérohémorragique
FBS : serum de fœtus de bovin
HDTMA : hexadécyl triméthyl ammonium
kb, pb ; kilobases, paires de bases
kDa : kiloDalton
LB : milieu de culture Luria-Bertani
MEM α : milieu essentiel minimum alpha
MH : milieu de culture Mueller-Hilton
min : minute
ORF : phase ouverte de lecture
PAI : îlot de pathogénicité
PCR : réaction de polymérisation en chaîne (polymerase chain reaction)
PZ : zone de plasticité
qsp : quantité suffisante pour
RM : système de restriction-modification de l'ADN.
rpm : rotation par minute
SER : serine
SDS : sodium dodecyl sulfate
SSC : Standard Saline Citrate

SSH : Hybridation soustractive

TE : Tris-HCl 10 mM, EDTA 1 mM

THR : thréonine

Tris : Tris hydroxymethyl amino méthane

U : unité

UPEC : *E. coli* uropathogénique

UV : ultra-violets

Chapitre I

Introduction

1 Description générale de l'espèce *Campylobacter jejuni*

Campylobacter a été décrit pour la première fois par le pédiatre et bactériologiste Théodore Escherich en 1880, à partir d'excréments d'enfant présentant des diarrhées. Bactérie classée tout d'abord dans l'espèce *Vibrio fetus*, il faut attendre Sebald et Veron qui proposent en 1963 de créer le genre *Campylobacter* (145). L'isolement de souches de *Campylobacter* à partir d'excréments humains a été couronné de succès grâce à l'élaboration d'un milieu sélectif en 1972 (40). Depuis, les bactéries du genre *Campylobacter* ont été reconnues comme bactéries pathogènes pour l'homme. L'espèce *Campylobacter jejuni* est la plus fréquemment impliquée dans les cas d'entérite (12,28). *C. jejuni* peut être trouvé en fonction des conditions environnementales sous forme de bâtonnets spiralés (de 1,0. à 6,0 μm de long et 0.2 à 0.5 μm de large, Figure 1), ou sous un aspect coccoïde pour les formes de résistance ou en phase de dégénérescence. *C. jejuni* présente fréquemment des flagelles à l'un ou aux deux pôles. La présence de flagelles associés à une forme spiralée confère à la bactérie un fort potentiel de motilité (92).

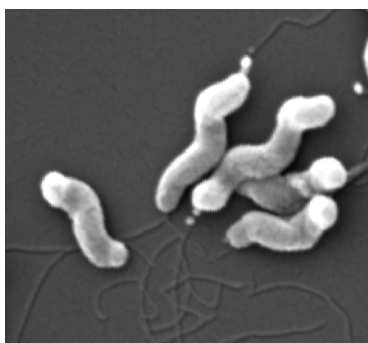


Figure 1 : Photo de *C. jejuni* par microscopie électronique (x 10000)

Le génome de la souche *C. jejuni* NCTC 11168, a été publié en février 2000 (123). La bactérie possède un chromosome circulaire de 1,64 Mb (G+C % = 30,6) codant pour 1654 ORFs. Le génome a la particularité de ne présenter aucune séquence d'insertion (IS), aucune séquence d'origine phagique, et de présenter un faible nombre de séquences hypervariables. Le génome présente un certain nombre de séquences répétées. Ces dernières correspondent à des répétitions de nucléotides G et C. Ces séquences sont

rencontrées au niveau de gènes codant pour des enzymes de biosynthèse ou impliquant une modification de structure de la membrane externe, ainsi qu'à des gènes à fonction inconnue.

Si le taux d'infection diminue régulièrement aux Etats-Unis depuis 1996, *Campylobacter* n'en reste pas moins le deuxième cas d'infection intestinale dans ce pays, juste derrière *Salmonella* (Figure 2).

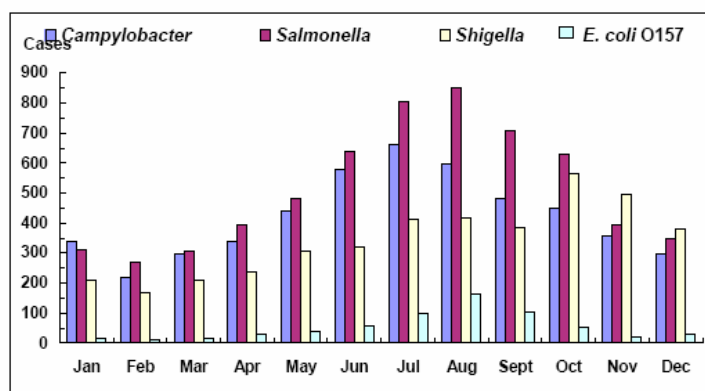


Figure 2 : Nombre de cas de maladies à entéropathogènes, relevés en 2002 aux Etats-Unis (source FoodNet)

Les infections à *Campylobacter* touchent environ un million d'américains chaque année, représentant 0.5% de la population du pays (1,2). Les infections causées par cette bactérie représentent donc un réel problème socio-économique.

C. jejuni est une bactérie microaérophile, qui nécessite une faible concentration atmosphérique en oxygène (3 - 5%) et une concentration en CO₂ de 3 à 15%. Les températures optimales de croissance vont de 37 à 42°C, températures correspondant aux températures de leurs hôtes (l'homme et le poulet, respectivement). L'incubation de *C. jejuni* varie de 2 à 5 jours.

Les sources majeures de contamination résultent de la consommation d'eau, de lait cru ou de viande de volaille (63). *C. jejuni* est une bactérie commensale de l'intestin de poulet. La bactérie colonise le mucus de la paroi intestinale de préférence dans le ceca et dans le petit intestin (Figure 3). En fonction de la souche et de ses relations avec le

poulet, il a été démontré de façon expérimentale qu'un nombre restreint de bactéries était nécessaire à la colonisation des intestins (32). Après colonisation, *C. jejuni* peut être récupéré jusqu'à environ 10^9 bactéries/g d'intestin. Connaissant les méthodes utilisées pour l'élevage de poulet destiné à la consommation humaine, il est aisé d'imaginer la facilité de propagation de *C. jejuni* via le contact de la nourriture avec les excréments au cours du processus d'abattage. La plupart des cas de campylobactériose sont sporadiques, cependant des cas de manifestation de *C. jejuni* apparaissent chaque année dans le monde entier (8).

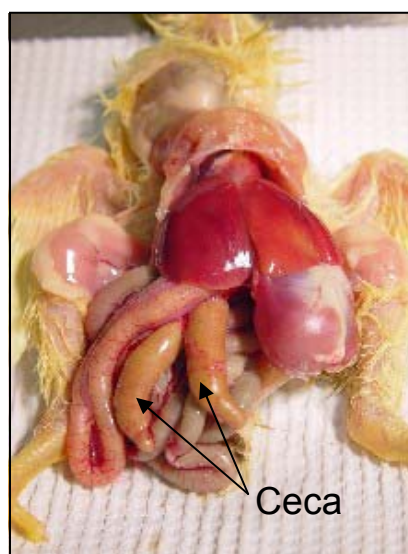


Figure 3 : Dissection de poulet mettant en évidence les ceca

Au Wisconsin par exemple, 75 personnes présentèrent en 2001 des troubles intestinaux en rapport avec une campylobactériose. Une coproculture a permis de caractériser dans 97 % des cas, la présence de *C. jejuni*. Presque tous les patients présentèrent des diarrhées associées à des crampes abdominales. Quelques patients ont également développé de la fièvre et/ou présentèrent du sang dans les selles. La source d'infection a été associée à la consommation de lait cru provenant d'une ferme locale (3).

Les aspects cliniques varient, sous forme de sévères diarrhées inflammatoires à des bénignes diarrhées dites aqueuses, sans signe d'inflammation (92). L'infection par *C. jejuni* n'est pas létale pour des adultes en bonne santé mais peut avoir des issues plus dramatiques pour des personnes immuno-déprimées, ou des enfants en bas âge (28). Des

complications post-infectieuses peuvent apparaître chez un faible nombre de patients. Ces complications sont l'arthrite, le syndrome de Guillain-Barré (GBS) (21). Environ 0.05% des infections par *Campylobacter* sont suivies de complication de type GBS (9). Il a été montré qu'environ 40% des patients développant ce syndrome présentent des évidences sérologiques ou culturelles d'une infection par *C. jejuni* (108). Les signes cliniques de GBS correspondent à une démyélination du système nerveux périphérique engendrant une paralysie aiguë du système neuromusculaire. Le développement de GBS est associé à une maladie auto-immune envers les structures des gangliosides présents à la surface des cellules nerveuses humaines. Cette réponse auto-immune est déclenchée après infection par certaines souches de *C. jejuni* présentant en surface des structures polysaccharides qui miment les structures des gangliosides hôte humains. Récemment Yuki et coll. ont démontré l'apparition de symptômes semblables à GBS chez des lapins infectés avec le LOS de type GM1 de *C. jejuni*, prouvant que cette maladie est due à une réaction autoimmunitaire (171).

2. Facteurs de pathogénicité de *Campylobacter jejuni*

2-1 Définition

Une bactérie pathogène est un microorganisme capable de causer une maladie à une plante, un animal ou un insecte. La pathogénèse est l'habilité à provoquer une maladie chez l'organisme hôte. De manière générale les microbes expriment leur pathogénèse à travers leur virulence, terme reflétant le degré de pathogénicité du microorganisme. Le degré de virulence d'un microorganisme pathogène est conféré par soit un déterminant génétique, biochimique ou une structure particulière provoquant une maladie chez l'organisme hôte. (Kenneth Todar, University of Wisconsin-Madison, Department of Bacteriology, www.textbookofbacteriology.net.)

Quatre mécanismes majeurs de pathogénicité bactérienne sont reconnus chez *C. jejuni* (161):

- Motilité (mécanisme largement étudié. Permet l'accès de la bactérie à la cellule hôte et sa pénétration) (111,115,163)
- Adhérence (probablement déterminant dans la colonisation) (88,142)
- Invasion (23,141)
- Production de toxine (162)

La publication de la séquence génomique de *C. jejuni* NCTC 11168 a permis des études fonctionnelles au sein de l'espèce (137,147). La séquence chromosomique présente quatre zones impliquées dans la biosynthèse de sucres: les deux systèmes de glycosylation protéique, la glycosylation des flagellines, la biosynthèse de structure extracellulaire de type lipooligosaccharide (LOS) et la biosynthèse de capsule. A l'exception d'un des systèmes de glycosylation protéique, tous les clusters de gènes impliqués dans ces processus ont été montrés comme étant hypervariables du fait de la présence de répétition G ou C dans leur séquence nucléotidique (123). Ces quatre mécanismes ont tous été montrés comme étant impliqués dans un mécanisme de virulence bactérienne (18,30,88,133).

2-2 Glycosylation de protéines

C'est depuis peu qu'il a été montré que les bactéries, comme les cellules eucaryotes, étaient capables de modifier leurs protéines par addition de résidus de sucre (ou glycosylation). Il apparaît en fait que les procaryotes présentent une variété de glycosylations supérieure à celle des eucaryotes. Le premier exemple de glycosylation de protéine bactérienne a été démontré à la surface de la bactérie *Halobacterium halobium* en 1989 (109). Les deux systèmes de modification protéique, N- et O-glycosylation, ont été identifiés avec une prédominance de liaison N pour les archéobactéries et de liaison O pour les eubactéries (132).

La N-glycosylation correspond à l'attachement d'une chaîne de résidus carbohydrates sur l'azote de la fonction amide d'un résidu asparagyl (ASN). La glycosylation de type O correspond à l'attachement d'une chaîne de carbohydrates sur la fonction hydroxyle d'un résidu serine (SER) ou thréonine (THR). La glycosylation de protéines est effectuée par

une enzyme spécialisée reconnaissant le résidu à glycosyler si celui-ci fait partie d'une séquence consensus spécifique à chaque organisme (132). Diverses fonctions ont été attribuées à la glycosylation de protéines chez les eucaryotes telles que la protection contre un clivage protéolytique ou la reconnaissance pour une adhésion cellulaire. Ces mêmes fonctions ont été postulées chez les procaryotes, mais dans la plupart des cas la preuve expérimentale reste à établir.

2-3 Le système de O-glycosylation

C. jejuni présente l'habilité à glycosyler ses protéines par les systèmes de N- ou O-glycosylation. Le système de O-glycosylation est attribué à la modification post-traductionnelle de la flagelline (composant protéique des flagelles) (150). Ces structures représentent une part très importante des protéines bactériennes glycosylées. Les résidus fixés représentent environ 10% de la masse totale de l'ensemble des protéines glycosylées. La modification post-traductionnelle des protéines de flagelline chez *C. jejuni* a été montrée par Doig *et al* (45) via la sensibilité des flagellines à l'oxydation au periodate et attachement aux molécules de lectine, molécules ayant une affinité pour l'acide sialique. L'analyse structurale des sucres fixés sur les flagellines a permis d'identifier la nature des résidus attachés. Ceux-ci correspondent à de nombreux monosaccharides analogues de l'acide pseudaminique et de ses dérivés (45). Ces composés présentent une réaction positive à une certaine molécule de lectine (45), observation ayant engendré à tort, l'hypothèse d'une présence d'acide sialique sur les flagellines.

La glycosylation des protéines de flagelle de la souche *C. jejuni* 81-176 concerne 19 résidus SER/THR du domaine central du filament. Aucune séquence protéique consensus n'a pu être déterminée. Néanmoins, les résidus glycosylés sont localisés dans une zone hydrophobique étroite, favorisant de façon hypothétique la glycosylation. Les flagellines sont connues comme étant des protéines dominantes dans le mécanisme de reconnaissance par le système immunitaire de l'hôte (90).

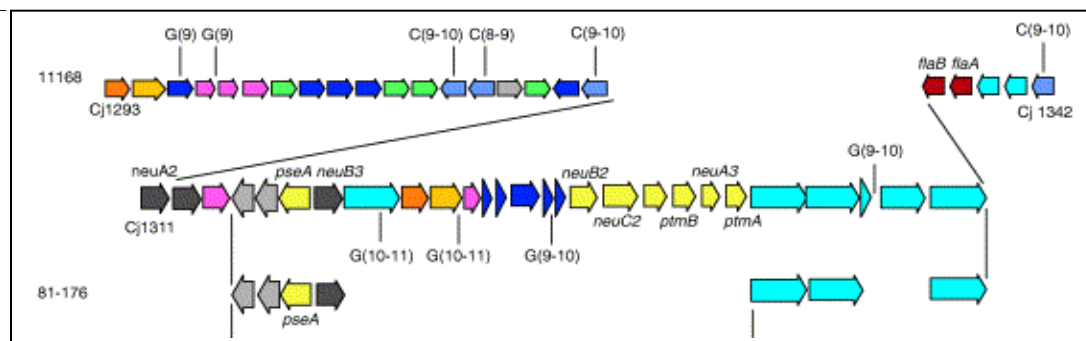


Figure 4 : Représentation de différents locus de modification des flagelles de 2 souches de *C. jejuni* (150). Lignes 1 & 2 *C. jejuni* NCTC 11168. Ligne 3, souche *C. jejuni* 81-176. Cette souche ne présente pas les gènes impliqués dans le processus de sialilation des protéines de flagelle.

Des analyses mutationnelles de gènes codant pour des enzymes impliquées dans la glycosylation des flagellines ont montré une implication de cette glycosylation dans la serospécificité des souches bactériennes (64,65). Le séquençage de NCTC 11168 a permis d'identifier un locus d'environ 50 gènes en amont des gènes *flaA* et *flaB* codant pour les flagellines. La plupart de ces gènes présentent des homologies avec les gènes de glycosylation de protéines bactériennes telles que des aminotransférases ou des acyltransférases. Ce locus, variable chez les souches de *C. jejuni* (Figure 4), est caractérisé par des séquences nucléotidiques présentant un nombre variable de répétitions de purines G ou C. La présence de ces variations peut entraîner un décalage du cadre de lecture lors de la traduction, engendrant une protéine tronquée non fonctionnelle. Ce système permet l'obtention d'une variabilité phénotypique potentiellement impliquée dans les mécanismes d'évasion immunitaire de l'hôte (78). Comparée à NCTC 11168, il manque à la souche 81-176 plusieurs gènes impliqués dans le processus de modification des protéines de flagelle tels que les gènes *ptm* ou les gènes impliqués dans la biosynthèse d'homologue d'acide pseudaminique (gènes *neu*) (150) (Figure 4).

2-4 Le système de N-glycosylation

La glycosylation de type N correspond au mécanisme de glycosylation de protéines le plus répandu chez *C. jejuni*. Les gènes impliqués dans ce processus ont été tout d'abord classés dans les mécanismes de biosynthèse de lipooligosaccharides (LOS) et de matériel capsulaire (60,93). Des analyses mutationnelles menées chez la souche 81-176 ont permis d'identifier la fonction réelle de ces gènes (150). Chez *C. jejuni*, un total d'environ 30 protéines a été suggéré comme étant glycosylé par ce mécanisme (170). La glycosylation implique l'attachement d'un heptasaccharide sur un résidu asparagine (Asn) présent dans la séquence consensus Asn-Xaa-Ser/Thr. La nature du résidu glucidique a récemment été montrée par spectrophotométrie de masse et par résonance magnétique nucléaire.

La voie de biosynthèse du système de glycosylation de type N- de *C. jejuni* est codée par un locus de 17 Kb constitué par les gènes *pglA* à *J*. Comparé aux mécanismes de glycosylation rencontrés chez *C. jejuni*, ce système est très conservé au sein de l'espèce et ne présente aucune zone GC.

Une voie de biosynthèse hypothétique basée sur les précédents travaux effectués chez *C. jejuni* en comparaison avec d'autres espèces bactériennes a été publiée (150). *PglB* est l'enzyme clef de cette voie de biosynthèse ([Figure 5](#)). Cette protéine présente des similarités avec une oligosaccharyl transférase STT3 de *S. cerevisiae* impliquée dans le système N-glycosylation (151). Cette enzyme identifiée chez *C. jejuni* représente le premier exemple de N-oligosaccharyltransférase identifié chez un procaryote. La voie de N-glycosylation est un processus de polymérisation via des glycosyltransférases d'un heptasaccharide sur un support lipidique (150). Une fois l'heptasaccharide transféré, le complexe lipo-heptasaccharide est tourné au travers de la membrane interne dans le périplasme via le transporteur WlaB (transporteur de type ABC) en présence d'ATP. Dans le périplasme, la protéine PglB catalyse le transfert du résidu glucidique du support lipidique sur un résidu Asn d'une protéine cible.

L'analyse mutationnelle de gènes impliqués dans ce mécanisme démontre une réduction d'attachement de la bactérie aux cellules-hôte in vitro, ainsi qu'une réduction de

colonisation (à faible dose) du cecum de poulet (88). Il a été suggéré que certaines des protéines glycosylées par ce système étaient impliquées dans le mécanisme d'attachement bactérien aux cellules épithéliales de l'hôte. L'analyse mutationnelle des gènes impliqués dans ce système est difficile à interpréter en raison de l'effet pléiotropique engendré par de telles mutations.

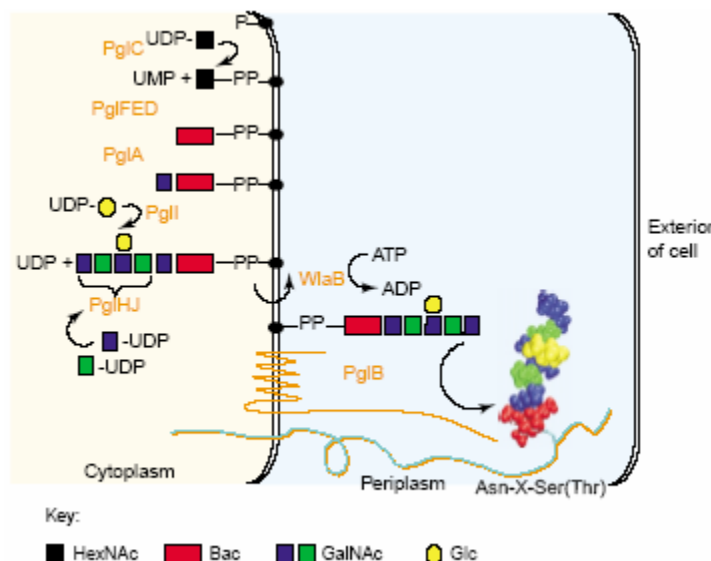


Figure 5 : Voie hypothétique de N-glycosylation de protéines chez *C. jejuni* (150).

2.5 Capsule

La capsule est une structure polysaccharidique communément trouvée à la surface des bactéries. Par sa localisation, la capsule représente le lieu d'interaction avec le milieu extérieur. Celle-ci est constituée de polysaccharides attachés de façon covalente à la membrane via une liaison phosphodiester. Les polysaccharides capsulaires peuvent également être trouvés libres dans le milieu de culture suite à une rupture d'une liaison phosphodiester trop faible. Les types de polysaccharides rencontrés sont assez hétérogènes parmi les différentes espèces de bactéries, ainsi qu'au sein d'une même espèce. Par exemple, 80 différents types de polysaccharides ont été identifiés chez *E. coli* (120). Cette diversité peut être expliquée par les propriétés biochimiques des

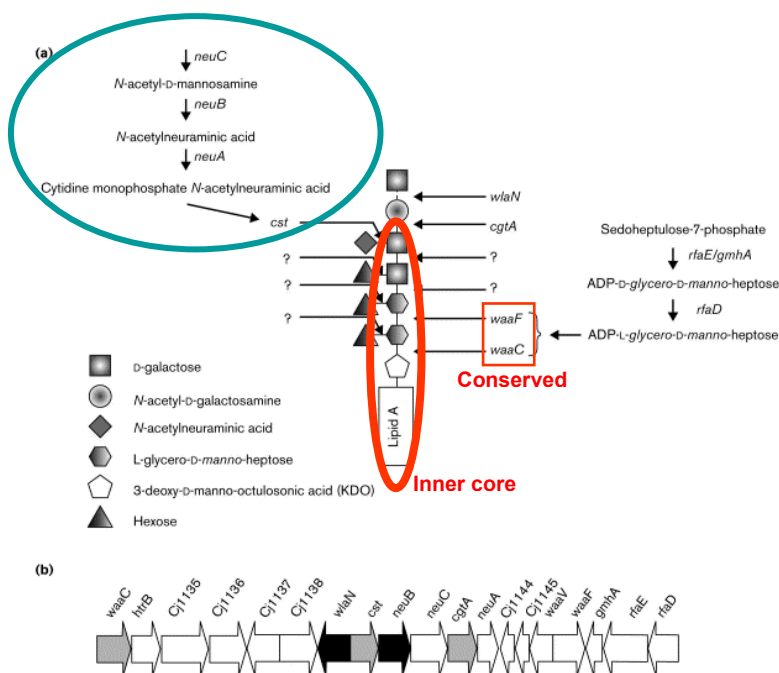
polysaccharides: ces structures glucidiques sont constituées d'homo- ou d'hétéropolymères d'unités de monosaccharides. Chacune d'elles est liée à sa voisine via une liaison glycosique impliquant un groupe hydroxyle. La possibilité de combinaison via de tels groupes explique la diversité de polysaccharides rencontrée. La structure de la capsule confère de nombreux avantages autant aux bactéries pathogènes que non pathogènes. La structure glucidique permet une rétention d'eau très importante (95% de la masse de la capsule). Cette barrière permet une prévention contre la dessiccation. Le processus de dessiccation a été montré chez *E. coli* comme un signal d'induction des enzymes impliquées dans la biosynthèse de la capsule. Les composants glucidiques de la capsule confèrent également une adhérence bactérienne aux cellules hôtes, ainsi que la formation de biofilm. Une autre propriété de la capsule est son implication dans la résistance au système immunitaire de l'organisme hôte. La capsule limite l'activation de la cascade du complément via deux différentes actions : la capsule masque les antigènes bactériens nécessaires à l'activation du complément et également limite l'accès des molécules actives du complément à la membrane bactérienne. La charge électrique présente au niveau de la capsule a été associée à la résistance à la phagocytose. De plus l'hyper-variabilité des structures capsulaires favorise l'évasion du système immunitaire de l'hôte. La première démonstration de la présence d'une capsule chez *C. jejuni* a été faite en 2000 par Karlyshev *et al* (89). Cela a été démontré par différentes évidences suggérant une structure capsulaire au lieu de lipopolysaccharides (LPS), comme suggéré auparavant (126). La démonstration a été faite entre autre par homologie de séquence de gènes impliqués dans ce processus avec les gènes de capsule *kps* d'*E. coli*, par analyse mutationnelle de ces gènes et par analyse structurelle de l'attachement des polysaccharides. Une réaction enzymatique démontre l'attachement des structures de polysaccharide capsulaire sur un phospholipide caractéristique. L'année suivante, le même groupe publia la première visualisation par microscope électronique d'une capsule d'une souche de *C. jejuni* par la coloration au bleu d'alcan (91). Depuis, la structure de base des composés polysaccharidiques a été déterminée mais peu de paramètres sont connus à propos des gènes impliqués dans la biosynthèse de cette structure. L'organisation génétique des gènes de capsule de *C. jejuni* NCTC 11168 apparaît être la même que celle décrite pour les bactéries Gram négatives présentant des capsules du

groupe B. Chez ces bactéries, le locus des gènes de biosynthèse de capsule est bordé par les gènes de transport des composants capsulaires (167). Ce locus, chez NCTC 11168, est représenté par les ORFs Cj1413c à Cj1448c (gènes *kps*). Au cours du séquençage du génome de NCTC 11168, plusieurs ORFs appartenant à ce cluster se sont avérées comme étant hyper-variables (Cj1420c, Cj1421c, Cj1422c, Cj1426c, Cj1429c), ou présentant des répétitions GC (Cj1437c) (123). Les polysaccharides constituant la capsule sont maintenant reconnus comme étant le principal antigène stable à la température impliquée dans la méthode de sérotypage « Penner » (18). Un mutant *kpsM*, gene impliqué dans l'exportation des composants de capsule chez *C. jejuni*, démontre une capacité réduite d'attachement et d'invasion in vitro des cellules INT 407 Eee. De plus, ce mutant présente une virulence réduite chez le modèle in vivo du furet (18). Récemment, il a été également montré que ce mutant perd la capacité de coloniser les intestins de poulet (85).

2.6 Lipooligosaccharide

Les LOS présentent une structure bipolaire. Chez *C. jejuni*, ces structures ont pour longtemps été considérées comme étant des LPS, mais l'absence d'une répétition d'une structure de type antigène-O ont permis de les caractériser en tant que LOS (100). Ces structures sont connues comme étant impliquées dans les mécanismes de résistance au sérum et dans l'évasion du système immunitaire de l'hôte (64). Chez *C. jejuni*, les LOS sont postulés comme étant à l'origine du développement des syndromes de Guillain-Barre et de Miller-Fisher, par mimétisme avec des structures gangliosidiques humaines (110). Ce mimétisme entraîne le développement d'une réponse auto-immune. Récemment, il a été montré que la liaison de la partie saccharidique sur le lipide était impliquée dans le déclenchement de la réponse auto-immune (134). Un certain nombre de types potentiels de LOS mimant les structures gangliosidiques humaines et induisant le syndrome GBS ont été caractérisés : GM1 (Figure 6), GM2, GD3, GD1a, GT1a, GQ1b (61). La présence de résidus d'acide sialique sur les LOS a été montré comme étant responsable dans le mime des gangliosides (62). L'étude de la composition chimique des LOS et des gènes impliqués dans leur biosynthèse a débuté très tôt, bien avant la publication du génome de NCTC 11168. Chez *C. jejuni* NCTC 11168, La biosynthèse

des LOS est représentée par les ORFs Cj1131 à Cj1152. Les gènes impliqués dans la biosynthèse de la partie externe (variable) présentent des



Current Opinion in Microbiology

Figure 6 : Schéma de biosynthèse d'un LOS de type GM 1 produit par la souche NCTC 11168 (a) et locus génétique correspondant (b) (101)

répétitions GC résultant en une possibilité de variations de structure au sein même d'une souche (61). Le gène Cj1139 appelé *wlaN* chez *C. jejuni* NCTC 11168 (similaire au gène *cgtB* d'*E. coli*) est une bêta-1,3 galactosyltransférase impliquée dans la biosynthèse des LOS (100). L'analyse biochimique démontre que cette enzyme est responsable de l'addition de bêta-1,3 galactose à une structure GM2, résultant en une structure GM1 (100). Si ce gène est muté, la bactérie présente une accumulation de LOS de type GM2 uniquement. Le passage du type GM2 au type GM1 se fait spontanément au cours de subcultures par perte d'un résidu G dans la zone de répétition GC.(100). Gilbert *et al* en 2002 (61) ont entrepris l'étude des clusters de biosynthèse des LOS chez 11 souches de *C. jejuni*. Cette étude a permis d'identifier au moins cinq paramètres conduisant à une variation des structures LOS: 1) la présence de répétition GC, 2) l'inactivation de gènes

par délétion, 3) l'insertion d'un nucléotide engendrant une enzyme non fonctionnelle, 4) la complémentation ou 5) la mutation de gènes engendrant de nouvelles propriétés (ex: substrat enzymatique). Les 11 souches testées présentent un total de 8 différents mimétismes de gangliosides. Parmi ces souches, il a été possible de reconnaître 3 classes de cluster LOS sur la base de leur composition et de leur organisation génétique (Figure 7) (61). Par l'étude de la souche TGH 9011, nous avons postulé un éventuel quatrième type (130). Les clusters LOS identifiés paraissent être encadrés par des gènes impliqués dans la biosynthèse de la partie interne des LOS, les gènes *waaC* et *htrB* à une extrémité, et les gènes *waaV* et *waaF* à l'autre extrémité.

Les gènes *waaC* et *waaF* codent pour des heptosyltransférases impliquées dans l'attachement du premier résidu heptose sur la structure Kdo attachée sur le lipide A. Le gène *htrB* code pour une enzyme de transfert d'acyltransférase sur le lipide A et *waaV* code pour une potentielle glycosyltransférase. Sa fonction reste cependant à démontrer. Quelques variations peuvent être notées à différentes positions de gènes impliqués dans la synthèse de la partie externe variable des LOS. Les gènes impliqués dans la biosynthèse (*neuA*, *B*, *C*) et transfert (*cstI*, *II*) d'acide sialique sont présents dans la plupart des souches étudiées. Quelques exceptions sont rencontrées dans la quatrième classe postulée, représentée chez les souches *C. jejuni* NCTC 11828 et TGH 9011 (130). La souche *C. jejuni* TGH 9011 est connue comme étant dénuée de LOS sialylé (15). L'analyse du cluster LOS démontre l'absence des gènes *neuA*, *B*, *C* et *cst*, gènes impliqués dans ce processus. La sialylation des structures LOS est connue comme étant un facteur important dans le mimétisme des structures gangliosidiques humaines (62). La sialylation est un facteur important de résistance au sérum et à l'évasion du système immunitaire (134).. Les mutants ne présentant pas ces structures ont été montré *in vitro* comme étant plus sensibles au sérum mais en aucun cas affectés dans le pouvoir d'invasion de cellules épithéliales *in vitro* (65).

Il serait intéressant de savoir si *C. jejuni* TGH 9011 est capable d'induire le syndrome GBS (130). Le gène *galE* a été caractérisé chez la souche 81116 en tant que UDP-glucose 4-épimerase catalysant l'interconversion d'UDP-galactose en UDP-glucose (59).

2.7 Toxines

Plusieurs types de toxines ont été caractérisés chez *C. jejuni* mais la production de la plupart d'entre elles est controversée. Deux classes de toxines bactériennes ont été décrites: entérotoxine et cytotoxine (162). Ces deux classes sont différenciées par leur mode d'action. L'enterotoxine est sécrétée par la bactérie et entre dans les cellules cibles de l'hôte via un récepteur spécifique. Cette entrée entraîne l'élévation d'AMP cyclique cellulaire (161). Les cytotoxines sont sécrétées par la bactérie. Une fois en contact avec la cellule hôte, cette cytotoxine a pour but de tuer les cellules de l'hôte (161). La production d'entérotoxine par des entéropathogènes a été suggérée être associée à des diarrhées aqueuses alors que la production de cytotoxines est associée au développement de diarrhée inflammatoire (161). Néanmoins cette hypothèse reste controversée par l'observation de bactéries asymptomatiques présentant néanmoins une activité cytotoxique, ce qui suggère que le type de diarrhée développée serait également dépendant du patient.

La production d'une entérotoxine de type choléra-like (CLT) par *C. jejuni* a été décrite pour la 1ère fois en 1983 par Ruiz-Palacios *et al.* (140). Les études suivant cette découverte restent cependant obscures, remettant en question la production d'une telle toxine (92).

C. jejuni a été également référencée comme produisant un hémolysine (162). La production de cette toxine a été principalement associée à des colonies vieillissantes: possibilité de libération de protéines cytoplasmiques conférant cette activité aux cellules mortes. Ces résultats restent obscurs et nécessitent d'avantage d'études.

A l'heure actuelle, *C. jejuni* est associé à la production d'une seule toxine très bien caractérisée : la « cytolethal distending toxin » (CDT).

En 1988, Johnson et Lior (84) sont les premiers à décrire la production d'une nouvelle toxine par certaines souches de *Campylobacter spp*. Cette toxine présente la particularité d'induire une augmentation du volume de cellules de hamster chinois (lignée CHO) in vitro, puis leur mort cinq jours après la mise en contact (84). Des études supplémentaires entreprises par le même groupe ont permis de caractériser la production de cette toxine par un grand nombre de bactéries entéropathogènes incluant *Campylobacter spp*. (84).

Les premières études génétiques relatives à la production de CDT ont été publiées en 1996 par Pickett *et al.* à partir de la souche *C. jejuni* 81-176 (129). Ces études ont permis d'identifier la présence de 3 gènes adjacents (gènes *cdt A*, *B* et *C*) présentant une forte similarité avec les protéines correspondantes de souches d' *E. coli* (129). Ces gènes sont également présents dans la souche NCTC 11168 (ORFs Cj0079, Cj0078, Cj0077).

La toxine CDT de *C. jejuni* induit l'augmentation de volume de cellules épithéliales intestinales par accumulation d'une forme inactive de la protéine CDC2 phosphorylée, enzyme clef régulant l'entrée en mitose (38). Des cellules épithéliales eucaryotes traitées *in vitro* avec CDT sont bloquées en transition G2/M du cycle cellulaire (96). Des travaux basés sur l'analyse des similarités au niveau des résidus d'acides aminés, montrent une relation de la protéine CdtB avec la famille de DNaseI (96). Des tests *in vitro* ont montré que la protéine CdtB purifiée présente une activité nucléasique (96). L'injection de cette protéine purifiée dans des cellules épithéliales *in vitro* était suffisante pour obtenir une augmentation du volume cellulaire. Ces résultats laissent penser que la protéine CdtB posséderait une activité nucléasique qui endommagerait l'ADN de la cellule hôte en jouant le rôle de nucléase, bloquant le cycle cellulaire au point de contrôle G2/M (Figure 8 et Figure 9). Les études menées sur *Actinobacillus actinomycetemcomitans* ont permis d'identifier la partie N-terminale de la protéine CdtB comme étant impliquée dans le transport actif de cette protéine dans la partie nucléaire de la cellule hôte (117). La partie C terminale de CdtB a, quant à elle, été identifiée comme étant responsable de l'activité nucléasique de CdtB (116). Le rôle exact de CDT reste obscur.

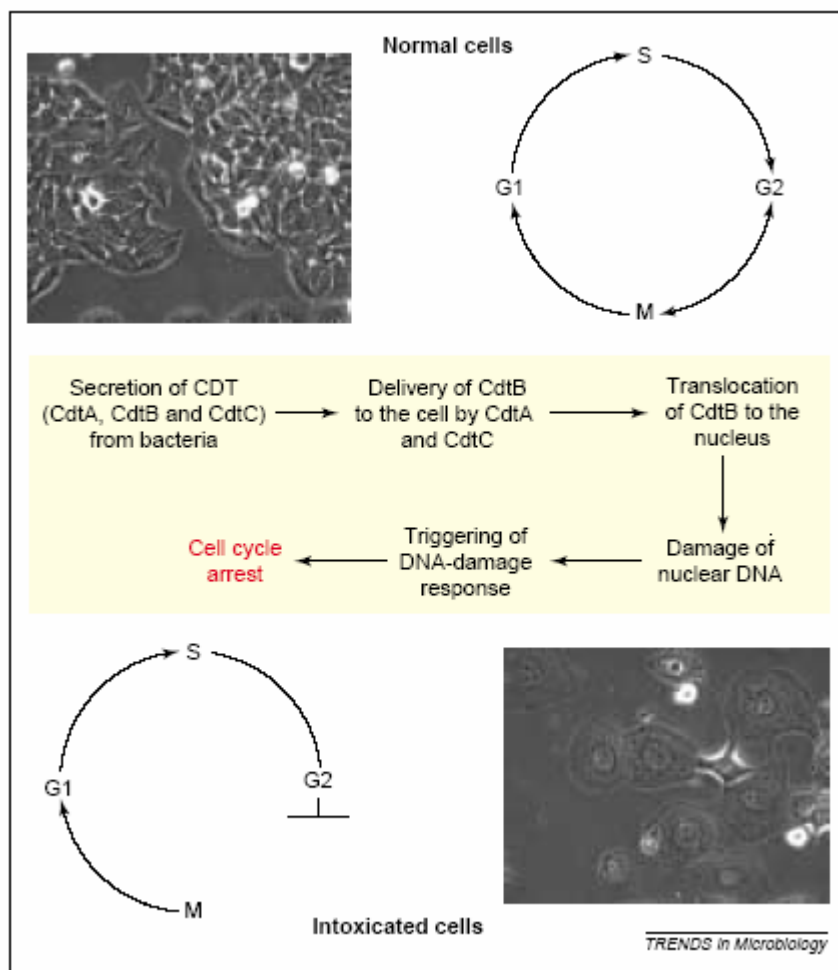


Figure 8 : Modèle du mécanisme d'action de la toxine CDT. (97). La protéine CdtB est délivrée au sein de la cellule hôte via CdtA et C. CdtB migre jusqu'au noyau ou elle exerce des cassures au niveau du chromosome menant à l'arrêt du cycle de division cellulaire.

Pourquoi la bactérie crée-t-elle de faibles cassures de l'ADN de la cellule hôte, bloquant celle-ci dans son cycle cellulaire ? Une hypothèse, formulée par Lara-Tajero et Galan (95) est que l'arrêt de multiplication des cellules épithéliales intestinales, ou du système immunitaire, permettrait une meilleure colonisation de la bactérie.

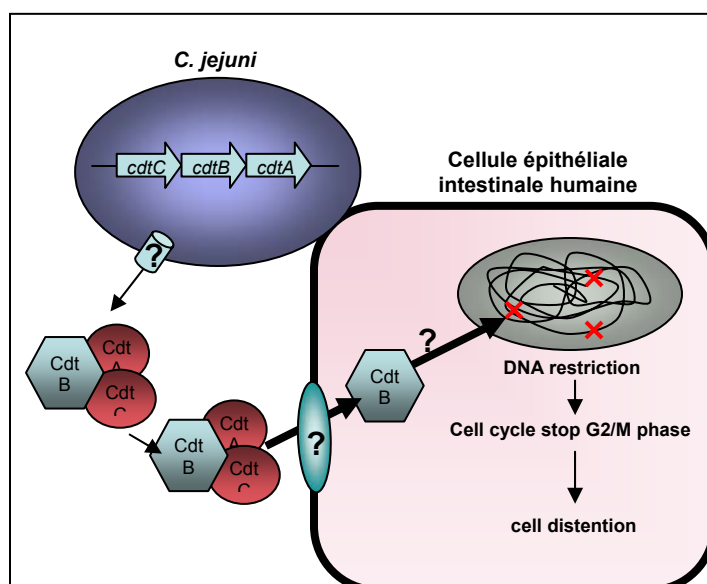


Figure 9 : schéma hypothétique récapitulatif de l'action de CDT sur les cellules épithéliales humaines

L'action de CDT produit par les bactéries pourrait influencer le renouvellement cellulaire, augmentant ainsi le temps de contact pour leur attachement. Malgré ces études, de nombreux points restent obscurs comme par exemple les mécanismes d'exportation de ces complexes de protéines par la bactérie, le type de récepteur reconnu par CdtA et B, et les mécanismes de translocation.

3. Vue d'ensemble des techniques comparatives génomiques : cas des bactéries entérotoxigènes.

3-1 Introduction

A la différence d'organismes à reproduction sexuée, la définition d'espèce dans le règne bactérien reste subjective. Les espèces bactériennes sont définies par le biais de plusieurs critères génomiques tels que l'hybridation ADN-ADN, les séquences ARN ribosomiques et les critères phénotypiques. Cette approche est dite polyphasique (35,158). Les souches

bactériennes présentant une hybridation ADN-ADN supérieure ou égal à 70% sont en général classées au sein d'une même espèce (158). Par ce critère, la diversité génétique de deux souches appartenant à la même espèce peut donc ainsi correspondre à 30% de la taille du génome de cette espèce. Dans le cas de bactéries présentant de larges génomes (8 Mb) comme *Burkholderia cepacia*, la diversité au sein de cette espèce peut atteindre en théorie jusqu'à 2.4 Mb, correspondant à environ 2200 gènes, soit par exemple la moitié du génome de la souche *E. coli* K12. Il est facile d'imaginer les différences phénotypiques conférées par cette diversité génomique entre les souches bactériennes d'une même espèce. Par exemple, l'espèce *E. coli* présente une multitude de souches avec d'importantes différences pathogéniques : la souche *E. coli* K12 n'est pas pathogène pour l'homme, la souche O157:H7 est entérohémolytique (EHEC) et la souche CFT03 est uropathogénique (UPEC). La séquence chromosomique de ces différentes souches permet d'identifier 0.9Mb de séquence unique entre les souches K12 et O157:H7 et 0.6Mb entre les souches K12 et CFT073 (127).

Les bactéries sont soumises à de nombreuses modifications génétiques au cours de leur croissance. Ces modifications sont conservées si elles s'avèrent bénéfiques ou neutres pour la bactérie dans son environnement (14). Le potentiel d'adaptation des bactéries via leur plasticité génomique est connu sous l'appellation de « Fitness » (135). Différents mécanismes interviennent dans l'évolution du génome bactérien (118):

- mutation ponctuelle
- réarrangement (inversion, translocation)
- délétion
- acquisition d'ADN étranger

Ce dernier mécanisme est admis comme étant le principal mécanisme d'évolution bactérienne (18).

3-2 Qu'ont besoin les bactéries pour devenir pathogènes?

L'acquisition d'une pathogénicité apparaît être multifactorielle et implique une étroite relation entre la bactérie et l'organisme hôte : habilité pour la bactérie à pénétrer et se

répliquer dans l'organisme hôte; capacité à utiliser de nouveaux nutriments, à résister vis-à-vis du système immunitaire de l'hôte et habilité à induire des signes pathologiques ou à tuer l'organisme hôte. Des études récentes de comparaison entre bactéries pathogènes et non pathogènes appartenant à une même espèce permettent d'aboutir à la conclusion que les bactéries pathogènes évoluent à ce statut par acquisition de larges structures d'ADN étranger dans leur génome mais également par perte de gènes (119). Par exemple, les souches de *Shigella* responsables de dysenterie ne présentent pas le gène *ompT* présent chez les souches non pathogènes d'*E. coli*, proche phylogénétiquement. L'introduction de ce gène par génie génétique chez les souches de *Shigella* supprime la virulence de ces souches par blocage de sa propagation intercellulaire (113). L'acquisition de larges structures d'ADN étranger au niveau chromosomique chez les souches pathogènes peut être reconnue comme un transfert d'îlots de pathogénicité (PAI) (68). Mais l'acquisition de séquences étrangères peut également se faire par acquisition de plasmides (169).

3-2-1 Acquisition de pathogénicité par des plasmides

Le genre *Yersinia* est représenté par 11 espèces (128). Trois d'entre elles sont pathogènes pour l'homme: *Y. pestis*, *Y. pseudotuberculosis* et *Y. enterocolitica* (les 2 dernières espèces citées sont de type entéropathogène) (66). La première est responsable de trois pandémies de peste humaine (128). Un représentant de cette espèce, la souche CO92, a été séquencée en 2001 (124). Un second représentant a été séquencé en 2002 (souche KIM) permettant la comparaison de deux génomes de bactéries appartenant à une même espèce (42).

Y. pestis est transmis par la piqûre de puces alors que les deux autres espèces sont transmises par ingestion orale à partir d'une source fécale conduisant à des gastro-entérites chez l'homme (169). Toutes ces souches ont l'habilité à se développer dans les tissus lymphoïdes humains, causant des dommages systémiques (128). Au niveau génétique, ces 3 souches présentent un large plasmide pYV appelé aussi pCD, conférant à

ces bactéries leur pathogénicité. Ce plasmide porte les gènes impliqués dans la sécrétion des protéines plasmidiques YOPs dans les cellules-hôte (169).

La sécrétion des protéines YOPs est calcium dépendante à 37°C. A une concentration supérieure à 2,5 mM en calcium, les bactéries ne secrètent pas les protéines YOPs (36). Les gènes *yop* sont dispersés sur le plasmide et apparaissent comme étant co-régulés. Jusqu'à présent, six protéines Yop ont été caractérisées. Ces protéines ont été impliquées dans la survie en milieu extracellulaire (inhibition de la phagocytose par les macrophages) ainsi que dans la prolifération des bactéries dans les tissus lymphatiques de l'hôte. Les protéines Yop H,E,T,O ainsi que YopJ ont été montrées comme bloquant la réponse pré-inflammatoire des cellules infectées (156).

Y. pestis est reconnue comme contenant un autre large plasmide nommé pMT1 (96.2 kb) ainsi qu'un 3^{ème} plasmide relativement plus petit nommé pPCP1 (9.6 kb) (124). Ces plasmides sont apparemment présents seulement chez l'espèce *Y. pestis* et absents des deux espèces entéropathogènes. Le plasmide pMT1 est connu comme portant un gene présentant des similarités avec le gene de toxine murine Ymt, toxine agissant en tant que phospholipase D, nécessaire à la survie des bactéries dans leur vecteur animal (puce) (169). Si *Y. pestis* est curé de ce plasmide, les bactéries sont incapables de coloniser la puce (76). Néanmoins il a été montré que des séquences chromosomiques étaient également nécessaires à la colonisation de la puce, comme le locus *hms* nécessaire à la colonisation du pro-ventricule de la puce (77). Le plasmide pPCP1 contient un activateur de plasminogène Pla impliqué dans la dissémination de *Y. pestis* dans le système sanguin juste après l'infection subcutanée (146). Les plasmides pMT et pPCP1 ne sont pas suffisants pour la transmission de la puce à l'homme mais peuvent expliquer la rapide évolution de l'espèce *Y. pestis* à ce processus d'infection (169). Les bactéries de type *Yersinia* sont un bon exemple d'acquisition de plasmides, résultant en une rapide évolution vers la pathogénicité.

3-2-2 Acquisition de pathogénicité par un PAI

3-2-2-1 Définition d'un PAI

L'augmentation du nombre de séquences génomiques de bactéries pathogènes montre que l'information génétique est composée d'un coeur de séquences conservées codant pour des gènes de ménage indispensables à la bactérie (67). Du fait de leur importance, ces séquences présentent un faible taux de mutation, et un taux relativement stable de composition G-C et d'usage de codon (67). En plus de ces régions conservées, le génome bactérien possède des régions présentant plus de variabilité (% GC, usage de codon). Ces séquences sont à l'origine de la diversité phénotypique bactérienne (67). Elles représentent dans la plupart des cas le résultat d'une acquisition d'ADN étranger par transfert horizontal. Ces séquences sont dénommées des « îlots génomiques ». L'acquisition d'îlots est reconnue comme étant le mécanisme le plus rapide d'évolution bactérienne (39). Dans le cas d'îlot génomique portant des gènes conférant une pathogénicité à la bactérie, ces régions sont appelées « îlots de pathogénèse (PAI) ». Le premier PAI bactérien a été décrit à la fin des années 1980 chez une souche UPEC d'*E.coli* (144). Les PAI sont maintenant très bien connus et caractérisés chez un grand nombre de bactéries. Ils présentent les caractéristiques suivantes (67) :

- sont présents chez les souches de bactéries pathogènes mais généralement absentes des souches non pathogènes de la même espèce,
- contiennent au moins un gène conférant la virulence à la bactérie,
- présentent un pourcentage GC et un usage de codon différents du reste du génome (il est à noter que cette particularité est difficile à distinguer si la souche donatrice de l'ADN étranger présente un pourcentage GC similaire à la souche acceptrice),
- sont fréquemment localisés à proximité de gènes de tRNA (ces gènes apparaissent comme étant un point d'ancrage d'ADN étranger),
- forment de larges zones comprises entre 10 et 200 kb,
- sont souvent bordés de répétitions directes (DR), vestiges de leur insertion par recombinaison,

- peuvent également présenter des traces de gènes conférant la mobilité à des fragments d'ADN tels que des intégrases, transposases ou séquences d'insertion (IS). Ces séquences confèrent une instabilité aux PAI ce qui peut expliquer leurs différentes localisations dans des génomes de bactéries proches phylogénétiquement.
- Finalement, les PAI sont en général des structures hétérogènes reflétant de nombreux événements d'acquisition d'ADN étranger ou réarrangements.

Un certain nombre de gènes impliqués dans la virulence bactérienne ont été montrés comme pouvant être portés par un PAI comme par exemple des systèmes de sécrétion, des toxines, des invasines (144).

3-2-2-2 Système de sécrétion

Il existe cinq types de sécrétion bactérienne, quatre d'entre eux ont été montrés comme pouvant être portés par un PAI :

- Le type I de sécrétion, ou type de sécrétion ABC, correspond au transport d'une molécule spécifique dans le milieu extracellulaire. Un exemple de ce type de transport est l'opéron *hly* impliqué dans la synthèse, l'activation et le transport d'alpha-hémolysine chez une souche UPEC (uropathogénique) d'*E. coli* (26).
- Le type II correspond au système général de sécrétion de protéines des bactéries pathogènes et non-pathogènes. Certaines protéines effectrices contenant un signal N-terminal sont exportées à travers la membrane via ce système. Les gènes codant le transporteur de type II n'ont jamais été caractérisés comme étant portés par un PAI, mais ce système est responsable de l'excrétion de plusieurs protéines effectrices présentant une action pathogénique codées par un PAI (148).
- Le type III de sécrétion est un assemblage complexe d'au moins 20 protéines. Il a été suggéré que ce système dérivait du système d'assemblage des flagelles. Il est utilisé par des bactéries pathogènes pour l'injection de protéines effectrices dans les cellules hôtes eucaryotes. Le type III de sécrétion est rencontré chez de nombreuses bactéries pathogènes comme *Salmonella* (PAI SP-1, SP-2), *Yersinia*, *Shigella* ou UPEC d'*E. coli*, mais absentes des bactéries non pathogènes de ces espèces (144).

- Le type IV de sécrétion est similaire au type III. Il est impliqué dans la translocation de molécules dans les cellules-hôtes eucaryotes. Ce système a été largement étudié dans le cadre du transfert du plasmide T d'*Agrobacterium tumefaciens* dans des cellules de plantes. Ce système est codé par des gènes de PAI chez plusieurs bactéries pathogènes comme *H. pylori* (responsable de la sécrétion de la protéine Cag), *Bordetella pertussis*, *Brucella suis* (144).
- Le type V de sécrétion est connu comme étant un auto-transporteur. La synthèse du système de transport et le substrat sont synthétisés sous forme d'une pré-protéine. Ce système est reconnu comme étant codé par certains PAI d'*E. coli* et de *Salmonella enterica*.

3-2-2-3 Invasine, moduline et effecteurs

Codées par des gènes de PAI, ces molécules jouent un rôle dans la déstabilisation des cellules eucaryotes de l'organisme hôte, favorisant l'invasion bactérienne. Les *Salmonella spp.* sont reconnues comme étant la majeure source de bactéries entéropathogènes transmises par la nourriture aux Etats-Unis (FoodNet report, 2002). *Salmonella spp.* présentent au moins 5 différents PAI (SP-1 à SP-5). Plusieurs de ces PAI portent des gènes nécessaires à l'invasion. Par exemple les protéines IpaB et IpaC sont codées par SP-1 et ont pour cible la membrane eucaryote des cellules de l'hôte. IpaA s'attache aux cellules hôtes au niveau de la protéine vinculine, résultant en une depolymérisation des molécules d'actine F (155). Le PAI LEE (locus of enterocyte effacement) de EPEC (*E. coli* entéropathogénique) et EHEC (*E. coli* entérohémolytique) produit une protéine Tir qui est transférée au niveau de la membrane de la cellule hôte (50). Cette protéine est supposée être impliquée dans l'adhésion bactérienne.

3-2-2-4 Toxines

Les PAI ayant une origine bactériophagique sont connus comme portant différents types de toxines. L'alpha-hémolysine (HlyA) de bactéries UPEC est codée par un PAI et transportée via le système d'excrétion de type I. Cette toxine forme un pore dans plusieurs types de cellules eucaryotes résultant dans la lyse de ces cellules (166). Les toxines cholériques et diphtériques sont également codées par des gènes de PAI.

3-2-2-5 Système d'acquisition de fer

Pour les bactéries pathogènes de mammifères, la disponibilité en fer est un facteur critique pour leur croissance. Plusieurs bactéries pathogènes présentent des PAI codant pour des mécanismes additionnels d'acquisition de fer, comme par exemple l'acquisition et/ou la synthèse de sidérophore. Les systèmes impliqués dans l'incorporation et la biosynthèse de yersiniabactine ou d'aérobactine ont été montré comme étant codés par des gènes portés par des PAI chez *Yersinia spp.* (29) et *Shigella*, respectivement. L'incorporation d'haemin a été également démontrée comme étant codée par un petit PAI de 5 kb chez EHEC et *Shigella* (144). La pluralité de mécanismes d'acquisition de fer n'est pas une marque des bactéries pathogènes. Les bactéries non pathogènes présentent également différents systèmes d'acquisition. Dans ce cas, ces pluralités augmentent la « fitness » de ces bactéries. La présence de mécanismes additionnels chez les bactéries pathogènes est reconnu comme faisant partie de la virulence (68).

3-3 Méthodes d'étude comparative des génomes bactériens

3-3-1 Introduction

La première séquence complète d'un ADN chromosomique bactérien a été obtenue en 1995 via une approche dite de « shotgun » (55). L'utilisation de cette technique était sensée être vouée à l'échec par de nombreux scientifiques pour la simple et bonne raison que le reassemblage de mini-séquences génomiques était largement au dessus des

capacités de calcul des ordinateurs de l'époque. Le séquençage de *H. influenzae* démontra l'inverse, permettant ainsi l'application de cette technique toujours actuelle au séquençage de génome (57). Elle est largement utilisée pour le séquençage de génomes procaryotes, ainsi que pour les génomes eucaryotes plus complexes tels que le génome de la drosophile (4) ou même le génome humain (159).

Cette technique est basée sur le séquençage de fragments d'ADN issus d'une banque génomique représentative de l'organisme à séquencer, et cela sans carte physique du génome. Tout d'abord, un morcelage aléatoire de l'ADN génomique via une digestion enzymatique ou une coupure physique en fragments de petite taille (quelques kb) est réalisé. Ces fragments d'ADN sont ensuite clonés dans un vecteur plasmidique pour la construction d'une banque génomique représentative (shotgun). En théorie le génome à séquencer via cette technique doit être représenté de 6.5 à 8 fois pour obtenir en probabilité 99.8% de couverture du génome (57).

Après séquençage, chaque séquence d'ADN est coupée des séquences plasmidiques ainsi que des extrémités présentant une faible qualité, et les séquences ainsi traitées sont assemblées en contigs. Les programmes informatiques utilisés pour ces assemblages identifient toutes les possibilités de recouvrement des séquences nucléotidiques dans les deux orientations. Les programmes les plus couramment utilisés pour l'assemblage de génome sont par exemple PHRAP ou TIGR Assembler (34).

Après séquençage de l'ensemble des clones de la banque génomique, la séquence génomique reste incomplète du fait de la sous-representation de certains clones dans la banque génomique. Plusieurs zones de taille variable restent à combler. Cette étape représente la tâche la plus fastidieuse du séquençage de génome. Dans un premier temps, les extrémités de chaque contig sont utilisées à l'identification de clones de la banque génomique. Les clones ainsi identifiés sont reséquencés pour combler une région manquante sur le génome. Après ce premier filtre, les zones manquantes restantes sont comblées via deux stratégies: la première implique la construction d'une seconde banque génomique dans un vecteur différent de celui utilisé. Cette stratégie permet d'identifier les zones génomiques manquantes, principalement dues à l'instabilité de tels fragments dans la banque génomique de départ. Ensuite, des amorces correspondant aux extrémités

de chaque contig sont synthétisées et utilisées comme sonde pour une recherche de clones dans la banque génomique de départ. Les clones sélectionnés sont séquencés.

La seconde stratégie correspond à l'utilisation d'amorces complémentaires aux extrémités de chaque contig, couplage de chacune d'elles par paire et utilisation pour une réaction de PCR. Le séquençage des produits PCR permet de combler les zones manquantes. Le génome complet est annoté c'est à dire examiné dans le cadre d'une recherche de gènes. Cette recherche est faite par utilisation de programmes informatiques. La méthode la plus classique pour repérer les gènes consiste simplement à trouver des phases ouvertes de lecture (commençant par un codon d'initiation sur les six sens possibles et se terminant par un codon stop) supérieur à une taille donnée. Cette recherche peut être combinée à la détection de signaux nécessaires à la traduction, en particulier le site de fixation du ribosome (RBS) (72). Les programmes les plus élaborés tiennent compte du contenu G+C de la troisième position de la phase ouverte de lecture avec la distribution habituelle dans les gènes de l'espèce étudiée. L'un des programmes les plus utilisés pour cette tâche est le programme Glimmer (Gene Locator and Interpolated Markov Modeler) (143). Ce programme utilise un modèle dérivé du modèle de Markov utilisé à l'identification des segments de séquence qui présentent un comportement statistique similaire à celui des régions codantes.

La recherche d'ORFs est suivie par l'annotation fonctionnelle de ces ORFs par comparaison des protéines codées avec une banque de protéines complètes.

3-3-2 Apport du séquençage d'autres bactéries entériques pathogènes.

3-3-2-1 génome de *Yersinia pestis*

3-3-2-1-1 *Y. pestis* biovar Orientalis souche CO92

Yersinia pestis a été responsable d'épidémies majeures de peste au cours du Moyen-Age. Si cette maladie n'est plus un problème dans les pays développés, l'étude de cette espèce a été relancée pour son utilisation potentielle en tant qu'agent de bioterrorisme et aussi du fait de la ré-émergence de la maladie dans des pays en voie de développement. L'espèce

Y. pestis est divisée en 3 biovars (Antiqua, Mediaevalis et Orientalis) (66). La première souche séquencée en 2001 (124) appartient au biovar Orientalis. Cette souche a été isolée d'un cas fatal humain après contact avec un chat infecté.

Le génome de cette souche est composé d'un chromosome de 4,65 Mb et de 3 plasmides identifiés comme prenant part à la pathogénicité (cf 3-2-1 Acquisition de pathogénèse via des structures plasmidiques). Le chromosome présente un grand nombre d'IS (140 au total) et une disparité de pourcentage GC indiquant de nombreuses acquisitions d'ADN étranger par transferts horizontaux. Trois zones arborant un pourcentage GC différent sont bordées de séquences IS. Dans ce cas, ces disparités de pourcentage GC correspondent principalement à des événements d'inversion et de translocation au sein du chromosome bactérien. Dans le cas des inversions, il a été montré que celles-ci se manifestaient au sein d'une même culture, reflétant ainsi l'extrême fluidité du génome de *Y. pestis*. Cette propriété est en partie conférée par le grand nombre d'IS. L'effet de ces réarrangements sur la pathogénèse bactérienne reste inconnue (124).

Des îlots chromosomiques spécifiques tel que *hms*, confèrent la survie dans l'organisme-hôte. Un système de sécrétion de type III, similaire au PAI SP-2 de *Salmonella*, a été détecté en plus de celui porté par le plasmide pYV. Les substrats et le mécanisme de fonctionnement de ce système restent inconnus. Un PAI de type HPI a également été détecté dans ce génome. Ce PAI est d'une taille de 30 kb et il est localisé physiquement près d'un locus de tRNA. Ce PAI est impliqué dans les systèmes de biosynthèse et d'incorporation du sidérophore, la yersiniabactine (29). Il a été montré que les mutations portant sur le récepteur de ce sidérophore diminuent fortement la virulence de la souche. Ce PAI contient de nombreux DR (répétition directe) et IS. La présence d'intégrase dans ce PAI chez certaines souches peut expliquer son instabilité.

Environ 150 pseudogènes ont été trouvés. La majorité d'entre eux représente des gènes non-fonctionnels (gènes présentant des homologies avec des gènes codant pour LPS, invasine, adhésine, impliqués dans une vie bactérienne de type entéropathogène).

Par acquisition de gènes étrangers, *Y. pestis* a modifié son style de vie par rapport à son ancêtre pour évoluer vers une adaptation à la colonisation de la puce comme organisme

hôte, et a progressivement perdu des gènes impliqués dans un style de vie d'une bactérie entéropathogène (124).

3-3-2-1-2 *Y. pestis* biovar Mediaevalis souche KIM

Le second génome d'un représentant de *Y. pestis* a été publié par Deng et coll. en 2002 (42). La souche KIM appartient au biovar Mediaevalis. Cette souche est associée à la seconde pandémie de peste humaine (42). Le séquençage de ce génome donne l'opportunité de comparer le génome de deux souches de *Y. pestis* appartenant à deux biovars distincts et admises comme étant séparées d'un ancêtre commun depuis seulement quelques centaines d'années.

La souche KIM présente trois plasmides communs à ceux de la souche CO92. Ces plasmides apparaissent comme étant relativement conservés entre les deux souches. Le chromosome de la souche KIM est plus petit de 50kb de celui de la souche CO92 et avec une absence plus faible d'IS (séquence d'insertion). Ces 2 génomes partagent 95% d'identité du point de vue nucléique. La différence de 5% est principalement due à des événements d'insertions isolés, tels que la présence de nombreux IS et d'un opéron tRNA supplémentaire chez la souche CO92.

Du point de vue de la structure chromosomique de ces souches, plusieurs événements de recombinaison intra-chromosomique peuvent être détectés (42). Ceux-ci sont localisés au niveau de trois zones. Cette constatation n'est pas vraiment une grande surprise. La plasticité détectée sur le génome de la souche CO92 pouvait laisser penser à un tel résultat.

La souche KIM possède 6 zones présentant des homologies avec des séquences d'origine phagique. Jusqu'à présent, aucune fonction de pathogenèse n'a été attribuée à ces séquences. Le génome de la souche KIM, comparé à celui de la souche CO92, possède en plus un second gène impliqué dans la réponse au stress (*katG*). Cette particularité représente le résultat le plus surprenant de l'analyse comparative de ces deux génomes de *Y. pestis* (42).

3-3-2-2 Séquences génomiques d'*Helicobacter pylori*

Les génomes de deux souches appartenant à l'espèce *H. pylori* isolées de patients présentant des troubles gastriques ont été séquencés. La première en date, *H. pylori* 26695, a été séquencée en 1997 et la seconde, J99, l'a été deux ans plus tard (10,153). Le séquençage de cette dernière a été le premier exemple de séquençage d'une souche bactérienne ayant un représentant de la même espèce déjà séquencé. *H. pylori* J99 a été isolé en 1994 d'un patient présentant un ulcère du duodenum. La souche *H. pylori* 26695, quant à elle, a été isolée avant 1987 d'un patient anglais présentant une gastrite. Cette dernière a été repliquée plusieurs fois en laboratoire avant d'être séquencée (10).

La comparaison génomique démontre une organisation génétique similaire. Entre 6 et 7% du patrimoine génétique est unique à une souche donnée. La variabilité nucléique est observée très majoritairement sur le 3^{ème} nucléotide de codon n'engendrant aucune modification majeure au niveau protéique en raison de la dégénérescence du code génétique. Du point de vue architectural, la disposition des gènes sur le chromosome est similaire entre ces 2 souches (10).

La séquence du génome de la souche *H. pylori* 26695 a été à l'origine publiée comme contenant 1590 ORFs, nombre ramené ensuite à 1552 du fait que certaines ORFs de taille inférieure ou égale à 150 nucléotides, présentaient des codons stop dans la partie codante des gènes correspondants chez la souche J99. De telles ORFs ne codent vraisemblablement pas pour des protéines (10). La souche J99 quant à elle, présente 1495 ORFs dont 89 sont absentes du génome de la souche 26695. Les deux génomes présentent un faible nombre d'IS (inférieur à 20). Il est à noter qu'en 2001 Janssen et coll., se fondant sur des analyses de groupage global, réduisent le nombre de gènes uniques à 110 et 52 pour les souches 26695 et J99, respectivement (82).

Les deux souches présentent le PAI *cag* encadré par les mêmes gènes et présentant seulement des différences mineures. Inversion et translocation chromosomiques apparaissent à faible taux : 10 inversions/translocations de 1 à 83kb chez J99 seraient suffisantes pour obtenir une structure chromosomique semblable à celle de 26695. Dans la plupart des cas, ces événements peuvent être associés à la présence d'IS, DR et gène de restriction modification (RM). Une région présentant un faible pourcentage GC (35%) a

été nommée zone de plasticité du fait que cette zone contient environ 40% des gènes uniques de chacune des deux souches. Les gènes rencontrés dans cette zone favorisent son classement en PAI. Les gènes uniques, en dehors de cette zone de plasticité, sont disposés de façon opéronique (82), avec un quart de ces gènes présents en opérons de deux gènes.

Alm et coll. (10) ont montré que des gènes uniques sont retrouvés à une même localisation physique sur le chromosome des deux bactéries, menant ainsi à l'hypothèse que *H. pylori* a une plasticité d'incorporation d'ADN étranger plutôt limitée (11,11).

Les gènes uniques caractérisés, de l'une ou l'autre des deux souches séquencées, correspondent principalement à des fonctions R/M. Il a été montré que ces systèmes spécifiques sont tous fonctionnels alors que chez les systèmes R/M communs, environ 30% d'entre eux sont non-fonctionnels. Le génome de *H. pylori* présente des répétitions homopolymériques qui pourraient être impliquées dans la modulation de l'expression de gènes. Ces répétitions sont retrouvées par exemple au niveau de gènes codants pour des protéines présentes à la surface externe de la membrane cytoplasmique et dans les systèmes R/M.

Des travaux plus récents s'attachant à la révision de la classification des ORFs des deux génomes d'*H. pylori* séquencés, identifient 485 protéines, sur un total de 1447, présentant une différence de taille entre ces deux génomes (25). Une grande partie de ces protéines (201/485) présentent 1 à 10 acides aminés de différence. Ces protéines appartiennent au métabolisme de l'ADN, à l'enveloppe cellulaire et à divers processus cellulaires. Il est à noter que la variation de taille de protéines hypothétiques doit être principalement due aux différences d'annotation des ORFs correspondantes (différents codons « start » ou stop, voire erreurs de séquençage) (25,25). La comparaison des séquences génomiques de souches bactériennes proches phylogénétiquement présente l'avantage de mettre en lumière les problèmes (et/ou les erreurs) rencontrés au cours du séquençage de génomes.

En conclusion, les séquences des génomes des souches 26695 et J99 apparaissent comme étant très proches les unes des autres. Les variations ainsi que les inversions/translocations au niveau chromosomique font part de la micro-évolution de *H. pylori*.

3-3-2-3 Le génome de *Salmonella enterica* sérovar typhi

Les bactéries de type *Salmonella enterica* sérovar typhi sont des entéropathogènes de l'homme responsables de la fièvre typhoïde (54). Ces bactéries sont responsables d'environ 16 millions d'infections dans le monde par année, dont 500.000 cas mortels (54).

La souche CT18 est le premier représentant de cette espèce à avoir été séquencé (122). Une seconde souche, Ty2, a été séquencée deux ans plus tard, permettant une analyse comparative (43).

Nom de la souche	Taille (Mpb)	(G+C) %	Nombre de genes	Nombre d'IS	Nombre de plasmide
<i>Salmonella enterica</i> CT18 ⁽¹⁾	4,81	52,09	4599	32	2
<i>Salmonella enterica</i> Ty2 ⁽²⁾	4,79	52,05	4546	29	0
<i>Shigella flexneri</i> 2a str. 2457T ⁽³⁾	4,60	50,90	4084	284	0
<i>Shigella flexneri</i> 2a str. 301 ⁽⁴⁾	4,61	50,90	4434	314	1
<i>Yersinia pestis</i> CO92 ⁽⁵⁾	4,65	47,64	4012	144	3
<i>Yersinia pestis</i> KIM ⁽⁶⁾	4,60	47,64	4198	122	3
<i>Helicobacter pylori</i> 26695 ⁽⁷⁾	1,67	39,00	1590	17	0
<i>Helicobacter pylori</i> J99 ⁽⁸⁾	1,64		1495		0
<i>Escherichia coli</i> K12 ⁽⁹⁾	4,64	50,80	4288		1
<i>Escherichia coli</i> O157:H7 EDL933 ⁽¹⁰⁾	5,53	50,40			0
<i>Vibrio cholerae</i> N16961 ⁽¹¹⁾	4,03	47,30	3885		0
<i>Vibrio parahaemolyticus</i> 2210633 ⁽¹²⁾	5,16	45,40	4832		0

Tableau 1 : tableau récapitulatif des données numériques des génomes bactériens. ⁽¹⁾ Parkhill et coll. (122), ⁽²⁾ Deng et coll. 2003, ⁽³⁾ Wei et coll. (164), ⁽⁴⁾ Jin et coll. (83), ⁽⁵⁾ Parkhill et coll. (124), ⁽⁶⁾ Deng et coll. (42), ⁽⁷⁾ Tomb et coll. (153), ⁽⁸⁾ Alm et coll. (10), ⁽⁹⁾ Blattner et coll. (24), ⁽¹⁰⁾ Perna et coll. (127), ⁽¹¹⁾ Heidelberg et coll. (75), ⁽¹²⁾ Makino et coll. (104).

La souche CT18 présente un large plasmide impliqué dans la virulence et un second plasmide présentant des similarités avec le plasmide pMT1 de *Y. pestis*. La souche TY2, quant à elle, ne présente aucun plasmide. Dans les travaux de Deng et coll.(43) relatifs à la publication du second génome (souche TY2), aucune donnée ne permet d'apprécier le potentiel de virulence de cette souche. En supposant que les souches TY2 et CT18 présentent le même potentiel de pathogénicité, les gènes de virulence chez TY2 devraient être trouvés sur le chromosome. Cependant, la comparaison des deux génomes ne montre aucune différence majeure au niveau chromosomique. Ces deux souches partagent 4516 gènes alors que seulement 29 gènes sont uniques à la souche TY2. Du point de vue architectural, la différence majeure se situe au niveau d'une inversion, postulée être le résultat d'une recombinaison, entre deux loci de rRNA. De manière générale, la structure chromosomique des deux souches est nettement mieux conservée, comparée à celle des génomes de *Y. pestis*. Ce qui peut être expliqué par un faible nombre d'IS (environ 30) (43). Comparées aux souches de *H. pylori*, les souches de *Salmonella* ne présentent pas un grand nombre de systèmes R/M.

Un faible nombre de gènes présentent de petites insertions ou des mutations ponctuelles qui ne perturbent pas le cadre de lecture. Les différences majeures entre ces deux génomes se situent au niveau de l'inactivation de la fonction de gènes par le biais de pseudogènes (43). Quelques uns, présents dans l'une des deux souches, ont été supposés jouer un rôle dans la pathogénicité codant pour des protéines sécrétées ou des glycosyltransférases (43). Un pourcentage des gènes mutés menant à des pseudogènes chez les 2 souches pourrait être expliqué par une adaptation à la réponse immunitaire de l'hôte (43). L'accumulation de pseudogènes peut être également expliquée par une adaptation à un mode de vie particulier comme par exemple un mode de vie intracellulaire (124) ou la survie dans les intestins de l'homme(43). La comparaison de ces deux génomes montrent une très grande homologie de séquence et d'organisation génétique (43).

3-3-2-4 Génomes de *Shigella flexneri*

Shigella est responsable de 160 millions de cas de dysenterie, et d'environ 1,1 million de cas mortels par année dans le monde (94). La shigellose, maladie conférée par *Shigella spp.*, représente un véritable problème de santé publique dans les pays en voie de développement (94). Comme *Salmonella spp.*, *Shigella* est référencée comme étant pathogène uniquement pour l'homme (164). Cette bactérie présente des PAI chromosomiques participant activement à la pathogenèse de la bactérie. Le genre *Shigella* a été séparé du genre *Escherichia* au début des années 70, mais les analyses taxonomiques récentes placent *Shigella* dans le genre *Escherichia* (27). Le séquençage de la souche *S. flexneri* 245T apporte l'opportunité de comparer la relation phylogénétique avec *E. coli* et potentiellement mettre en lumière les bases génétiques à l'origine des différences pathogéniques et de modes de vie de ces bactéries (*Shigella* est un pathogène intracellulaire alors qu'*E. coli* est pathogène extracellulaire) (83).

Le séquençage d'une seconde souche de *S. flexneri*, la souche 2a301, isolée à une autre période, a été effectué la même année, apportant ainsi la possibilité de comparaison entre les deux souches (164).

Le séquençage de la souche 245T montre que son organisation chromosomique et sa composition génétique sont très proches de celle de *E. coli* K12 (non-pathogène), démontrant une fois de plus que ces deux genres ne devraient en faire qu'un (83). La comparaison des deux génomes de *S. flexneri* montre un génome de plus petite taille pour la souche 245T, différence principalement due à la présence additionnelle de séquences IS dans le génome de la souche 301. Les séquences sont très similaires mais toutefois présentent des ponctuations de différences nucléotidiques (1400 au total) et quelques réarrangements chromosomiques apparemment dus à la présence d'IS. Les îlots génétiques sont similaires entre les deux souches mais présentent des localisations chromosomiques différentes. Chaque souche présente ses propres pseudogènes arborant des fonctions tel que stockage de fer, glycosyltransférases putatives, transport de protéines etc. La présence de ces pseudogènes, différents chez les deux souches, montre une adaptation différente envers leur environnement. De façon générale aucune insertion/délétion majeure ne peut être notée entre ces deux génomes (164).

S. flexneri 245T présente une taille chromosomique de 4,6 Mb codant potentiellement 4084 gènes, avec 205 d'entre eux uniques à cette souche comparé à *E. coli* K12 (83). Comparé au génome d'*E. coli* O157:H7, la structure chromosomique de *S. flexneri* 245T est grossièrement similaire à celle de K12 avec des ponctuations d'îlots nucléotidiques, le nombre de ces îlots étant plus petit chez 245T par rapport à O157:H7 et majoritairement différents. Ces deux souches partagent uniquement 30 gènes en commun des gènes absents de K12 (83).

Le génome de *S. flexneri* 245T présente de multiples réarrangements comparé au génome de K12, incluant de larges inversions chromosomiques. Le génome de *S. flexneri* présente un record d'IS pour un génome bactérien (247 complets et 67 incomplets). Comparé au génome de K12, le génome de 245T présente 37 îlots codant au moins pour un gène qui ne soit pas une transposase. Le tiers de ceux-ci correspond à l'insertion de génome phagique. Un total de 372 pseudogènes a été identifié sur le génome de *S. flexneri* (83).

Tout comme *Y. pestis*, *S. flexneri* est connu comme portant un large plasmide. Ce plasmide de 218 kb est impliqué dans la colonisation de cellules épithéliales (164). *S. flexneri* contient également deux petits plasmides multi-copies. Il est intéressant de noter que quatre plasmides ont été trouvés dans la souche 2457T. Cette souche présente, en plus des trois plasmides cités précédemment, un plasmide de 165 kb nommée pSf-R27. Il correspond à un plasmide de type R proche du plasmide pMT1 de *Y. pestis* (164). Ce plasmide confère une résistance à la tétracycline. Après recherche de ce plasmide chez différentes souches de *S. flexneri*, la souche 2547T s'est avérée la seule à le porter, menant à l'hypothèse que ce plasmide ait été accidentellement introduit après isolation (164).

3-3-2-5 Séquençage du génome de la souche *E. coli* O157:H7

Cette souche entéropathogénique d' *E. coli* a été isolée à la suite d'une épidémie d'entérite due à la consommation de viande contaminée dans l'Etat du Michigan aux USA (139). On estime que cette souche possède un ancêtre commun avec la souche non pathogène de *E. coli* K12 datant d'environ 4.5 millions d'années (138). Le génome

d'E.coli O157:H7 a été publié en 2001 (127) et a permis l'identification des gènes spécifiques de virulence par comparaison avec la souche K12 non pathogène.

Le génome de la souche O157:H7 présente un chromosome de 5,5 Mb, soit plus grand de 0.89 Mb par rapport à celui de la souche K12. De façon générale, les deux génomes présentent une structure chromosomique identique sur 4,1 Mb, codant 3574 gènes. Cette structure commune est parsemée d'îlots d'insertion spécifiques à l'une ou l'autre des deux souches. La souche O157:H7 contient 1,34 Mb de séquences spécifiques, alors que la souche K12 n'en contient que l'équivalent de 0,53Mb. Perna et coll. (61) ont estimé la présence de 1387 gènes spécifiques à la souche O157:H7, incluant des gènes potentiellement impliqués dans la virulence de cette souche. Le génome O157:H7 présente 9 larges îlots (>15 kb) spécifiques. Ces îlots portent manifestement des gènes impliqués dans la virulence bactérienne comme ceux codant pour une toxine de macrophage, une toxine RTX, deux clusters d'uréase, une adhésine, un système de sécrétion de type III et des protéines potentielles d'excrétion, présentant des similarités avec une toxine impliquée dans l'invasion de *Salmonella* et *Shigella* (127). Les îlots spécifiques à O157:H7 de petite taille présentent des gènes impliqués dans la biosynthèse de fimbriae, l'incorporation de fer ou dans des systèmes d'adhésion autres que celui des fimbriae (127).

Les gènes spécifiques à la souche K12 présentent, pour 57% d'entre eux, une fonction attribuée par similarité avec des séquences protéiques connues. Il est à noter que la souche K12 porte un gène impliqué dans l'utilisation du citrate ferrique qui est considéré être un facteur de virulence. Il n'est pas très clair si K12 est pathogène pour d'autres organismes que l'homme, ou si ces séquences correspondent à des reliques de pathogénèse (127).

La différence de GC% des îlots spécifiques à l'une ou l'autre des souches laisse penser que la plupart de ces îlots ont été acquis de façon indépendante plutôt qu'hérités d'un ancêtre commun. Sur les séquences conservées entre ces deux souches, le polymorphisme nucléotidique est principalement localisé sur la 3^{ème} paire de base des codons. Ces

séquences conservées ont également été identifiées comme présentant des variabilités touchant les gènes impliqués dans la relation hôte-pathogène ou dans le système R/M. Ces régions pourraient correspondre à des événements de duplication de gènes chez l'ancêtre commun aux deux souches, avec une évolution différentielle des gènes dupliqués en fonction de l'environnement.

3-3-2-6 Séquences des génomes de *Vibrio paraheamolyticus* et *Vibrio cholerae*

Le choléra est reconnu pour être responsable de 7 pandémies depuis sa caractérisation en 1817 (48). Cette maladie se traduit par une dysenterie prononcée, causée directement par la toxine cholérique produite par les souches toxigéniques de *V. cholerae*. L'espèce *V. paraheamolyticus* est composée de bactéries marines à l'origine de gastroentérite impliquant communément des diarrhées inflammatoires et septicémiques (104). Le séquençage d'un représentant pathogène de chacune de ces deux espèces représente un bon modèle d'étude comparative génomique, permettant de mieux comprendre les mécanismes pathogéniques particuliers à chacune d'elles.

Le séquençage du génome de la souche *V. cholerae* El Tor N16961 révèle de nombreux facteurs de pathogénèse comme les gènes impliqués dans la production de toxine cholérique, d'hémolysine, protéase, lipase, et d'un système de sécrétion de type IV (74).

Les deux souches de *Vibrio* présentent toutes deux une paire de chromosomes d'une taille de 3,29 Mb et 1,88 Mb pour *V. paraheamolyticus* et 2,96 Mb et 1,07 Mb pour *V. cholerae*. La plus grande différence entre ces 2 souches concerne la taille du petit chromosome, plus grand de 0,8 Mb chez *V. paraheamolyticus*. Le plus petit des chromosomes, d'après le contenu de ses gènes, peut être considéré comme ayant un rôle dans l'adaptation environnementale des bactéries. Le petit chromosome de *V. paraheamolyticus* ne présente que très peu de séquences avec un GC% particulier ou la présence d'IS, suggérant que la différence de taille n'est pas due à l'acquisition de séquences étrangères, mais probablement à la délétion de gènes au cours de l'adaptation de *V. cholerae* (104).

Il apparaît que le réarrangement chromosomique prend part au niveau intra- mais également inter-chromosomique. La plupart des réarrangements intra-chromosomiques ne modifient pas la distance relative des gènes de l'origine de réplication. Il apparaît également que malgré de nombreux réarrangements apparents, la localisation physique des gènes est très conservée entre ces 2 souches. Une région particulière, nommée « super-intégron », est trouvée au niveau du génome de *V. cholerae*. D'une taille de 126 kb, elle contient 179 cassettes. Ce super intégron a été postulé comme étant un piège à gènes, absorbant de nouveaux gènes étrangers. Les gènes faisant partie de cette structure diffèrent totalement entre ces 2 souches (104).

Un PAI de 80 kb a été découvert au niveau du petit chromosome de *V. parahaemolyticus* (104). Ce PAI code pour une hémolysine thermostable, caractéristique de souches pathogènes (86). De nombreux autres gènes pouvant être impliqués dans la pathogenèse ont été également trouvés au niveau de ce PAI, comme par exemple des gènes impliqués dans la synthèse d'un système de sécrétion de type III pouvant potentiellement délivrer des protéines dans les cellules de l'hôte. Un autre ensemble de gènes codant pour un système de sécrétion de type III, a été également détecté en dehors du PAI. Finalement un certain nombre de gènes potentiellement impliqués dans la pathogenèse de *V. cholerae* ont été détectés en dehors du super intégron. Le séquençage de ces deux génomes apporte de nombreuses informations génétiques à propos de la virulence de ces souches (104). La comparaison permet l'obtention d'informations sur l'évolution et l'adaptation de ces organismes à leur environnement, mais n'apporte pas d'information supplémentaire flagrante à propos de leur virulence respective.

3-4-2-7 Comparaisons des génomes de bactéries du genre *Campylobacter*

Campylobacter jejuni NCTC 11168 a été la première souche du genre *Campylobacter* à être séquencée (123).

Le génome de cette souche est composé d'un chromosome circulaire de 1,64 Mpb présentant un nombre total de 1654 ORFs. Quatre vingt quatorze pour cent du génome code pour des protéines, en faisant à l'époque le génome connu le plus dense. Le pourcentage GC de *C. jejuni* NCTC 11168 est de 30,6%. Deux régions chromosomiques étendues, présentent un pourcentage GC relativement plus faible par rapport au reste du génome. Ces régions correspondent aux gènes codant pour des protéines impliquées dans la biosynthèse de LOS (25,4%) et de la capsule (26,5%).

Une des particularités du génome de cette souche est la quasi-absence de répétitions d'ADN : seules 4 séquences répétées sont présentes sur l'ensemble du génome. Le génome ne porte aucune trace d'IS fonctionnel (à l'exception de *Cj0752* qui présente des similarités en partie avec IS605 d'*H. pylori*), de transposons ou de prophages. Une des autres particularités de ce génome est la faible proportion de gènes organisés en structure opéronique ou en clusters.

Le génome présente également un certain nombre de répétition de type GC. Ce type de séquences a été montré, chez d'autres bactéries, comme étant à l'origine de variations de structures antigéniques, créées par le glissement de l'ADN polymérase au cours de la réplication.

Récemment, quatre nouvelles souches de *Campylobacter* ont été séquencées, donnant l'opportunité de comparer ces séquences d'un point de vue génomique avec la séquence de la souche de *C. jejuni* NCTC 11168 (56).

Ces souches sont :

- La souche *C. jejuni* RM1221, cette souche a été isolée d'une carcasse de poulet et montrée comme étant capable d'envahir des cellules épithéliales in vitro, ainsi que de coloniser les ceca et l'épiderme de poulet (107).
- La souche *C. coli* RM2228, également isolée d'une carcasse de poulet, a été sélectionnée pour ses propriétés de multirésistance aux antibiotiques (152). Bien que les souches de *C. coli* soient des bactéries commensales du poulet, cette espèce présente une prévalence pour le cochon (106).

- La souche *Campylobacter upsaliensis* RM3195 est un isolat clinique provenant d'un enfant de 4 ans diagnostiqué comme porteur du syndrome GBS (106). Les souches de *C. upsaliensis* sont souvent retrouvées associées aux chiens et aux chats (69).
- La souche *Campylobacter lari* RM 2100 est un isolat clinique (152). Cette espèce est prévalente chez les oiseaux et peut être trouvée dans l'eau de source, dans l'eau de mer, et chez les crustacés (51).

Les données résultant du séquençage des génomes de *Campylobacter* sont regroupées dans le tableau 2

	Taille du chromosome (Mbp)	% GC	Nombre d'ORFs	Nombre de plasmides	Nombre d'IS intactes	Nombre de répétitions GC
<i>C. jejuni</i> NCTC 11168	1,64	30,55	1634	0	0	48
<i>C. jejuni</i> RM1221	1,78	30,31	1835	0	0	46
<i>C. coli</i> RM2228	1,68	31,37	1764	1	3	47
<i>C. lari</i> RM2100	1,5	29,64	1554	1	0	22
<i>C. upsaliensis</i> RM3195	1,66	34,54	1782	2	0	209

Tableau 2: synthèse des données de séquençage des génomes du genre *Campylobacter*.

La structure du génome de *C. jejuni* RM1221 présente une grande synténie avec celle de *C. jejuni* NCTC 11168, avec cependant la différence de présenter 4 zones de type prophagique/îlots génomiques nommés CMLP1, CJIE2, CJIE3 et CJIE4 (56). CMLP1 est présumé provenir de l'insertion d'un prophage de type mu. CJIE2 et CJIE4 présentant tous deux des gènes associés à des phages menant à l'hypothèse d'une origine phagique. CJIE3 est présenté comme un îlot génomique. Cette région codant pour 62 protéines, proviendrait de l'insertion d'un megaplasmide de type pCC178 de *Campylobacter*. La

majorité des protéines codées par les régions de type phagique ont de fonctions inconnues, rendant délicat l'attribution de fonction bactérienne à ce type de région (56).

La structure génomique de la souche de *C. coli* RM2228 présente une grande synténie avec celle de RM1221. Par contre les structures génomiques *C. lari* et *C. upsaliensis* apparaissent différentes de celle des souches de *C. jejuni* séquencées (56).

Tout comme *C. jejuni* NCTC 11168, *C. jejuni* RM1221 est caractérisé par l'absence d'IS intacte tout comme *C. lari* et *C. upsaliensis*, à la différence que ces deux dernières ne présentent pas la relique d'IS située à proximité du gène *tonB* chez les deux premières. La souche de *C. coli* RM2228 présente, quant à elle, 3 copies intactes d'IS de type IS605 sur son chromosome (56).

Toutes les souches de *Campylobacter* séquencées sont caractérisées par un certain nombre de répétitions GC (tableau 2). *C. upsaliensis* RM3195 présente un très grand nombre de telles séquences comparé aux autres souches. Sur un total de 209 séquences, soixante et une sont retrouvées au niveau d'ORFs codant pour des protéines hypothétiques, ou des protéines de structure ou de biosynthèse de l'enveloppe cellulaire, ou des protéines associées à la virulence de *Campylobacter*. Il n'est pas compris pourquoi un grand nombre de répétitions GC se trouvent au niveau de régions non codantes (56).

Les spécificités et différences entre ces génomes de *Campylobacter* se retrouvent au niveau d'ORFs codant pour des système R/M, des gènes de biosynthèse des LOS, ou des gènes impliqués dans la biosynthèse de la capsule ou de la résistance à différents antibiotiques.

Il a été trouvé chez *C. upsaliensis* RM3195 un nouveau locus putatif de virulence, présentant des identités de séquences avec les gènes d'*Haemophilus influenzae* (*licABCD*). Chez cette dernière, ce locus est impliqué dans la modification des LOS (165).

L'information générée par le séquençage de ces génomes a également été utilisée pour une comparaison phylogénétique de ces bactéries avec les autres bactéries de la clade epsilon déjà séquencées : *Wolinella succinogenes* DSMZ 1740, *H. pylori* J99, *H. pylori*

26695 et *Helicobacter hepaticus* ATCC 51449. Il apparaît que *C. coli* RM2228 est plus proche de *C. jejuni* RM1221 que de *C. jejuni* NCTC 11168 (56).

Le séquençage de ces génomes permet d'obtenir une vue plus globale des génomes de *Campylobacter*.

3-3-3 Hybridation soustractive (SSH)

3-3-3-1 Introduction

La technique d'hybridation soustractive (SSH) permet l'identification de séquences d'ADN uniques à une souche donnée (44). Comme nous l'avons vu dans la partie consacrée au séquençage de génome, deux souches appartenant à une même espèce partagent une grande partie de leur patrimoine génétique, par ailleurs ponctué d'événements d'insertion ou de délétion. Par comparaison avec une souche de référence, il est possible de focaliser sur les séquences uniques à la souche que l'on veut tester, cela sans séquençage complet de cette dernière. Si la souche de référence est une souche dont le génome a été déjà séquencé, la détection de faux-positifs correspondant à des séquences présentes chez les deux souches, en est facilitée.

3-3-3-2 Technique

L'hybridation soustractive peut être appliquée à partir d'ADN génomique, ou d'ADN-complémentaire (cDNA). Dans le cas d'utilisation d'ADN génomique (Figure 10), l'ADN de la souche à tester est fragmenté par une méthode physique ou enzymatique, en vue de l'obtention de fragments de petite taille (de 0.1 à 2 kb). Ces fragments sont ensuite « polis » (ajustés ? ? ?) à leurs extrémités.

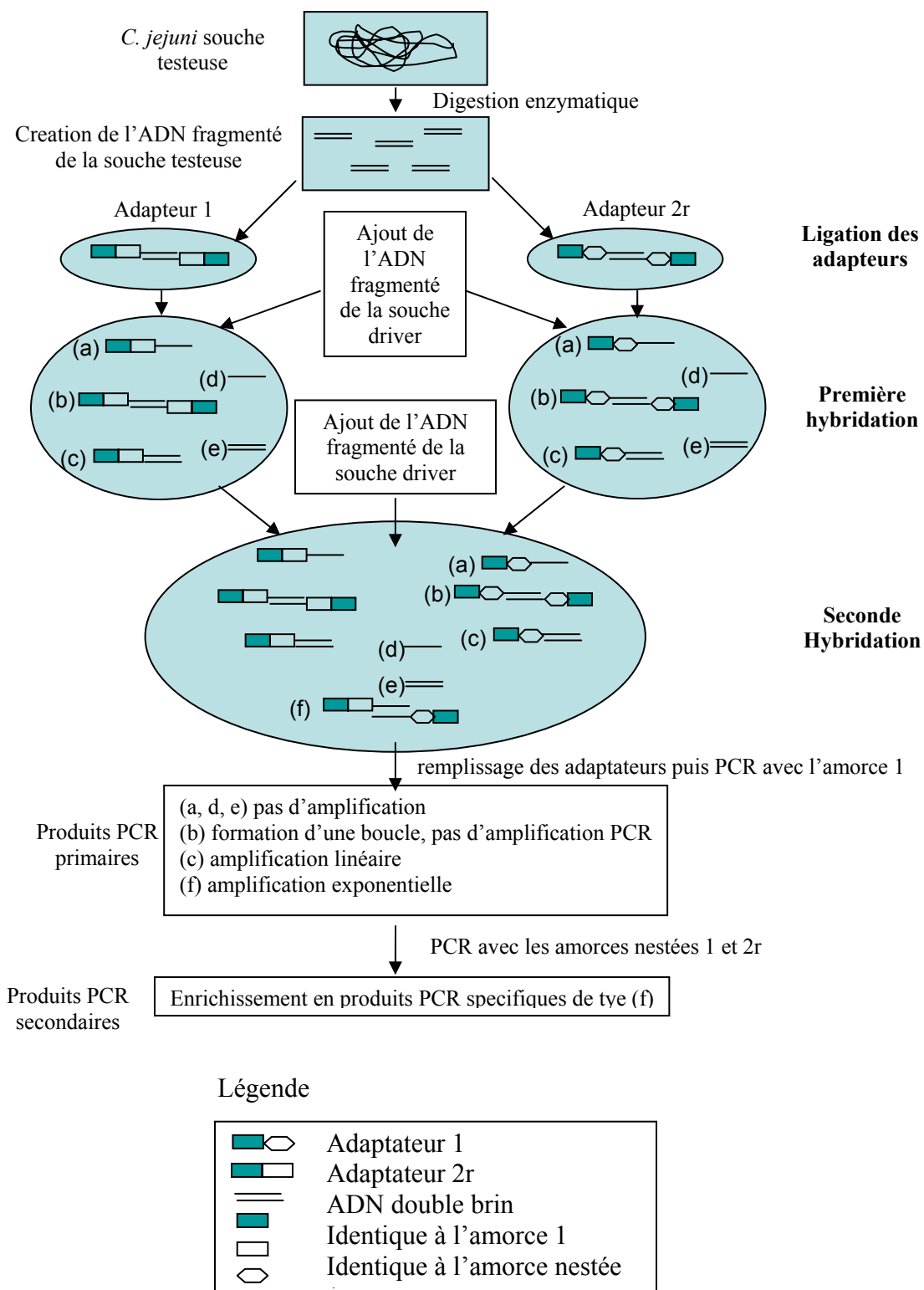


Figure 10 : Schéma explicatif de la technique d'hybridation soustractive.

L'ADN est ensuite divisé en deux lots marqués de façon indépendante avec des adaptateurs de nature différente, se liant à l'extrémité 5' des fragments d'ADN. Les deux types d'adaptateurs partagent une séquence commune reconnue par une amorce (par exemple A).

Les deux lots d'ADN sont mélangés séparément avec de l'ADN fragmenté de la souche référence en vue d'une première hybridation. Une seconde hybridation est effectuée par addition à un nouveau lot d'ADN de référence du mélange des deux lots d'ADN de la souche à tester. Les extrémités simple-brin de l'ADN à tester sont rendues double-brin et une PCR utilisant l'amorce de type A est faite sur ce mélange. Les fragments d'ADN double-brin issus d'un complexe des deux lots à tester seront amplifiés de façon exponentielle. Les hybrides correspondant à une hybridation « ADN à tester-ADN de référence » sont amplifiés de façon linéaire, alors que les hybrides entre ADN de référence ne sont pas amplifiés. Cette étape est suivie par une amplification PCR à l'aide des amorces spécifiques à chacun des adaptateurs.

3-3-3-3 Applications

La séquence de *H. pylori* 26695, publiée en 1997 (153), a permis la comparaison de séquences par SSH avec la souche J166, d'origine simièsque (7). Dix huit séquences ont été isolées comme étant uniques à J166. Sept de celles-ci sont impliquées dans des systèmes R/M et neuf d'entre elles n'ont pas de fonction attribuée. La comparaison des deux génomes de *H. pylori* séquencés (cf 3-4-2-3 Séquences génomiques d'*Helicobacter pylori*) nous apprend que 6 à 7 % de séquences sont uniques à une souche. Sur un génome de la taille de *H. pylori*, ceci représente théoriquement une cinquantaine de gènes, alors que dans l'étude de Akopyants et coll. (5), seules 18 ORFs ont été isolées. L'explication d'une telle différence pourrait être que les deux souches étudiées sont très proches phylogénétiquement, ou que le pourcentage de couverture de SSH est très faible (7).

Plus récemment Han et coll. (71) ont publié une expérience de SSH basée sur la comparaison de 2 souches de *H. pylori* non séquencées. L'une de ces souches a été isolée d'un patient présentant un ulcère à l'estomac et la seconde d'un patient asymptomatique. Cette étude a permis d'identifier 23 séquences uniques à la souche isolée du patient présentant un ulcère de l'estomac. Quinze d'entre elles se sont avérées appartenir à un plasmide de 3,5 kb (plasmide pHP0100). En conclusion, les deux études de SSH présentées permettent apparemment de mettre en évidence qu'une partie de la variabilité génétique des souches d'*H. pylori*.

En 2003, Agron et coll. (5) ont utilisé les deux souches d'*H. pylori* séquencées (J99 et 26695), chacune d'elles étant utilisée comme souche à tester, avec la mise en œuvre de quatre enzymes de restriction pour l'obtention des fragments d'ADN. Un total de 3264 et 3744 fragments, respectivement, ont été isolés comme étant uniques aux souches J99 et 26695. Respectivement 90 et 86% des gènes uniques des souches J99 et 26695 ont pu être identifiés, prouvant que la méthode SSH est une méthode fiable pour la découverte des séquences uniques au sein d'un génome bactérien. Il est tout de même à noter que le nombre de clones séquencés au cours de cette étude est très important ce qui rend cette étude difficilement comparable aux travaux de SSH précédemment publiés.

En 2002, Walker et Verna (160) profitèrent de la séquence de la souche non-pathogène *E. coli* K12 pour l'étude par SSH d'une souche de *Shigella flexneri*. Ces deux bactéries partagent 80 à 90% de séquences génomiques communes, alors que le genre *Shigella* est connu pour contenir de larges PAI. Ces deux particularités font de *S. flexneri* un bon candidat pour une étude par SSH. Les auteurs n'ont pas utilisé une SSH basée sur une amplification PCR comme décrit précédemment, mais une méthode par marquage de l'ADN de référence avec de la biotine. Cette technique est appliquée plus spécifiquement à la détection de large zone d'insertion comme des PAI. Après mélange, dénaturation et renaturation des ADN fragmentés des deux souches, l'échantillon est filtré sur colonne d'affinité contenant des molécules de streptavidine. Les séquences non retenues correspondent à des séquences uniques à la souche à tester. Ce passage permet donc d'éliminer les fragments d'ADN communs entre les deux souches. Après sept étapes de sélection, limitant la contamination par de l'ADN correspondant à des séquences

communes, les ADN sont attachés à un adaptateur et amplifiés par PCR. Enfin, les produits PCR sont utilisés comme sonde pour l'identification de clones d'une banque génomique cosmique de la souche *Shigella* à tester. Cinquante cosmides ont été sélectionnés de cette façon et trois seulement se sont avérés être uniques à la souche testée. Les inserts de ces 3 cosmides correspondent à un contig de 42 kb, présentant de nombreuses séquences de PAI. Cette étude représente un très bon exemple d'identification de PAIs chez une bactérie proche phylogénétiquement d'une bactérie non pathogène.

Une étude de SSH utilisant *E. coli* K12 en tant que référence et une souche aviaire pathogénique d'*E. coli* a été publiée en 2002 par Stocki et coll. (149) : 62 séquences uniques ont été identifiées sur 101 sélectionnées. Plus de la moitié de ces séquences présentent des relations avec des séquences phagiques ou plasmidiques. Ces séquences présentent une grande disparité de pourcentage GC menant à l'hypothèse que ces séquences pourraient être des PAI potentiels, acquis par transfert horizontal.

En 2002, Ahmed et coll (6) ont publié des travaux de SSH sur une souche de *C. jejuni* capable d'envahir les intestins de poulet (souche 81116), la souche de référence étant la souche NCTC 11168, séquencée et non-colonisatrice du poulet. Un total de 24 fragments uniques à la souche 81116 a été identifié. Six d'entre eux portent des gènes codant pour des protéines ayant des similarités avec des systèmes R/M, six autres avec des protéines hypothétiques, six autres avec des protéines de fonction inconnues, et deux autres présentant des similarités avec des gènes impliqués dans la résistance à l'arsenic. Les clones restants portent des gènes impliqués dans la synthèse de cytochrome c-oxidase III, dTDP-glucose 4,6-dehydratase, γ -glutamyl transpeptidase et une protéine impliquée dans la résistance aux infections phagiques.

La technique d'hybridation soustractive est une bonne méthode pour l'isolation de séquences uniques. Cependant, cette technique présente le désavantage de conduire aussi à un grand nombre de clones faux-positifs, ce qui représente un gaspillage de séquençage et un pourcentage de gènes uniques détectés, par rapport à l'ensemble des gènes uniques

du génome, généralement faible. De plus, la technique n'est pas appropriée à la détection de petits fragments génétiquement variables, comme des insertions ou des variabilités nucléiques. La technique ne donne également aucune indication sur la localisation de ces séquences uniques sur le génome bactérien.

3-3-4 Comparaison génomique par utilisation de puces à ADN

La disponibilité de séquences complètes de génomes bactériens donne l'opportunité d'étudier les différences génomiques et l'évolution de bactéries proches phylogénétiquement, par l'utilisation de puces à ADN. Si la technique permet d'étudier en une seule expérience l'expression de l'ensemble des gènes bactériens, les puces à ADN peuvent également être utilisées pour une hybridation compétitive entre l'ADN d'une souche bactérienne proche phylogénétiquement de la souche utilisée pour la construction de la puce et l'ADN de cette souche de référence.

Cette comparaison permet l'identification rapide des gènes présents dans le génome de la souche testée et de la souche de référence, ainsi que les gènes uniques à la souche de référence. L'analyse de souches appartenant à une même espèce permet de définir le patrimoine génétique commun à cette espèce. Ces études sont maintenant très répandues pour toutes les espèces bactériennes ayant au moins une souche dont le génome a été séquencé.

En 2000, Salama et coll. (74) ont utilisé une puce à ADN composite, portant les gènes des souches *H. pylori* J99 et 26695. Cette puce contient donc l'ensemble des gènes communs aux deux souches, ainsi que les gènes uniques à celles-ci. Cette puce a été utilisée pour l'étude de 15 isolats cliniques de *H. pylori* dans le but de détecter de nouveaux déterminants génétiques de virulence de l'espèce. En raison de la relative conservation entre les deux souches séquencées, l'étude permet de cartographier les gènes manquants au niveau des deux génomes. Deux zones, dites de plasticité (PZ), ont été ainsi détectées. Un PAI a également été montré comme étant une zone présentant une grande variabilité entre les souches. Les deux PZ ont été distinguées par leur différence de pourcentage GC. Elles représentent une mosaïque de gènes spécifiques à chacune des différentes souches, acquis par de multiples événements d'acquisition indépendante.

Cette analyse montre que le PAI est moins variable que les 2 PZ et qu'il correspond plus probablement à un événement unique d'insertion. Ces résultats suggèrent que le typage des souches d'*H. pylori* sur la base du gène *cag* du PAI, n'est pas le plus adapté. Effectivement, le PAI ne définit pas une lignée d'évolution. Les souches possédant ce PAI ne sont pas phylogénétiquement plus proches les unes des autres que des souches d'*H. pylori* dénuées de ce PAI. Malheureusement, le manque d'information phylogénétique concernant les souches cliniques utilisées ne permet pas une corrélation entre la présence ou l'absence de gènes de virulence, et l'origine géographique de ces souches.

Le même type de puces à ADN a été utilisé par Israel et coll. (81) pour une étude comparative de deux souches d'*H. pylori* induisant des symptômes cliniques différents chez le modèle gerbille de Mongolie. L'hybridation indépendante de l'ADN de ces deux souches avec un mélange d'ADN des deux souches d'*H. pylori* séquencées, permet l'identification des contenus génétiques des souches testées. L'analyse montre une large délétion du PAI atténuant l'induction d'IL-8 chez les cellules hôtes et réduisant l'inflammation gastrique, en accord avec les études montrant une délétion partielle ou complète du PAI.

L'utilisation de puces à ADN a prouvé son efficacité dans le cas de *Salmonella spp.* Le genre *Salmonella* est composé de plus de 2200 sérovars classés en sept sous-espèces de *Salmonella enterica* et 7 sous-espèces de *Salmonella bongori* (20). Deux représentants de la sous-espèce *S. enterica*, bactéries très fortement adaptées à leur hôte mammifère, ont été séquencés la même année, *S. enterica* sv. *Thyphimurium* LT2 (43) et *S. typhi* CT18 (122). Le génome de CT18 a été le premier à être utilisé pour l'étude de comparaison génomique via l'utilisation de puces à ADN. Un représentant des 7 sous-espèces de *Salmonella bongori* a été testé indépendamment (131) sur la base d'une puce à ADN de la souche LT2. La majorité des gènes manquants de la souche LT2 au niveau des différentes souches testées sont des séquences plasmidiques ou des gènes d'origine phagique. Certains groupes de gènes métaboliques sont absents chez différentes souches, probablement à cause d'une spécialisation des différentes sous-espèces à leur environnement. Basé sur la relation phylogénétique des souches testées, il apparaît que la

plupart des clusters de gènes sont distribués dans différentes sous-espèces, reflétant une acquisition indépendante ou l'héritage d'un ancêtre commun suivi de la perte des différents lignages au cours de l'évolution (131).

Le génotypage par utilisation de puces à ADN montre également des avantages dans le cas d'études de bactéries impliquées dans des pandémies. En 2002, Dziejman *et al.* (48) ont étudié l'évolution de l'espèce *V. cholerae* par utilisation d'une puce à ADN présentant l'intégralité des ORFs identifiés chez la souche séquencée El Tor 16961, isolée de la 7^{ème} épidémie de choléra. Cette puce a été utilisée pour une hybridation compétitive avec les ADNs de la souche El Tor prépandémique, de la souche pandémique El Tor, de la souche O139 récemment isolée et associée à des épidémies de choléra, et de souches de *V. cholerae* non-toxigéniques. Ces expériences avaient pour but d'identifier les gènes impliqués dans la virulence de la souche responsable de la septième pandémie de choléra. Les résultats montrent que le génome de la souche séquencée est proche de celui de la souche prépandémique ainsi que des autres souches étudiées. Une zone de 22 gènes uniques à la souche El Tor N 16961, a pu être isolée. La moitié de ces gènes est regroupée au niveau d'une zone présentant un GC% faible par rapport au reste du génome. Cette zone a été nommée "*Vibrio* seventh pandemic island I" (VSP-I). La plupart des gènes de VSP-I code pour des protéines à fonction inconnue. Certains de ces gènes révèlent de possibles séquences laissant penser à une possible acquisition via des éléments mobiles. Huit autres gènes uniques sont localisés au niveau d'une seconde zone présentant un pourcentage GC faible, nommée VSP-II. Comme VSP-I, la plupart de ces gènes codent pour des protéines à fonction inconnue. Par contre, l'un des gènes de VSP-II code pour une protéine présentant des similarités avec un facteur de transcription (48).

Cette étude permet l'identification des îlots VSP-I et VSP-II contenant probablement les gènes impliqués dans le pouvoir pathogène de la souche responsable de la septième pandémie. Elle démontre le pouvoir des puces à ADN pour une identification rapide et efficace de gènes spécifiques à des souches impliquées dans des pandémies récentes.

L'espèce *C. jejuni* présente plusieurs évidences de diversité génotypique et phénotypique (123). L'utilisation d'une puce à ADN représente une technique adaptée à la découverte

de la variabilité génotypique de cette espèce. Le premier exemple d'une telle étude a été obtenu par utilisation d'une puce à ADN construite à partir d'une banque génomique de la souche séquencée *C. jejuni* NCTC 11168 (46). Onze souches ont été testées vis-à-vis de la souche *C. jejuni* NCTC 11168, mettant en évidence les gènes communs à ces souches. Un total de 1300 gènes, soit 78.3% des gènes, sont communs à l'ensemble des souches. Ces gènes correspondent vraisemblablement aux gènes « de ménage » de l'espèce *C. jejuni*. Plusieurs des gènes de *C. jejuni* NCTC 11168 sont variables chez les différentes souches testées. Néanmoins, il est possible de détecter des zones chromosomiques présentant de grandes variabilités selon les souches. Ces zones contiennent des gènes impliqués dans la biosynthèse de structures exposées à la surface membranaire de la bactérie comme les gènes des flagelles, des lipooligosaccharide LOS, des capsules, ainsi que des gènes impliqués dans des systèmes RM et des gènes impliqués dans l'acquisition de fer.

Leonard et coll. (98,99) ont utilisé une puce à ADN correspondant au dépôt des produits PCR de l'ensemble des gènes de la souche NCTC 11168. Les souches de *C. jejuni* isolées d'épidémies cliniques sont communément typées par RAPD ou PFGE, associé à un sérotypage. Dans un premier temps, Leonard *et al* ont étudié 15 isolats cliniques de *C. jejuni* et ont comparé les résultats obtenus avec les résultats de RAPD. Les classifications de ces souches d'un point de vue phylogénétique, basées sur les résultats d'hybridation sur la puce à ADN mènent aux mêmes conclusions que celles obtenues par RAPD (98). L'hybridation sur les puces à ADN permet de distinguer de façon plus précise les souches testées par une technique de typage clinique. D'autres travaux menés par ces auteurs ont consisté à tester l'hybridation de différentes souches impliquées dans le développement du syndrome de Guillain-Barré sur la puce à ADN de NCTC 11168. Cependant, aucun profil particulier n'a pu être obtenu (99).

Dans le cas du séquençage du génome d'une souche hautement pathogène, la construction d'une puce à ADN correspondant à l'ensemble du génome de cette bactérie permet l'isolation rapide des gènes impliqués dans la virulence de cette souche. Malheureusement cette technique ne permet pas l'identification des gènes uniques à la souche testée non séquencée.

3-3-5 Puces à ADN de type « shotgun »

Une technique efficace et rapide, adaptée à l'analyse de souches bactériennes non séquencées, fait défaut. L'utilisation de puces à ADN à souches séquencées est d'un grand intérêt pour l'analyse des génomes de bactéries phylogénétiquement proches. Malheureusement, bien que mettant en lumière les différences génétiques au sein des bactéries testées, la technique n'apporte aucune information à propos des séquences uniques des souches testées. Ces séquences comme nous l'avons vu précédemment, confèrent la variabilité phénotypique des souches bactériennes. Ces différences s'avèrent très intéressantes dans le cas de souches pathogènes comme par exemple des souches cliniques ou des souches à l'origine de pandémies. La méthode SSH pourrait répondre à cette attente. Mais comme déjà décrit, cette technique souffre de plusieurs désavantages (3-4-2-3 Applications). Très récemment a été développé un nouveau type de puces à ADN correspondant au dépôt de l'ensemble des clones d'une banque génomique représentative d'une souche bactérienne non séquencée. Une telle puce à ADN est dite de type "shotgun" (73). La technique a été appliquée pour la première fois à *Plasmodium falciparum*, alors que le génome de cet organisme n'était pas encore complètement élucidé (73). Cette puce à ADN correspondait au dépôt de 3648 fragments aléatoires utilisés pour l'étude de l'expression génétique de *P. falciparum* au cours de ses différentes phases de développement. A la suite de cette étude, ce type de puce a été utilisé pour la comparaison de génomes et l'identification de gènes uniques pour des souches particulières non séquencées.

Succinctement et de façon pratique, les petits fragments de moins de 2 kb d'une banque génomique aléatoire représentative d'une souche bactérienne non séquencée, sont amplifiés par PCR à l'aide d'une paire d'amorces communes, hybridant avec des séquences du vecteur. Après purification, ces produits PCR sont individuellement déposés sur un support en verre pour la construction de la puce à ADN. Cette technique a été appliquée avec succès à quelques espèces bactériennes telles que *E. coli* (173), l'archéobactérie *Haloferax volcanii* (172) et *Haemophilus influenzae* (112).

Au cours des travaux développés dans ce manuscrit, nous avons amélioré et utilisé la technique de puces à ADN de type « shotgun » à l'analyse des deux souches *C. jejuni* TGH 9011 et 81-176, dans le but d'identifier les séquences uniques à ces souches par rapport à la souche séquencée NCTC 11168.

Chapitre II

Matériels et Méthodes

1 Souches bactériennes, lignées cellulaires et conditions de culture

Les souches *C. jejuni* ATCC 43431 (également référencée sous l'appellation TGH 9011) et *C. jejuni* NCTC 11168 ont été obtenues auprès de l'“American Type Culture Collection” (Manassas, VA 20108 USA) et de la “National Collection Type Cultures” (61 Colindale Avenue, London NW9 5HT, Angleterre), respectivement. La souche *C. jejuni* 81-176 a été gracieusement fournie par Pr Pickett. Les souches sont cultivées à 37°C en milieu Mueller-Hilton (MH; Gibco, USA) liquide ou solide (agar 15 g/L), ou en milieu biphasique correspondant à un flacon contenant un fond de milieu MH solide recouvert d'un volume équivalent de milieu MH liquide. Ce milieu permet l'obtention d'une croissance de *C. jejuni* plus élevée qu'en milieu liquide. La phase solide de ce milieu permet un attachement des bactéries, augmentant ainsi le rendement de la culture. Tous les milieux sont stérilisés à l'autoclave pendant 20 min à 120°C. Les cultures sont maintenues en atmosphère microaérophile contrôlée (83% N₂, 4% H₂, 5% O₂ et 8% CO₂), dans une chambre de culture MACS-VA500 (Don Whitley, West Yorkshire, Angleterre). Les différentes souches d'*E. coli* utilisées au cours de ces travaux ont été cultivées en milieu Luria-Bertani (LB) (Gibco, USA) liquide ou solide à 37°C. Les souches bactériennes sont conservées à -80°C dans leur milieu de culture supplémenté avec 15% de glycérol. La lignée cellulaire épithéliale embryonnaire INT-407 a été obtenue auprès de l'“American Type Culture Collection”. Cette lignée cellulaire a été maintenue dans un milieu minimum essentiel (MEM- α), supplémenté avec 10% de sérum foetal bovin (Gibco, USA). Les cellules sont cultivées en atmosphère enrichie à 10% en CO₂, à 37°C.

2 Mesures de concentration d'ADN

La concentration d'ADN est mesurée par fluorimétrie grâce à l'incorporation spécifique et quantitatif du fluorochrome PicoGreen (Molecular Probes, Leiden, Pays-Bas). Une courbe étalon est faite à l'aide d'un standard d'ADN de concentration connue. Des dilutions en tampon TE du standard sont effectuées pour obtenir une gamme de concentration en ADN comprise entre 1 et 150 ng/ μ l. Soixante-quinze microlitres de chaque dilution sont distribués dans une plaque de 96 puits en plastique noir (Costar, Réf.

3694). Un puits supplémentaire contiendra 1 μ l de l'échantillon à tester, dilué dans 74 μ l de TE. La solution mère de PicoGreen est diluée 200 fois dans du TE en quantité suffisante pour pouvoir ajouter 75 μ l à chaque puits. Les solutions sont ensuite mélangées et lues par le fluorimètre (Victor3TM Multilabel Counter, Perkin Elmer, USA). La concentration de l'échantillon d'ADN est alors déterminée par comparaison de la valeur de fluorescence obtenue avec les valeurs de fluorescence de la courbe étalon.

3 Test d'invasion bactérienne.

La lignée INT-407 (cellules embryonnaires intestinales humaines) est mise en culture dans un milieu MEM α (Invitrogen, USA), 10% sérum albumine de bovin (FBS, Hyclone, USA), en plaque de 24 puits de 1 ml. Pour cela une préculture de cellules est incubée 24 h à 37°C en atmosphère contrôlée (10 % CO₂). Cette durée est suffisante à l'obtention d'un tapis confluent d'environ 5.10⁵ cellules/puits. Le milieu de culture est remplacé par 0.4 ml d'une suspension bactérienne (cultivée en milieu bi-phasique) en phase exponentielle de croissance diluée dans du milieu MEM α . Le rapport bactérie/cellule INT 407 utilisé varie de 10 à 90. Après 3 heures d'incubation à 37°C en atmosphère contrôlée, de la gentamycine est ajoutée au milieu d'incubation (concentration finale 100 mg/l), qui est remis à incuber pendant 1 h à 37°C. Cette étape permet de tuer les bactéries n'ayant pas colonisé les cellules épithéliales. Après ce traitement, les cellules INT-407 sont lavées à deux reprises avec une solution de Hanks (Gibco-BRL, USA) afin d'éliminer toute trace de gentamycine. Les cellules épithéliales sont ensuite décrochées du fond du puits par incubation 4 min en présence de 20 μ l de trypsine à 0.25% dans du MEM α . Un millilitre d'une dilution à 0,1% de Triton dans la solution de Hanks est ajouté à chaque puits. La plaque est ensuite incubée 20 min à température ambiante. Cette étape permet la lyse des cellules INT-407 et la libération des bactéries. La suspension bactérienne est récupérée, homogénéisée 2 min en Vortex, avant d'être étalée sur milieu MH. Après croissance, le pourcentage de bactéries ayant eu la capacité d'envahir les cellules INT-407 est déterminé. Il est à noter que le temps d'incubation de 3 heures ne permet pas aux bactéries de se multiplier. Ainsi, le pourcentage d'invasion est calculé par

le rapport du nombre de bactéries invasives divisé par le nombre de bactéries inoculées initialement.

4 Tests préliminaires à la fabrication des puces à ADN de type "shotgun"

4-1 Elimination de l'ADN chromosomique

Cette étape est effectuée dans le but d'éliminer l'ADN chromosomique d'*E. coli* des solutions d'ADN plasmidiques. La présence de cet ADN « parasite » aurait en effet pour conséquence d'augmenter potentiellement le bruit de fond de l'hybridation. Cette élimination est effectuée par une réaction utilisant une nucléase spécifique pour l'ADN chromosomique (Plasmid-Safe™ ATP-Dependent DNase, Epicentre, USA), dans les conditions décrites par le fournisseur.

4-2 Restriction/linéarisation du plasmide

Cette réaction est faite par utilisation de l'enzyme *EcoRI* qui coupe le plasmide en deux endroits de part et d'autre du site de clonage, libérant ainsi l'insert. Un tel essai nous a permis de déterminer si une linéarisation permettait une meilleure accessibilité de l'ADN lors des hybridations sur la puce.

4-3 Dénaturation de l'ADN par chauffage

Cette étape est réalisée en utilisant un appareil PCR. La solution d'ADN plasmidique est chauffée 5 min à 99°C puis refroidie sur glace avant d'être additionnée d'un volume équivalent de DMSO.

4-4 Préparation, marquage et hybridation des sondes

Les conditions de dépôt d'ADN sur la puce, le marquage de l'ADN des souches *C. jejuni* NCTC 11168, TGH 9011 et 81-176, ainsi que les conditions d'hybridation et de détection du signal d'hybridation sont décrits dans le sous-chapitre 6-2 (Marquage de l'ADN).

5 Construction de puces à ADN des souches *C.jejuni* 81-176 et TGH 9011

5-1 Extraction d'ADN chromosomique de *C. jejuni*

La méthode suivie est une modification de la méthode d'Ausubel et coll. (58). Cent microlitres d'une culture de 48 h de *C. jejuni* en milieu MH sont étalés sur milieu MH gélosé (10 boîtes), incubées à 37°C en condition micro-aérophile jusqu'à obtention d'un tapis confluent. Le tapis bactérien est récupéré par addition de 10 ml de milieu MH après remise en suspension à l'aide d'une pipette Pasteur à l'extrémité recourbée. Les bactéries sont ensuite transférées dans un volume de 10 ml de milieu MH puis centrifugées à 4000 rpm pendant 5 min. Le culot bactérien est repris dans 19 ml de TE pH 8 et additionné de 1 ml de SDS 10%. Cette suspension est ensuite transférée dans un tube Falcon (50 ml) et placée à 37°C pendant une heure après addition de 70 µl de Rnase A (10 mg/ml). Cent microlitres de protéinase K (20 mg/ml) sont ajoutés au mélange et placé à 50°C pendant 3 heures. 3,6 ml de NaCl 5 M sont ensuite ajoutés. Le tout est agité puis additionné de 3ml d'un mélange de 10% CTAB (cétyl tri-ammonium bromure) dans 0,7M NaCl, et à nouveau mélangé vigoureusement. La solution est ensuite placée à 65°C pendant 20 min. A cette étape la solution doit être fluide. Un volume équivalent d'une solution de phénol/chloroforme (1:1) est ajouté, le tout étant agité vigoureusement puis centrifugé 10 min à 7000 rpm. Le surnageant est récupéré et additionné d'un volume équivalent de chloroforme. Le tout est agité vigoureusement puis centrifugé 10 min à 7000 rpm. Après récupération du surnageant, celui-ci est additionné de 0,6 volume d'alcool isopropylique. L'ADN est récupéré à l'aide d'une pipette Pasteur scellée (non courbée) par agitation délicate. La pelote d'ADN récupéré au bout de la pipette Pasteur est lavée par un passage dans de l'éthanol à 70% puis séchée à l'air libre. L'ADN est repris dans 2 ml d'eau

distillée. De façon générale la quantité d'ADN récupérée est très importante (de l'ordre de 500 mg).

5-2 Nébulisation

Le nébulisateur utilisé est de la société IPI Medical Products Inc. (Cat. N° 4101, USA).

Au fond du nébulisateur on place la solution suivante :

- ADN 50 µg
- tampon TM 10 fois concentré 200 µl
- glycérol 50% qsp 2ml

(Tampon TM X10 : 5 ml 1M Tris-HCl, pH 8 ; 1,5ml 1M MgCl₂ et H₂O qsp 10 ml). La nébulisation est effectuée sur glace à 10 psi d'azote pendant 5 min, en vue de l'obtention de fragments d'ADN d'une taille moyenne de 1 kb. L'ensemble du nébulisateur est centrifugé rapidement à 1500 rpm pour collecter les projections de liquide engendrées lors de la nébulisation. Les 2 ml de la suspension de nébulisation sont repartis de façon équivalente dans 4 tubes de 2 ml. Un volume d'acétate de potassium 3 M correspondant à dix % du volume de la solution est ajouté, ainsi que 2,5 volumes d'éthanol pur. Le tout est mélangé au Vortex et placé à -80°C pendant 30 min, ou toute la nuit à -20°C. Le mélange est ensuite centrifugé à 13 000 rpm à 40°C pendant 15 min. Le surnageant est jeté et le culot d'ADN est lavé avec 500 µl d'éthanol à 70%. Après avoir jeté l'éthanol 70%, le culot est laissé séché puis suspendu dans 40 µl d'eau et la concentration d'ADN est estimée par la méthode décrite dans le sous-chapitre 2

5-3 Clonage dans un vecteur plasmidique

5-3-1 Obtention d'extrémités franches

Les fragments d'ADN issus de nébulisation sont traités pour obtenir des extrémités franches et pouvoir les introduire dans un vecteur de clonage de type pCR 4Blunt-TOPO (Invitrogen). Le protocole suivi est celui du kit "TOPO shotgun subcloning » (Invitrogen).

5-3-1-1 Principe

Le plasmide est fourni sous forme linéaire avec, attaché à son extrémité 3', une topoisomérase du virus *Vaccinia*. Ces topoisomérases ont la capacité de se fixer à des sites spécifiques de l'ADN et de cliver celui-ci au niveau des liaisons phosphodiester, après le motif 5'-CCCTT. L'énergie libérée lors de cette réaction permet la formation d'une liaison covalente entre l'extrémité 3' du brin coupé et le résidu tyrosyl de l'enzyme. Cette liaison est ensuite cassée par l'attaque de l'extrémité 5' formée lors du clivage de l'ADN par la topoisomérase, ce qui a pour résultat de libérer l'enzyme et de lier l'insert au vecteur.

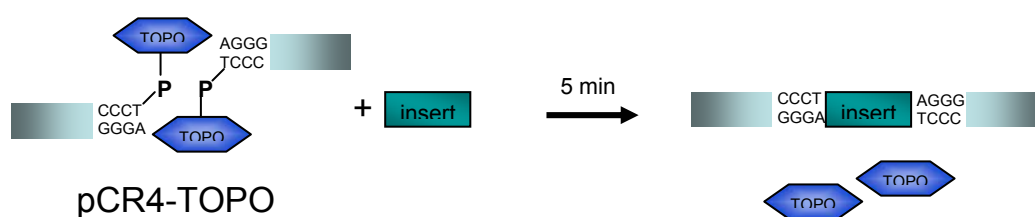


Figure 1 : Schéma de ligation de l'insert dans le vecteur pCR4-TOPO.

Pour le clonage, ce vecteur nécessite un insert à bouts francs dephosphorylés. Pour cela, un mélange des enzymes T4 DNA et Klenow polymérases permet de combler les extrémités simple-brin de l'ADN. De plus l'activité 3' 5' exonucléasique de l'enzyme T4

DNA polymérase dégrade les extrémités 3' simple brin. Cette étape est suivie d'une déphosphorylation réalisée par une phosphatase alcaline.

5-3-1-2 Réaction

Les conditions suivantes ont été utilisées pour la création d'inserts à bout francs dephosphorylés :

- ADN 1 à 3 µg
- Tampon X10 5 µl
- 250 mM dNTPs 5 µl
- Klenow DNA pol. (8U) 2 µl
- T4 DNA pol. (8U) 2 µl
- BSA (1 mg/ml) 1 µl
- H₂O qsp 50 µl

Cette réaction est réalisée à température ambiante pendant 30 min puis stoppée par chauffage 20 min à 75°C.

5-3-2 Déphosphorylation du vecteur

Les solutions suivantes sont ajoutées directement au produit de ligation :

- 40 µl ddH₂O
- 5 µl tampon de déphosphorylation
- 5 µl de Phosphatase intestinale de veau (correspondant à 5 unités)

La réaction est faite à 37°C durant une heure. L'ADN est ensuite purifié par le kit "mini-elute" (QIAGEN) selon les indications du fournisseur. L'ADN est repris dans un volume final de 9 µl de ddH₂O.

5-3-3 Ligation

Le vecteur est fourni avec le kit. La ligation est réalisée par la topoisomérase fixée aux extrémités du vecteur.

La réaction de ligation est faite dans les conditions suivantes :

- ADN à bouts francs (environ 25 ng/μl) 4 μl
- Sels (1,2 M NaCl, 0.06 M MgCl₂) 1 μl
- Vecteur (10 ng/μl) 1 μl

La réaction est faite 5 minutes à température ambiante et le tout est purifié par le kit mini-elute en vue d'une électroporation (l'élimination des sels est une condition cruciale pour l'électroporation). Après purification, l'ADN est repris dans 9 μl et concentré par évaporation à un volume de 3 μl.

5-4 Transformation par électroporation

Les 3 μl de la solution de ligation sont ajoutés doucement sans pipetage ni formation de bulles à 50 μl de cellules électrocompétentes d'*E. coli* TOP10™ (Invitrogen, USA) décongelées sur glace. Le mélange est transféré dans une cuve d'électroporation de 2 mm conservée sur glace. L'électroporation est réalisée avec les paramètres suivants : 129 ohms de résistance et 2,5 kv de tension de décharge (de l'ordre de 5 ms).

Les cellules compétentes sont ensuite rapidement mélangées dans 1 ml de milieu SOC (2% tryptone, 0,5% extrait de levure, 10 mM NaCl, 2,5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose), transférées dans un tube de 10 ml de capacité incubé à 37°C sous agitation (200 rpm) pendant 1 heure. Un volume de 10 à 15 μl de la suspension bactérienne est étalé sur boîte de Pétri contenant du milieu LB gélosé additionné de 50

µg/ml d'ampicilline. Quarante microlitres d'une solution à 40 mg/ml de X-GAL ont été préalablement étalés (l'addition d' IPTG n'est pas nécessaire).

5-5 Conservation des clones

Les clones positifs (blancs) sont repiqués à l'aide de cure-dents et transférés individuellement dans un puits d'une plaque à 96 puits contenant 150 µl de milieu LB supplémenté avec 100 µg/ml d'ampicilline. Les bactéries sont cultivées 16 h à 37°C sous agitation (200 rpm) avant d'être additionnées de 7% de DMSO puis stockées à -80°C.

5-6 Préparation d'ADN plasmidique pour la construction d'une banque génomique

Les souches d'*E.coli* sont inoculées dans 1,6 ml de milieu LB 2x concentré supplémenté avec de l'ampicilline (50 µg/ml) dans une plaque de culture (capacité 2 ml) de 96 puits. La plaque est couverte avec du papier d'aluminium stérile et mise à incuber à 37°C, 320 rpm pendant 20 h. La plaque est ensuite centrifugée 15 min à 1500 rpm. Après avoir éliminé le surnageant, le culot bactérien est suspendu par pipetage dans 70 µl d'une solution de GET (30 mM glucose, 15 mM Tris-HCl pH 8, 30 mM Na₂EDTA, 60 mg/L RNase A). Après homogénéisation, 70 µl d'une solution de lyse (0,2M NaOH, 1% SDS) sont ajoutés. Le tout est agité 1 min puis laissé à température ambiante pendant 2 min. Soixante dix microlitres d'une solution contenant 14% d'acide acétique glacial et 3,6 M d'acétate de potassium sont ajoutés à la suspension de lyse et mis à agiter 2 min.

Le lysat bactérien est récupéré délicatement par pipetage. Pour cela la pipette est descendue le long de la paroi du puits jusqu'à atteindre le fond. La totalité des lysats est prélevée et dispensée dans les puits correspondants d'une plaque de filtration "Costar PVDF lysate clearing plate" (0,2 µm, Costar, Ref. 3504). La plaque Millipore FB contenant 140 µl par puits d'une solution 6,1 M d'iodure de potassium est placée au fond d'un Manifold, puis la plaque de filtration PVDF est placée au-dessus (Figure 2). Un vide

de 8 Inches (soit environ 200 mm) de Hg est appliqué pendant 3 min. La suspension de lyse est filtrée par différence de pression à travers la plaque Costar PVDF et est ensuite récupérée dans la plaque Millipore FB. A ce stade, la plaque Costar PVDF est jetée.

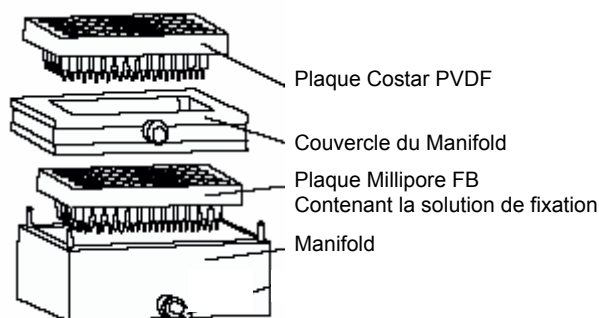


Figure 2 : Elimination des débris cellulaires. Les plasmides sont récupérés dans une plaque Millipore FB contenant une solution de fixation.

La suspension de lyse débarrassée des débris cellulaires est mélangée par pipetage avec la solution précédemment placée dans la plaque Millipore FB. Après homogénéisation, la plaque Millipore FB est placée au dessus du manifold et filtrée par application d'un vide d'environ 600 mm de Hg pendant 1 min. Les puits de la plaque sont lavés 2 fois avec 200 μ l d'éthanol 80% (Figure 3).

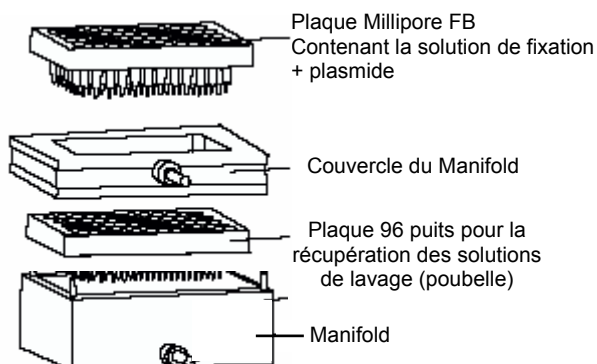


Figure 3 : Concentration et lavage des plasmides. Les plasmides sont fixés sur les filtres de la plaque Millipore FB.

L'éthanol résiduel est séché par centrifugation de la plaque pendant 10 min à 1000 rpm. Le matériel plasmidique est ensuite solubilisé par addition de 75 μ l d' H_2O dans chacun des 96 puits de la plaque et transféré par filtration dans une plaque de 96 puits stérile (Figure 4) par application d'un vide de 600 mm de Hg pendant 1min.

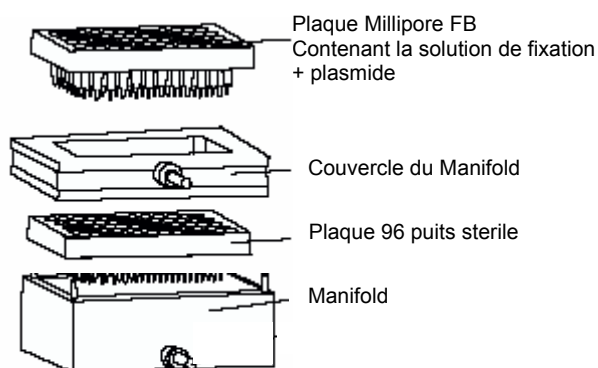


Figure 4 : Elution de plasmides.

Entre 55 et 60 μ l de solution plasmique de concentration 25 à 100 ng/ μ l sont récupérés. A la suite de chaque purification plasmidique, un échantillon de 3 μ l est déposé sur gel d'agarose 1% afin de s'assurer de la qualité du produit obtenu ainsi que pour l'estimation de sa concentration. Après purification, les plaques contenant les plasmides sont stockées à -80°C jusqu'à utilisation.

5-7 Dépôts des ADN plasmidiques

Après avoir vérifié que les plasmides pouvaient être déposés sur la puce sans traitement additionnel, la puce est réalisée par dépôt de l'ensemble de la collection de plasmides soit 9600 clones. Pour cela, les plasmides purifiés conservés en plaque de 96 sont transférés, de façon ordonné, en plaque de 384 puits (Genetix Catalog #X6004,USA) en quantités

suffisantes pour obtenir entre 500 et 1500 ng d'ADN par puits. Les plaques sont mises à évaporer en étuve sèche à 37°C. Après évaporation, 10 µl d'une solution SSC 3X (0,45 M NaCl, 45 mM citrate de sodium) sont ajoutés dans chaque puits, aboutissant ainsi à une concentration plasmidique comprise entre 50 et 150 ng/µl.

Vingt cinq plaques de 384 puits sont utilisées au total par souche bactérienne. Une fraction du contenu de chacun des puits est déposée sur une lame de microscope traitée par de l'aminosilane (CMT GAPS-II ; Corning inc.; Corning N.Y.). Le dépôt est effectué par un robot construit par la compagnie Molecular Dynamics en atmosphère contrôlée à 60% d'humidité. Le pattern de dépôt est constitué de 12 grilles comprenant 12 rangs de 62 colonnes et un rang de 56 colonnes, soit au total 9600 spots pour l'ensemble de la puce. Chaque spot a une taille moyenne de 250 µm. Chacun d'entre eux au sein d'une grille est séparé de son plus proche voisin par 300 µm (Figure 5). Après dépôt de l'ensemble de la librairie plasmidique, l'ADN est fixé sur le support en verre par chauffage en étuve sèche à 80°C pendant 4 heures. Les puces sont ensuite conservées en atmosphère sèche à température ambiante jusqu'à utilisation.

6 Hybridation des puces à ADN

6-1 Prétraitement des puces

Avant utilisation, les puces sont trempées dans une solution contenant SSC 5X, 25 % formamide, 0,1% SDS et 1% BSA, pendant 45 min à 42°C. Cette étape permet de dénaturer les brins d'ADN à la surface de la puce à ADN rendant ceux-ci accessibles à la sonde. La BSA a pour but de saturer la surface de la puce à ADN ne présentant pas d'ADN et limiter les fixations aspécifiques des sondes. Les puces sont ensuite lavées avec de l'eau distillée et séchées par centrifugation.

6-2 Marquage de l'ADN

6-2-1 Principe

Pour chaque hybridation, 2 μg d'ADN de la souche utilisée pour la construction de la puce à ADN et 2 μg d'ADN de la souche *C. jejuni* NCTC 11168, sont marqués de façon indépendante avec un fluorochrome. Le marquage est réalisé par incorporation d'un nucléotide de type aminoallyl-dUTP lors d'une synthèse aléatoire par l'enzyme de Klenow. Ce nucléotide sert d'ancrage à un fluorochrome de type succinimidyl ester Cy3 ou Cy5 (Amersham, USA).

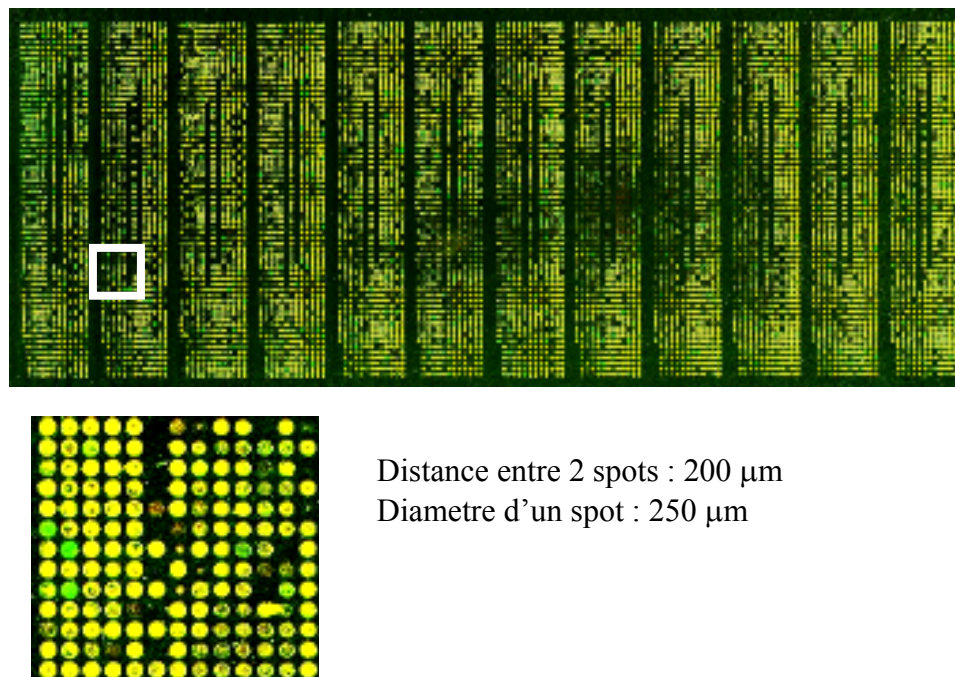


Figure 5 : puce à ADN après hybridation.

6-2-2 Amorçage aléatoire ou « Random priming »

Les ADN sont traités par "random priming" par utilisation d'un kit (Bioprime DNA labelling system, Invitrogen) dans les conditions suivantes :

- ADN (2 µg) 21 µL
- Amorces aléatoires (2,5 mM) 20 µL

Cette solution est portée à 99°C pendant 5 minutes pour dénaturation, puis gardée sur glace pendant 5 minutes. Cinq microlitres d'un mélange de dNTP contenant de l'aminallyl dUTP (1,2mM dGTP, dATP, dCTP ; 0,6mM aminallyl dUTP et 0,6 mM dTTP) sont ajoutés ainsi qu'un microlitre d'une enzyme de Klenow (40 U/µl). Le tout est mis à incuber à 37°C pendant 2 heures. Les amines non incorporées sont éliminées par utilisation d'une colonne de type Microcon 30 (Millipore). Pour cela, la solution de marquage est additionnée de 450 µl de H₂O bi-distillée et mise à centrifuger pendant 8 min à 10.000 g. La solution ayant passé à travers le microcon est jetée, le microcon est remplacé dans le tube et, additionné de 450 µl H₂O, est mis à centrifuger 10 min à 10.000 g. Le microcon est ensuite retourné dans un nouveau tube eppendorf et centrifugé 15 secondes afin de récupérer la solution d'ADN retenu par le filtre.

6-2-3 Marquage fluorescent

L'ADN est ensuite mis à évaporer à sec et repris dans 10 µl d'une solution de bicarbonate de sodium 0,1 M et 10 µl de NHS-ester Cy3 ou Cy5 (Amersham). Ce mélange est incubé 1 h à l'obscurité à température ambiante. Puis le marquage de l'ADN est arrêté par addition de 4,5 µl d'hydroxylamine 4 M et remis à incuber 15 min à l'obscurité à température ambiante. Les fluorochromes non incorporés sont éliminés par utilisation du kit Mini-elute (QIAGEN). L'ADN est élué dans 10 µl H₂O. La solution d'ADN, à ce stade, présente une coloration bleu ou rose due respectivement au marquage par le fluorochrome Cy3 ou Cy5, et reflétant ainsi la bonne efficacité de l'opération. Les

ADN marqués par les fluorochromes Cy3 ou Cy5 sont mélangés et additionnés de 2,5 µl d'ADN de sperme de saumon (10 mg/ml, Invitrogen), 9 µl de formamide, 9 µl de SSC 20X et 0,36 µl de SDS 10%. Cette solution d'hybridation est soigneusement mélangée par pipetage avant d'être dénaturée 2 min à 99°C puis maintenue à 42°C jusqu'au dépôt sur le support de puces. L'hybridation est faite entre lame et lamelle (Grace Bio-labs). La puce est ensuite placée dans une chambre d'hybridation (ArrayIt, USA) pour une incubation à 42°C pendant 16 h. La puce est ensuite désassemblée puis lavée. Les lavages consistent en une incubation 5 min à 42°C dans une solution SSC 2X, 0,1% SDS puis 10 min dans une solution SSC 0,1X, 0,1% SDS suivi de 4 lavages successifs de 1 min dans une solution SSC 0,1X. La puce est ensuite rincée avec un léger jet d'eau puis séchée par centrifugation.

7 Révélation de l'hybridation et sélection des spots uniques

7-1 Révélation de l'hybridation

La révélation du signal d'hybridation est effectuée par un scanner confocal de type ScanArray3000 (Perkin Elmer). Les puces sont scannées à une résolution de 10 µm et à une longueur d'onde de 532 nm et 635 nm, ce qui permet de détecter respectivement l'émission de fluorescence des molécules Cy3 et Cy5. La quantification du signal d'hybridation de chacun des deux ADNs est effectuée par le programme GenePix Pro 4 (Axon Instrument).

7-2 Sélection des spots spécifiques à la souche testée.

Les opérations sont effectuées à l'aide du programme Excel (Microsoft) et le programme OriginPro7 (OriginLab Corporation). La sélection des spots uniques est effectuée grâce à 3 expériences d'hybridation indépendantes. Tout d'abord, sont éliminés les spots correspondant à l'ADN utilisé pour la construction de la puce et présentant un signal d'hybridation inférieur à 2 fois la déviation standard du bruit de fond. Ensuite, les

valeurs du signal d'hybridation aux 2 longueurs d'onde sont normalisées par une analyse de régression linéaire. Cette méthode est basée sur 3 étapes itératives permettant l'identification des spots ayant une séquence commune chez les 2 souches étudiées.

- La première étape consiste en l'application d'une régression linéaire au graphe correspondant à la moyenne d'intensité géométrique Cy3 contre la moyenne d'intensité géométrique Cy5. Cette régression est utilisée pour la transformation des données brutes des signaux d'hybridation. Pour chaque spot, le rapport transformé des intensités d'hybridation Cy3 et Cy5 a été calculé de la façon suivante : $m = \text{Cy3/Cy5}$.
- La 2^{ème} étape correspond au classement décroissant des spots selon leur valeur m et élimination des 20 spots présentant la valeur m la plus basse (spots uniques à la souche ayant servi à la construction de la puce).
- La troisième étape correspond à la re-transformation des valeurs d'intensité des signaux d'hybridation bruts (moins les 20 spots éliminés), et de la re-calculation d'une régression linéaire similaire à l'étape 1.

Ces 3 étapes sont répétées de façon itérative jusqu'à obtention d'un coefficient de régression linéaire (r) proche de 1. La dernière régression obtenue est utilisée pour la normalisation des intensités d'hybridation. Nous avons choisi de façon arbitraire de séquencer comme spot unique les spots présentant une valeur $m < 0.5$ et $0.5 < m < 0.6$.

Seuls ont été retenus les clones remplissant par deux fois au minimum sur les trois essais effectués, les conditions citées précédemment.

8 Sélection des ORFs de la souche NCTC 11168 absents de TGH 9011

8-1 Sélection

La sélection des clones uniques à la souche NCTC 11168 révélés par utilisation de la puce à ADN de cette même souche, est effectuée par le programme GACK (Kim et al (<http://falkow.stanford.edu>)). Ce programme utilise un algorithme basé sur la distribution du ratio des signaux d'intensité pour l'estimation de la probabilité de présence ou d'absence du clone correspondant dans la souche testée. La détermination des ORFs de la

souche NCTC 11168 absents de TGH 9011, a été basée sur la moyenne de 3 hybridations.

8-2 Vérifications

Pour la vérification de l'absence ou présence des ORFs de NCTC 11168 chez TGH 9011, des PCR correspondants à une sélection d'ORFs, les uns déterminés absents et les autres présents, ont été faites sur l'ADN de TGH 9011 par utilisation des mêmes amorces PCR utilisés pour la construction de la puce à ADN de NCTC 11168. Au total, 416 PCR ont été effectuées, 88 sur des ORFs déterminées comme étant absentes de TGH9011, et 328 sur des ORFs déterminées comme étant présentes chez les 2 souches.

9 Détermination des clones correspondant au plasmide pTet de *C. jejuni* 81-176

Les plasmides pVir et pTet de la souche *C. jejuni* 81-176 cultivée sur milieu MH gélosé et supplémenté par 100 µg/ml de tétracycline (sélection du plasmide pTet conférant la résistance à cet antibiotique) ont été extraits par l'utilisation du kit MidiPrep (Qiagen). Ces plasmides sont ensuite marqués par le fluorochrome Cy3 via la technique décrite pour le marquage indirect de l'ADN. Les plasmides marqués sont utilisés pour une hybridation compétitive sur la puce à ADN de 81-176, avec l'ADN de cette souche marqué par le fluorochrome Cy5. Les spots identifiés via le programme GACK comme étant plasmidiques sont séquencés. Les séquences sont ensuite éliminées des parties correspondantes au plasmide ainsi que les parties de mauvaise qualité. Celles-ci sont ensuite soumises individuellement à comparaison via BLASTN à une banque génomique totale non redondante (NCBI). Les séquences alignant sur toute leur longueur avec la séquence de pVir et présentant un pourcentage d'identité égal à 100% sont considérées comme correspondantes à ce plasmide. Les autres séquences sont considérées comme appartenant au plasmide pTet. Les séquences correspondantes à chacun de ces 2 plasmides sont ensuite assemblées séparément en contigs via utilisation du programme Lasergen (DNASTar).

10 Séquençage des clones et annotation.

Les clones caractérisés comme étant uniques à la souche testée ont été séquencés à partir du stock de plasmides purifiés ayant servi à la création de la puce. Le séquençage a été effectué au « DNA Sequencing Facility of the Oklahoma Health Science Center, Oklahoma City), sur un appareil de séquençage de type ABI prism 3700 capillary sequencer (Applied Biosystems, USA).

La première étape consiste à éliminer et à reséquencer éventuellement les mauvaises séquences. Les séquences sélectionnées sont ensuite éliminées des parties correspondantes au plasmide ainsi que les parties de mauvaise qualité par utilisation du programme Lasergen (DNASTar). Ces séquences sont comparées à une banque de donnée génomique par BLASTN à l'adresse suivante : <http://www.ncbi.nlm.nih.gov/BLAST/>. Les séquences présentant un alignement sur l'intégralité de leur séquence avec le génome de *C. jejuni* NCTC 11168 sont éliminées. Les séquences ayant passé ces différents filtres sont assemblées en contigs par l'utilisation du programme Lasergen. Les contigs formés sont soumis individuellement au programme Artemis (<http://www.sanger.ac.uk/Software/Artemis/v5/>), à la recherche de phases ouvertes de lecture (ORFs) supérieures à 100 paires de bases. Les ORFs identifiées sont ensuite examinées manuellement, à la recherche en 5' du premier "codon d'initiation potentiel AUG, UUG, CUG (si la séquence en 5' présente un codon-stop en amont de l'ORF trouvé par le programme). Les protéines potentiellement codées par ces ORFs sont soumises à une analyse comparative avec une banque de données protéiques via le software BLASTP. Le résultat de BLASTP permet d'attribuer une fonction potentielle aux ORFs identifiées.

11 Etudes des gènes régulés par la protéine Fur, par utilisation de la puce à ADN de TGH 9011

11-1 Clonage du gène *fur* dans un vecteur d'expression.

Nous avons choisi le vecteur pASK-IBA7 (IBA, Allemagne, Figure 6). Ce vecteur permet l'expression de protéines cytoplasmiques chimériques présentant une séquence marqueur (strep-Tag II, WSHPQFEK) à leurs extrémités NH₂ terminales.

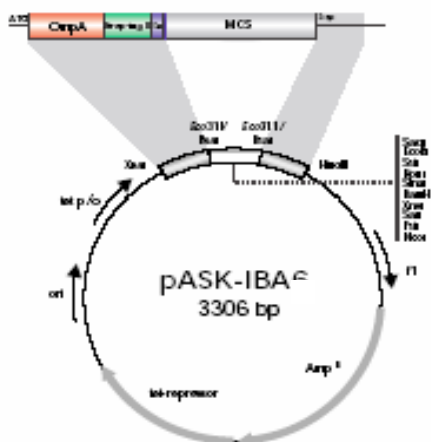


Figure 6 : Vecteur pASK-IBA

Cette séquence est utilisée pour la purification de la protéine sur colonne d'affinité. Strep-Tag II est un marqueur présentant une grande affinité envers la molécule Strep-Tactin, un dérivé de streptavidin (Figure 7). L'affinité de ce marqueur sur la molécule de Strep-Tactin est d'environ 100x supérieure à son affinité envers la streptavidin.

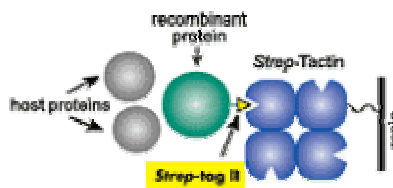


Figure 7 : Représentation schématique de la liaison du marqueur Strep-Tag II sur la molécule de streptactin (phase solide)

Cette propriété permet, en une seule étape, la purification d'une grande quantité de protéine recombinante. Le marqueur Strep-Tag est donné par le constructeur comme interférant peu avec les propriétés de la protéine recombinante. L'excision du marqueur reste toutefois possible.

11-1-1 Conception d'amorces

Le gène *fur* de TGH 9011 est cloné dans le vecteur d'expression pASK-IBA7. Ce plasmide permet l'expression cytoplasmique de façon inductible de la protéine recombinante via le promoteur *tet*. Il présente également une zone multiple de clonage en aval du promoteur *tet* et de la séquence codante pour le marqueur Strep-Tag II. Le clonage en phase du gène *fur* est effectué par insertion au niveau du site de restriction *BsaI*. En vue de ce clonage, des amorces PCR du gène *fur* incluant un site de restriction *BsaI* (formule) à leurs extrémités sont créées via le programme Primer D'signer 1.1 (programme disponiblesur www.iba-go/iba_us/download.html). Les amorces suivantes ont été commandées (Sigma-Genosys):

Amorce directe : ATGGAGGTCTCAGCGATGCTGATAGAAAATGGAATATGA

Amorce indirecte: ATGGTAGGTCTCATATCATATTTTTTACCTTTGCTTTTTTGATTATTAC

11-1-2 PCR

Le gène *fur* est amplifié par PCR par l'enzyme *PfX* (Invitrogen, USA). Cette enzyme est préférée à la Taq polymérase car il présente une activité de relecture limitant le nombre

d'erreurs introduites au cours de la réaction. La réaction est faite dans les conditions suivantes:

- Amorce directe (1 nM): 1 μ l
- Amorce indirecte (1 nM) : 1 μ l
- dNTP (10 mM): 2 μ l
- MgCl₂ (50 mM) : 1,75 μ l
- Tampon 10X : 2,5 μ l
- ADN TGH 9011 (0,1 mM) 1 μ l
- *PfX* 0,5 μ l
- H₂O qsp 25 μ l

30 Cycles: 94°C 1'
 58°C 45''
 68°C 45''

Le produit PCR est déposé sur gel d'agarose 1% pour vérification. Une bande de 505 pb, correspondant à la taille du gène *fur* (471 pb) et des queues insérées sur les amorces (17 pb pour chaque amorce) est attendue. Le produit PCR est ensuite purifié par le kit QIAquick PCR purification kit (QUIAGEN, Cat No. 28106, Etats Unis) selon les conditions décrites par le fournisseur.

11-1-3 Clonage.

Tout d'abord le vecteur et le produit PCR *fur* sont digérés séparément par l'enzyme *BsaI* dans les conditions préconisées par le fournisseur (Biolabs, Etats Unis). Les 2 produits de digestion sont ensuite déposés séparément sur gel d'agarose 0,8%. Après migration, les bandes correspondantes aux produits de digestion de ces produits (3236 pb pour le vecteur et 489 pb pour l'insert) sont extraits du gel et purifiées individuellement par le kit Prep-A-Gene DNA purification system (BIO-RAD, Etats Unis).

Le vecteur et l'insert *fur* sont ensuite liés par l'enzyme T4 ligase (Invitrogen) en utilisant un ratio vecteur/insert de 1/3 et en suivant les conditions décrites par le fournisseur. La ligation est menée à 16 °C pendant 18 h. Le produit de ligation est ensuite utilisé pour la transformation de cellules *E. coli* DH5 α (TOP10, Invitrogen, USA) par électroporation dans les conditions décrites par le fournisseur. Après une phase de récupération des cellules 1 h à 37 °C en milieu LB, les cellules sont déposées sur boîte de Pétri LB + Amp 100 μ g/ml, et incubées toute une nuit à 37 °C.

11-1-4 Vérifications des clones

Certaines vérifications sont faites avant purification de la protéine, afin de s'assurer de sa qualité. Deux clones sont sélectionnés pour vérification. Le plasmide est purifié par lyse alcaline suivi d'une digestion par l'enzyme *Sau3AI* et d'un dépôt sur gel d'agarose. Tout d'abord la présence d'un insert de la taille escompté de 505pb (471 pour *fur* et 34 pour les 2 queues) est vérifiée par PCR par utilisation des amorces ayant servi à l'amplification du gène. Ensuite une confirmation de cette amplification est faite par digestion par les enzymes de restriction *BclII* et *Sau3AI*. Ces 2 enzymes coupent l'insert 1 seule fois libérant 2 bandes de 317 et 188 pb. L'enzyme *BclII* est sensible à la méthylation et n'est pas en mesure de couper le plasmide.

11-2 Obtention de la protéine Fur

Après avoir isolé un clone présentant une séquence nucléotidique sans erreur, celui-ci est mis en culture dans le but d'obtenir la protéine purifiée.

11-2-1 Culture

Cent millilitres de milieu LB supplémenté à 100 mg/l d'ampicilline, sont inoculés avec 2 ml d'une pré-culture bactérienne de 14 heures. Cette culture est incubée à 37 °C sous agitation jusqu'à obtention d'une D.O. à 550 nm égale à 0,5. A ce stade 10 μ l d'anhydrotetracycline (AHT) à 2 mg/ml dilué dans du diméthylformamide (DMF) sont

ajoutés. La suspension est ensuite remise à incuber 3 h supplémentaires dans les mêmes conditions. L'anhydrotétracycline est un inducteur du promoteur *tet* situé en amont du gène cloné sur le vecteur pASK-IBA7.

11-2-2 Lyse

Les bactéries sont récupérées par centrifugation 12 min à 4500 x g à 4 °C et remises en suspension dans du tampon W (100 mM Tris pH 8; 150 mM NaCl) conservé à 4 °C, pour être lysées par l'utilisation d'une presse de French. Pour cela, le culot bactérien correspondant à 100 ml de culture est repris dans 20 ml de tampon W froid puis soumis à 3 étapes de lyse à 700 psi. La suspension de lyse est conservée sur glace tout au long du processus de lyse. Les composés insolubles sont écartés par centrifugation à 14.000 rpm pendant 15 min à 4 °C. A ce stade le surnageant contenant la protéine Fur est prêt à être traité sur colonne d'affinité.

11-2-3 Purification

L'étape de purification est faite par chromatographie d'affinité contenant une résine Strep-Tactin. Cette résine présente une grande affinité pour la séquence marqueur fusionnée à la protéine. La chromatographie d'affinité est faite à 4°C tout au long du processus de purification. Succinctement la solution de lyse est passée sur une colonne contenant 3 ml de résine strep-Tactin. Après passage complet de cette solution à travers la colonne par gravité, la résine est lavée 5 fois avec 3 ml de tampon W (0.1 M Tris-HCl, pH 8 ; 0,15 M NaCl). La protéine recombinante est ensuite éluée par addition de 1,5 ml d'une solution de desthiobiotin 2.5 mM (Sigma, USA). Cette étape d'élution est répétée 4 fois. Les 3 derniers passages sont regroupés et conservés à -20°C jusqu'à utilisation. La colonne est ensuite régénérée par addition de 3 fois 10 ml de tampon R (tampon W : 0.1 M Tris-HCl, pH 8 ; 0,15 M NaCl contenant 1 mM de HABA (2-[4'-hydroxybenzeneazo]benzoic acid)). Ce tampon contient un compétiteur de couleur rouge se fixant sur la résine permettant la régénération de la colonne d'affinité. Avant usage le compétiteur est élué par lavage avec 2 fois 7 ml de tampon W.

11-2-4 Vérification

Après purification un échantillon de l'éluat contenant la protéine Fur recombinante est prélevé puis analysé sur un gel d'acrylamide 15% (pour 21 ml: 7.88 ml de polyacrylamide (29:1) 40%; 5.25 ml de tampon L (1.5M Tris pH 8.8: 0.4% SDS), 7.72 ml d'eau; 0.125 ml d'une solution à 10% d'ammonium persulfate; 0.025 ml de TEMED). Le gel est coulé dans l'appareil d'électrophorèse. Une fois polymérisé, un gel de concentration (pour 5 ml: 0.56 ml de polyacrylamide (29:1) 40%; 1.25 ml de Tampon U (0.5 M Tris pH 6.8; 0.4% SDS); 3.14 ml d'eau; 0.02 ml d'une solution à 10% d'ammonium persulfate; 0.03 ml de TEMED) est coulé au-dessus du gel de séparation. Après polymérisation les échantillons sont déposés sur gel après addition de 50% de tampon de dépôt (10% glycérol, 2,5% SDS, 0,1% de bleu de bromophénol, 5% β -mercaptoéthanol) et chauffage dans un bain-marie pendant 10 min. Après migration à 50 V toute la nuit, les protéines sont révélées au bleu de Coomassie. Pour cela, le gel est retiré de la plaque de verre puis coupé dans un des coins en biseau pour pouvoir identifier les pistes. Le gel est immergé dans un bain contenant le colorant au bleu de Coomassie (0,5 g/l. dans 250 ml d'isopropanol, 100 ml d'acide acétique et 650 ml d'eau distillée) au moins 5 h à température ambiante sous agitation. Le gel est ensuite transféré dans des bains successifs d'une solution de décoloration contenant 40% méthanol et 10% acide acétique. La révélation d'une bande de 18,1 kb par le bleu de Coomassie reflète la présence de la protéine Fur dans l'éluat.

11-3 Synthèse d'anticorps anti-Fur

En vue de la détection de la protéine Fur sur notre puce à ADN, nous avons synthétisé des anticorps polyclonaux anti-Fur. Pour cette synthèse nous avons fait appel aux services Evoquest de la compagnie Invitrogen. La génération des anticorps est faite à partir de deux lapins. La production d'anticorps est suivie d'une purification sur colonne d'affinité créée à partir de la protéine Fur. La production et la construction de la colonne

d'affinité nécessitent 7 mg de protéine Fur pure, reprise dans un tampon sans trace d'autre composé organique. Pour cela une étape de dialyse a été nécessaire pour l'élimination de la molécule de desthioibiotin présente dans le tampon d'élution. Nous avons groupé les différentes fractions d'élution de la protéine Fur, réduit le volume par évaporation au SpeedVac à température ambiante, et dialysé cette solution sur les chambres Slide-A-Lyzer (PIERCE, USA), en utilisant le tampon W comme tampon d'échange et en suivant les indications fournies par la compagnie. La concentration de la protéine Fur est estimée par utilisation du kit BCA Protéine assay reagent (PIERCE, Cat No. 23225), selon les indications fournies par la compagnie.

11-4 Immunoprécipitation de la protéine Fur liée à l'ADN chromosomique

11-4-1 Préparation des cellules, fixation, lyse bactérienne et ultrasonation

Une culture bactérienne de *C. jejuni* est faite en milieu microaérophile en milieu MH biphasique de 10 ml jusqu'à obtention d'une DO₆₀₀ de 0,6 à 0,8. Un millilitre d'une solution de monophosphate de sodium 1 M, pH 6, ainsi que 11 ml de formaldéhyde 10% sont ajoutés à la culture bactérienne séparément. La suspension est incubée 20 min à température ambiante avec agitation, suivie de 30 min sur glace. Les bactéries sont collectées par centrifugation et lavées 2 fois avec une solution de PBS (0,15 M NaCl, 2,71 mM Na₂HPO₄ · 7 H₂O, 1,54 mM KH₂PO₄ [pH 7,4]). Le culot bactérien est remis en suspension dans 10ml de tampon de lyse (Tris 10 mM pH 8 ; 20% saccharose ; NaCl 50 mM ; EDTA 10 mM et 20 mg/ml de lysozyme), et incubé 30 min à 37°C. Cinq millilitres de tampon d'immunoprécipitation (HEPES-KOH 50 mM, pH 7,5 ; NaCl 150 mM ; EDTA 1 mM ; Triton X-100, 1% ; déoxycholate de sodium 0,1% ; SDS 0,1% ; PMSF 1 mM fraîchement préparé) sont ajoutés et le tout est remis à incuber 10 min supplémentaires à 37°C. La suspension est ensuite ultrasonée 5 fois 10 secondes sur de la glace. Cette étape permet de casser l'ADN en fragments d'une taille d'environ 700 pb.

11-4-3 Immunoprécipitation

Vingt deux microlitres d'une solution d'anticorps anti-Fur (1,9 mg/ml) sont ajoutés à la suspension bactérienne ultrasonée et mis à incuber à température ambiante 90 min avec une légère agitation. Trente microlitres de billes magnétiques présentant à leur surface de la protéine-A (MagnaBind Protein A Beads, PIERCE, USA) sont ajoutés à la suspension et remis à incuber dans les mêmes conditions (60 min supplémentaires). Les billes sont collectées à l'aide d'un support magnétique (MagnaBind Magnet, PIERCE, USA) et lavées 5 fois avec le tampon d'immunoprécipitation et 2 fois avec de l'eau distillée. Le complexe Anticorp-ADN est détaché des billes magnétiques par utilisation d'un tampon d'élution (Immunopurer IgG Elution Buffer, Pierce, USA). Les anticorps sont détachés de l'ADN par incubation de ce complexe toute une nuit à 65°C, après addition de 10% en volume de SDS 10%. L'ADN est ensuite purifié sur colonne QIAGEN selon les conditions dictées par le fabricant.

Chapitre III

Identification des gènes spécifiques à la souche
Campylobacter jejuni TGH 9011

1 Introduction

Depuis la publication en février 2000 de la séquence génomique de *C. jejuni* NCTC 11168 (123), de nombreuses évidences suggérant une grande variabilité génotypique au sein de cette espèce ont été reconnues (123). Cette variabilité pourrait en partie expliquer l'hétérogénéité des symptômes liés à l'infection par *C. jejuni*, allant d'une colonisation asymptomatique à une diarrhée inflammatoire, ou encore au développement du syndrome de Guillain-Barré (130).

Dans le but d'identifier les séquences uniques d'une souche bactérienne de *C. jejuni* en s'affranchissant d'un séquençage systématique de son génome, nous avons mis au point un nouveau type de puce à ADN. Ce type de puce est généré par dépôt des plasmides constituant une banque génomique totale d'une souche de *C. jejuni* non séquencée. Cette approche a été conduite sur une souche de *C. jejuni* TGH 9011 également référencée sous ATCC 43431). L'hybridation compétitive ADN/ADN entre la souche expérimentale utilisée pour la construction de la puce et la souche séquencée NCTC 11168 nous a permis d'isoler et de caractériser les séquences spécifiques à la souche testée. Contrairement à *C. jejuni* NCTC 11168, cette souche est capable d'envahir des cellules épithéliales humaines *in vitro* (80,130). Cette différence phénotypique, par rapport à la souche type séquencée, suggère la présence chez cette souche de gènes spécifiques codant pour des protéines impliquées dans le mécanisme d'invasion cellulaire. L'identification de ces gènes est essentielle à l'étude du mécanisme invasif et devrait permettre d'approfondir nos connaissances sur la pathogenèse des *Campylobacter*.

L'identification des séquences uniques à *C. jejuni* TGH 9011 a nécessité quatre étapes successives : (i) création d'une banque génomique représentative de chacune des 2 souches bactériennes, (ii) fabrication des puces à ADN, (iii) hybridation compétitive sur les puces à ADN génomique provenant de la souche ayant servi à la construction de celle-ci et de l'ADN génomique de la souche séquencée NCTC 11168, (iv) analyse des résultats et séquençage des clones identifiés comme étant absents chez NCTC 11168, puis identification des ORFs de ces clones.

Avant la construction des puces nous avons testé différentes conditions de traitement de l'ADN plasmidique avant dépôt, dans le but d'optimiser son accessibilité. Les résultats de

ces tests sont présentés en résultats additionnels à la première publication. A la suite de ces tests préliminaires, nous avons construit une banque génomique représentative de 9600 clones pour chacune des souches étudiées. Après purification des plasmides, chacun d'eux a été déposé sur une lame de verre revêtue d'aminosilane. Les puces à ADN résultantes ont été utilisées pour l'hybridation de l'ADN de la souche utilisée pour la construction de la puce et de l'ADN de NCTC 11168 marqués avec des fluorochromes distincts. Après analyse de l'hybridation, les inserts des clones sélectionnés ont été séquencés puis assemblés en contigs. Les phases ouvertes de lecture ont été ensuite identifiées. Les différents produits protéiques codés par ces ORFs ont été comparés à une banque protéique total par BLASTP pour identification fonctionnelle.

L'étude de la souche TGH 9011 (également référencée sous ATCC 43431) est présentée dans la publication « Identification of *Campylobacter jejuni* ATCC 43431 specific genes by whole microbial genome comparison ». Cette étude a permis l'identification de 130 ORFs spécifiques à la souche, représentant 84 kb d'ADN supplémentaire par rapport à la souche-type. Une fonction potentielle a été attribuée à environ 65% de ces ORFs. Les protéines codées par ces ORFs font principalement partie de mécanismes démontrés comme étant impliqués dans la pathogénèse de *C. jejuni*. La majorité des gènes impliqués dans la glycosylation des LOS, de nombreux gènes impliqués dans la formation de la capsule, des systèmes de restriction/modification ainsi que des protéines impliquées dans le transport membranaire ont été identifiés. *C. jejuni* TGH 9011 présente également 28 ORFs, majoritairement de fonctions inconnues, ayant des homologies avec des protéines codées par un îlot de pathogénicité d'*Helicobacter hepaticus*.

Cette étude utilise un type de puces à ADN correspondant à des clones plasmidiques. L'étude des souches *C. jejuni* TGH 9011 démontre l'efficacité de l'utilisation de ce type de puces pour l'identification rapide de la diversité génomique bactérienne par comparaison avec une souche séquencée.

2 Publication

[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

Identification of *Campylobacter jejuni* ATCC 43431-Specific Genes by Whole Microbial Genome Comparisons

Frédéric Poly, Deborah Threadgill, and Alain Stintzi

Journal of Bacteriology, 2004, Vol. 186, No. 14, Pages 4781-4795

Pages 4781-4795 :

- La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.
- Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur :

http://jb.asm.org/cgi/content/full/186/14/4781?maxtoshow=&HITS=10&hits=10&RESULTFORMAT=&fulltext=identification+of&searchid=1115219929219_2758&stored_search=&FIRSTINDEX=0&volume=186&issue=14&journalcode=jb

- Il est également possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP: peb.sciences@scd-ulp.u-strasbg.fr

3 Résultats additionnels

Introduction

La création d'une puce à ADN de type « shotgun » par dépôt direct des produits plasmidiques sans étape d'amplification PCR, n'avait jamais été décrite à l'époque où nous avons entrepris la construction de la puce à ADN de la souche TGH9011. Nous avons donc testé différentes conditions de traitement de l'ADN plasmidique avant de construire la puce, dans le but de déterminer les conditions optimales d'accessibilité de la sonde. Pour cela, nous avons fabriqué une première puce à ADN expérimentale, composée de spots correspondant à de l'ADN unique à la souche NCTC 11168, et de l'ADN commun aux deux souches TGH 9011 et NCTC 11168. Cette puce à ADN a été utilisée pour une hybridation compétitive entre l'ADN des souches NCTC 11168 et TGH 9011 marquées avec des fluorophores distincts. La qualité d'hybridation de ces différents traitements a été comparée avec les résultats obtenus à partir des produits PCR des inserts correspondants. Les différents traitements de l'ADN avant dépôt ont été ::

- Élimination de l'ADN chromosomique contaminant

Cette étape est réalisée dans le but d'éliminer l'ADN chromosomique d'*E. coli* potentiellement présent dans la solution plasmidique. Ceci est réalisé par une réaction utilisant une enzyme plasmid-safe (Epicentre). Cette enzyme a une activité endonucléasique importante pour l'ADN linéaire double brin mais faible pour l'ADN circulaire double brin de petite taille.

La réaction a été réalisée dans les conditions décrites par le fournisseur.

- Restriction du plasmide

Cette réaction est faite par l'enzyme *EcoRI* (Invitrogen) coupant le plasmide à deux endroits à proximité du site de clonage. Cette restriction est faite dans le but de linéariser le plasmide, augmentant ainsi l'accessibilité de la sonde vis à vis de celui-ci. La réaction a été réalisée dans les conditions décrites par le fournisseur.

- Dénaturation de l'ADN par chauffage

La dénaturation du plasmide est uniquement faite sur les plasmides circulaires non linéarisés. Elle a pour but de dénaturer le plasmide qui devrait être ainsi plus accessible à la sonde. Cette étape est réalisée en appareil PCR. L'ADN plasmidique est chauffé 5 min à 99 °C puis refroidi sur glace avant d'être déposé sur le support en verre de la puce.

Tests	Résultats des traitements étudiés				Hybridation
	Enzyme Plasmid-safe	EcoRI	purification *	99°C	
A	+	+	+	-	+++
B	+	+	-	-	0
C	+	-	+	+	+++
D	+	-	+	-	+++
E	+	-	-	+	0
F	+	-	-	-	0
G	-	+	+	-	++
H	-	+	-	-	0
I	-	-	-	+	++
J	-	-	-	-	++
PCR	-	-	-	-	+++

Tableau 2 : tests préliminaires de traitement des clones avant construction de la puce à ADN. +/- test effectué/non effectué. Colonne résultats : +++ bonne hybridation, ++ hybridation satisfaisante, 0 hybridation nulle. PCR : échantillon PCR purifié.

*purification faite après traitement enzymatique Plasmid-safe et/ou *EcoRI*.

Observations expérimentales (Tableau 2)

- les spots correspondants à l'ADN non purifié après traitement par les enzymes *EcoRI* ne présentent aucune hybridation. La purification après un traitement enzymatique est donc nécessaire.
- Le chauffage des échantillons juste avant dépôt sur la puce ne permet pas l'obtention d'une meilleure hybridation.

- La linéarisation du plasmide par l'enzyme *EcoRI* n'améliore pas de façon significative l'hybridation des sondes. Ce traitement n'est donc pas nécessaire.
- Le traitement par l'enzyme Plasmid-Safe™ (Epicentre, USA), ne permet pas l'obtention d'une hybridation plus intense.

La dégradation de l'ADN chromosomique d'*E.coli* éventuellement présent, permet d'obtenir une solution plasmidique plus pure et en théorie un plasmide sur la puce plus accessible pour la sonde. De plus, la présence de cet ADN peut également engendrer un certain bruit de fond par fixation aspécifique de la sonde. Ce traitement n'améliorant pas significativement la qualité de l'hybridation, on peut conclure que la purification du plasmide discrimine de façon satisfaisante l'ADN plasmidique de l'ADN chromosomique. Dans notre étude nous nous intéresserons à des signaux d'hybridation relativement forts. L'hybridation parasite avec la présence éventuelle d'ADN chromosomique d' *E.coli* n'engendrant aucun bruit de fond, un traitement préliminaire par cette enzyme ne s'avère donc pas nécessaire.

Bien que les produits PCR présentent une hybridation de meilleure qualité que les solutions plasmidiques correspondantes, cette amélioration ne permet pas la justification de la mise en oeuvre (PCR, Dépôt sur gel pour vérification, purification) de la synthèse de l'ensemble des inserts par cette technique à partir des plasmides purifiés.

Conclusion

En résumé, les différents traitements testés (DNA free enzyme, *EcoRI* et chauffage) n'augmentent pas l'efficacité d'hybridation de la sonde. La puce à ADN de type « shotgun » peut donc être faite en déposant directement les solutions de plasmides issues des purifications sur plaques Millipore. Ainsi l'ensemble des 9600 plasmides constituant la banque d'ADN génomique de TGH 9011 est utilisée directement pour servir à la construction de la puce à ADN de type « shotgun ».

Chapitre IV

Identification des gènes spécifiques à la souche
Campylobacter jejuni 81-176

1 Introduction

A la suite du succès de l'étude des gènes uniques de la souche *C. jejuni* TGH 9011, nous avons appliqué cette approche expérimentale de puce à ADN de type "shotgun" à la souche *C. jejuni* 81-176.

La souche 81-176 est l'une des souches de *C. jejuni* les plus étudiées. Des essais cliniques menés sur de nombreux modèles animaux et sur l'Homme ont montré que cette souche était inductrice de diarrhées inflammatoires (23). Tout comme *C. jejuni* TGH 9011 cette souche est capable d'envahir des cellules épithéliales humaines *in vitro* (80,130).

L'étude de *C. jejuni* 81-176, présentée dans la seconde publication « Genomic Diversity in *Campylobacter jejuni*: Identification of *C. jejuni* 81-176 Specific Genes », a permis l'identification de nouvelles séquences contenant 120 ORFs, représentant un total de 108 kb. Seule l'une d'entre elles est commune à la souche TGH 9011. La structure génétique de 81-176 est donc différente de celle de TGH 9011. La souche 81-176 comporte deux larges plasmides, pVir et pTet. Le premier a été séquencé récemment. Les clones de la librairie génomique contenant la séquence d'ADN du deuxième plasmide, pTet, ont été identifiés par son hybridation sur la puce. Les résultats complets concernant ce plasmide sont présentés en résultats additionnels faisant suite à la publication relative à la souche 81-176. Ce plasmide contient, tout comme pVir, un grand nombre de gènes impliqués dans le système de sécrétion de type IV. Tout comme chez TGH 9011, nous avons identifié sur le chromosome de 81-176 de nombreux gènes impliqués dans des mécanismes de pathogénèse, tels que des gènes impliqués dans la glycosylation des LOS, dans la formation de la capsule, et certains gènes présentant des homologies avec des protéines de restriction/modification.

2 Publication

Genomic Diversity in *Campylobacter jejuni*:
Identification of *C. jejuni* 81-176 Specific Genes

Frédéric Poly^{1,2}, Deborah Threadgill³, and Alain Stintzi^{1*}

¹Department of Veterinary Pathobiology, College of Veterinary Medicine
Oklahoma State University, Stillwater, OK 74078

²Dynamique, Evolution et Expression de Genomes de Microorganismes,
Université Louis Pasteur/CNRS FRE 2326, 67000 Strasbourg, France

³Department of Genetics, University of North Carolina
Chapel Hill, NC 27599-7264

Accepted in Journal of Clinical Microbiology

Since the publication of the complete genomic sequence of *C. jejuni* NCTC 11168 in February 2000, evidence has been compiling that suggests *Campylobacter jejuni* species exhibit high genomic diversity. In order to investigate this diversity, the unique genomic DNA sequences from a non-sequenced *Campylobacter* strain, *C. jejuni* 81-176, were identified by comparison with *C. jejuni* NCTC 11168, using a shotgun DNA microarray approach. Up to 63 kb of new chromosomal DNA sequences unique to this pathogen were obtained. Eighty-six ORFs were identified by the presence of uninterrupted coding regions encoding a minimum of 40 amino-acids. In addition, this study shows that the whole-plasmid shotgun microarray approach is effective and provides a comprehensive coverage of DNA regions that differs between two closely related genomes. The two plasmids harbored by this *Campylobacter* strain, pTet and pVir, were also sequenced with a coverage of 2.5 and 2.9-fold respectively, representing 72 and 92% of their complete nucleotide sequences. The unique chromosomal genes encode proteins involved in capsule and LOS biosynthesis, restriction and modification systems, and respiratory metabolism. Several of these unique genes are likely associated with *C. jejuni* 81-176 fitness and virulence. Interestingly, the comparison of *C.*

jejuni 81-176 unique genes with those of *C. jejuni* ATCC 43431 revealed a single gene which encodes a probable TraG-like protein. The product of this gene might be associated with the mechanism of *C. jejuni* invasion into epithelial cells. In conclusion, this study extends the repertoire of *C. jejuni* genes, and thus will permit the construction of a composite and more comprehensive microarray of *C. jejuni*.

Campylobacter jejuni causes an acute diarrheal disease with a variety of clinical symptoms such as fever, diarrhea, headache, abdominal pain, myalgia, vomiting and blood in feces (36). In addition, this bacterium is associated with the development of Guillain-Barré syndrome, an autoimmune mediated disorder of the peripheral nervous system (36). Until now, Penner serotyping, flagellin typing and pulsed-field gel electrophoresis (PFGE) have been the methods of choice for *Campylobacter* subtyping (37). The capsular polysaccharide is the major determinant for the heat-stable Penner serotyping scheme (18). Up to 60 Penner serotypes have been described, and certain serotypes appear to be preferentially associated with particular clinical outcomes (31). For instance, strains belonging to the O:19 serotype seem to predominate among *C. jejuni* isolated from Guillain-Barré patients (21). Epidemiological studies rely on these serotyping schemes. While these technologies are routinely used in clinical laboratories, there is a great interest in developing genomic-based approaches for in depth molecular typing.

The wealth of genome sequences information has driven the development and use of microarray approaches for the analysis of strain relatedness. In these genome-wide approaches, the “target” (not sequenced) genome is interrogated by the microarray for the presence or absence of genes belonging to the sequenced strain. Several reports, using this method, described an extensive variability in the gene content between the sequenced strain *C. jejuni* NCTC 11168 and strains of diverse origins (12, 23, 24, 30). While these studies proved to be extremely valuable for the identification of the minimum set of *C. jejuni* core genes, they failed to identify genotypic traits associated with phenotypic characteristics, such as serotypes, clinical outcomes, isolate origins or virulence.

Undoubtedly, the success of this methodology for strain typing or genotypic trait identification relies on the completeness and diversity of the arrayed “probe” for a particular species. Currently only one strain of *C. jejuni* has been completely sequenced (*C. jejuni* NCTC 11168) (29), and another one (*C. jejuni* RM 1221) is being sequenced at the TIGR institute. The addition of new genome information from *C. jejuni* strains of different origins will facilitate the identification of genotypic markers and improve the power of this genotyping method.

Recently, we have described a novel approach to identify non-redundant sequences of *C. jejuni* ATCC 43431 by combining the construction of a shotgun library and the power of microarray technology(32). This genome-wide approach consists of five consecutive steps: (i) construction of a shotgun library representative of the size of the genome, (ii) arraying of the whole plasmids on glass slides, (iii) competitive genomic hybridization, (iv) identification of the plasmids carrying unique DNA segments, and (v) DNA sequencing. A total of 84 kb of new DNA sequences unique to ATCC 43431 were identified (32). These sequences encode up to 130 genes, representing an addition of approximately 7.5% of novel genes to the current set of known *C. jejuni* genes. These unique genes likely constitute ideal targets for molecular markers of ATCC 43431 and other closely related strains.. However, in order to better define common and unique genes between *C. jejuni* strains, additional DNA sequence information from other *C. jejuni* isolates are needed.

In the present study, we propose to further increase the repertoire of genes belonging to the *C. jejuni* species by identifying the genes unique to *C. jejuni* 81-176. This isolate has been extensively studied and shown to carry 2 large plasmids, named pVir and pTet (1,

6). Recently, the DNA sequence of the pVir plasmid has been released (2). Interestingly, this plasmid harbors 54 predicted open reading frames encoding orthologs of components of *Helicobacter pylori* type IV secretion apparatus (2). Mutational analysis of 21 of these genes identified 7 genes encoding proteins required for a high level of invasion into INT407 cells. Interestingly, the pVir plasmid alone is insufficient to increase the low *C. jejuni* NCTC 11168 invasion ability (2). Given that the plasmid pTet seems to be non-essential for *C. jejuni* 81-176 invasion potential since pTet deficient 81-176 is still invasive, the chromosome of this strain likely carries genes encoding proteins involved in the invasion process. The identification of this chromosomal set of genes will provide a valuable resource to study *Campylobacter* invasiveness.

Materials and Methods.

Campylobacter strains. The two strains of *C. jejuni* used in this study, *C. jejuni* NCTC 11168 and 81-176, were cultured in Mueller-Hinton medium and incubated at 37°C under microaerophilic conditions (83% N₂, 4% H₂, 8% O₂ and 5% CO₂) maintained by using the MACS-VA500 variable atmosphere workstation (Don Whiteley, West Yorkshire, England). *C. jejuni* 81-176 was originally isolated from a human disease outbreak in 1985 and has been shown to cause inflammatory colitis in two human challenge studies (10, 20). Likewise, this isolate caused acute diarrheal illness in monkey (33), mouse (4), and ferret animal models (1).

Construction of the C. jejuni 81-176 shotgun microarray. The *C. jejuni* 81-176 shotgun microarray was constructed as previously described for *C. jejuni* ATCC 43431 (32). Briefly, the construction of this microarray involved 4 consecutive steps: (1) construction of a genomic library of *C. jejuni* 81-176, (2) plasmid extraction and purification, (3) assessment of the plasmid quality and concentration by agarose gel electrophoresis, and (4) printing of the whole plasmids on microarray slides. The genomic library was constructed by cloning nebulized genomic DNA fragments of 400 to 800 bp into the pCR[®]4Blunt-TOPO[®] vector from Invitrogen, according to the manufacturer's protocol and as previously described (32). This library was produced in TOPO10 *E. coli* and replicated into 96-well plates containing LB broth and ampicillin (100 µg/ml). Plasmids from 9600 clones were purified using the 96 well filter plates Millipore Multiscreen (Millipore MAFB NOB 50) and Costar PVDF (0.2 µm Costar) following the manufacturers' instructions. The yield of plasmid purification was estimated by agarose

gel electrophoresis, and plasmid concentrations were adjusted to approximately 100 ng/ul by dilution in 3xSSC buffer. Whole plasmids were printed on aminosilane-coated glass slides (CMT GAPS-II from Corning Inc., Corning, N.Y.) using a Molecular Dynamic arrayer. Plasmids were immobilized on slides by baking at 80°C for 4 hours. The diameter of each spot was about 200 µm with a spacing of 250 µm apart center from center.

Preparation of fluorescently labeled probes, hybridization, scanning and data analysis.

Fluorescently labeled probes were prepared from 2 µg of genomic DNA by random priming using the BioPrime DNA labeling system from Invitrogen in the presence of aminoallyl-dUTP, followed by coupling with either the Cy3 or Cy5 monoreactive dye (Amersham Pharmacia, NJ) as previously described (32). Before hybridization, the slides were pre-hybridized in 1% BSA, 25% formamide, 5xSSC buffer, and 0.1% SDS, for 45 min at 42°C. Next, the slides were briefly rinsed with water and spun dried. Both labeled probes were combined in a 36 µl hybridization solution consisting of 25% formamide, 5xSSC buffer, 0.1% SDS and 25 µg of salmon sperm. This hybridization solution was heated at 99°C for 2 min, cooled down to 42°C, and then placed on the microarray slide and covered by a coverslip (Fisher). The microarray was placed in a humidified chamber (ArrayIt™), and incubated at 42°C. After 16 hours, the slides were washed for 5 min with 2xSSC, 0.1% SDS, which was preheated to 42°C. Next, the slides were washed at room temperature in 0.1xSSC, 0.1% SDS for 10 min, followed by 4 washes of 1 min in 0.1xSSC. Finally, the slides were briefly rinsed in water and spun dried.

The microarray slides were scanned using a laser-activated confocal scanner (ScanArray 5000) at 532 and 635 nm. Fluorescent signal values were obtained by using the GenePix Pro4 software (Axon Instrument, Foster City, Calif.). Spots were removed from further analysis if they were localized in a region of hybridization abnormalities or if the signal intensity in the channel corresponding to *C. jejuni* 81-176 was 2 times below the standard deviation of the background. The fluorescence intensities were normalized by applying an iterative linear regression analysis as described previously (32). The microarray hybridization was repeated three times. The 3 microarray data were merged, and the mean of the fluorescence ratio (Cy5/Cy3) for each spot was computed. Then, the spots were sorted by descending ratio (signal intensity corresponding to *C. jejuni* 81-176 / signal intensity corresponding to *C. jejuni* NCTC 11168) and the DNA inserts were gradually sequenced in 96 well plates formats. The sequencing was stopped when the DNA sequence of at least 90% of the clones from a 96 well plate exhibited 100% identity with the genomic sequence of *C. jejuni* NCTC 11168.

DNA sequencing, ORF finding and annotation. DNA sequences were determined in both directions using the M13 universal primers by the DNA sequencing facility of the Laboratory of Genomics and Bioinformatics at University of Oklahoma Health Science Center (OUHSC, Oklahoma City). Vector trimming and sequence assembly were performed automatically by using the Lasergene software (DNASTar, Inc.). Each contig was manually edited and checked for residual error or misassembling. The sequences exhibiting 100% identity with *C. jejuni* NCTC 11168 were identified by BLASTN. These sequences correspond to false positives and were therefore discarded from further

analysis. Potential open reading frames were identified by using Artemis (available at <http://www.sanger.ac.uk/software/artemis>). Similarities searches were performed with the BLASTP and BLASTX programs against the NCBI non-redundant database, and the genes were grouped based on functional categories.

Plasmid extraction, labeling and sequencing. The plasmids pVir and pTet were extracted together using the Qiagen plasmid Midi kit following the manufacturer's instructions. The plasmids were labeled with the Cy3 monoreactive dye and applied to the shotgun array as described above for genomic DNA. Probes with fluorescent signal intensities above three times the standard deviation of the background were considered to carry DNA fragments belonging to one of the two plasmids. The corresponding probes were retrieved from the clone library and the inserts sequenced. The DNA sequences, which belong to the pTet or pVir plasmids, were assembled into contigs using the Lasergene software (DNASTar, Inc.) as described above. The BLASTN program was used to identify the sequences exhibiting 100% identity with the pVir (NCBI accession number AF226280) or pTet plasmid (NCBI accession number AY394561).

PFGE preparation and analysis Agarose plugs containing 81-176 and NCTC 11168 were prepared as described previously (33) Plugs with bacterial genomic DNA were digested with three restriction enzymes chosen based on an in silico digestion of NCTC 11168 to yield 10 or less fragments (www.insilico.com). Digests were performed with AsiSI, BsiWI, and SalI according to manufacturer's specifications, with AsiSI and SalI digests at 37°C and BsiWI at 55°C. Electrophoresis conditions were the same for the

three digests, and were as previously described for SalI, initial switch time of 50 sec. and final switch time of 90 sec.(33). These conditions are optimized to separate the concatemerized lamda markers (MW range 48.5Kb-1Mb, BioRad) used to estimate the bacterial genomic fragment sizes. Band sizes were determined as before (UV Doc It system, LabWorks Image Acquisition and Analysis Software) (33) and bands resulting from the two 81-176 plasmids were not visualized for any of the digests. The smallest fragment size seen for 81-176 was greater than 100 Kb and thus larger than either plasmid. The amount of genomic DNA in the plugs was estimated to be no more than 0.3 µg, which contributed to the lack of visible plasmid bands. The average genome size for 81-176 and 11168 was estimated from the fragment sizes for the three enzymes and was $1.736 \pm .0066$ Mb for 81-176 and $1.735 \pm .011$ Mb for 11168.

Nucleotide sequence accession numbers. The 58 chromosomal DNA fragments unique to *C. jejuni* 81-176 were submitted to the NCBI database and assigned accession numbers AY681239 to AY681296.

Results

Construction of the *C. jejuni* 81-176 shotgun microarray. Our objective was to identify the set of genes unique to *C. jejuni* 81-176 as compared to the genome-sequenced strain *C. jejuni* NCTC 11168 (29). The approach adopted for this study was previously developed for the identification of *C. jejuni* ATCC 43431 unique genes (32). This approach combines the construction of a whole plasmid DNA microarray obtained from a shotgun library of *C. jejuni* 81-176, as well as the competitive hybridization to the array of genomic DNA from strains *C. jejuni* 81-176 and NCTC 11168.

The sizes of the *C. jejuni* 81-176 and NCTC 11168 chromosomes were estimated to be similar in length ($1,735 \pm 10$ kb) by pulsed-field gel electrophoresis analysis (data not shown). However, in contrast to *C. jejuni* 11168, the 81-176 strain carries two large plasmids named pVir and pTet (2, 6). While the complete DNA sequence of the 37.5 kb pVir plasmid has been previously reported (2), the genetic information of the pTet plasmid, with the exception of the presence of a gene conferring tetracycline resistance, was initially unknown. However, the complete pTet plasmid was sequenced by Batchelor *et al.* during the course of the present study (6). Its sequence and annotation were recently made accessible at the NCBI database under the accession number AY394561.

A shotgun library of the *C. jejuni* 81-176 genome consisting of 9600 individual clones and comprising both the chromosomal and plasmidic DNAs was constructed in the *E. coli* strain TOP10 as previously described (32), and arrayed on glass slides. The mean DNA insert size was estimated (subsequently subjected to DNA sequencing and analysis) to be approximately 0.5 kb, and ranged from 0.1 to 1.5 kb. Based on the size of the genome (1,724 kb which corresponds to the sum of the sizes of the chromosome and the

two plasmids), the insert (0.5 kb), and the shotgun library (9,600 clones), the probability of a DNA fragment to be present on our array is equal to 0.94 ($P = 1 - e^{(9,600 \times \ln(1-0.5/1,724))}$).

Identification of clones carrying unique DNA sequences. The unique DNA sequences of *C. jejuni* 81-176 belonging to the chromosome and the two plasmids were identified by two different sets of microarray hybridization experiments.

The first set of experiments allowed the identification of every unique DNA sequence present on our shotgun array. These DNA sequences were identified by competitive hybridization on the array of genomic DNA from *C. jejuni* 81-176, with genomic DNA from *C. jejuni* NCTC 11168. The microarray hybridization was repeated three times. The microarray data was normalized using a linear regression analysis as previously described (32), and the mean of the fluorescent intensities ratio [(fluorescent intensity corresponding to the genomic DNA of *C. jejuni* 81-176)/(fluorescent intensity corresponding to the genomic DNA of *C. jejuni* NCTC 11168)] was computed. The ratios were sorted in descending number and the inserts of the first 96 plasmids with the highest ratio value were sequenced. The DNA sequences obtained were compared to the *C. jejuni* NCTC 11168 genome using the BLASTN program, and the inserts of the next 96 plasmids with the highest ratio value were sequenced until more than 90% of the sequences within a block of 96 plasmids shared 100% identity over their entire length with *C. jejuni* NCTC 11168. A total of 1056 plasmids were sequenced generating approximately 130 kb of DNA sequences unique to *C. jejuni* 81-176.

The second set of microarray experiments allowed the identification of the unique DNA sequences belonging to the pVir and pTet plasmids. The plasmids were extracted from *C.*

jejuni 81-176 using the Qiagen plasmid midi kit, labeled with the Cy3 fluorescent dye, and co-hybridized on the shotgun array with Cy5 labeled genomic DNA. Spots with a signal intensity above 2 times the standard deviation of the background in the Cy3 channel were selected as containing DNA fragments belonging to one of the two plasmids. With this criterion, 388 clones were selected and their insert sequenced. Of the 388 sequences obtained, 212 and 176 of them were identified as belonging to the pVir and pTet plasmids respectively by using the BLASTN program at the NCBI. Furthermore, given that 388 clones contained DNA inserts of plasmidic origin, 668 plasmids among the 1056 selected by the first set of microarray experiments should contain DNA inserts of chromosomal origin. To note, this second set of microarray experiments permitted the efficient differentiation between chromosomic and plasmidic DNA sequences. Indeed, only one clone out of the 668 that were assigned as containing an insert of chromosomal origin was found to harbor a DNA sequence belonging to one of the plasmids.

Sequence assembling and assessment of the shotgun microarray approach to identify unique DNA fragments. The DNA sequences were first split based on their plasmidic or chromosomic origins, and second assembled into contigs using the SeqMan II program from Lasergene (DNASTAR, Inc.). A total of 86 contigs, yielding 130 kb of DNA sequences unique to *C. jejuni* 81-176, were generated by this assembly process. The size distribution of these contigs for the chromosome and the two plasmids is shown in Table 1. There are 12 and 13 gaps remaining in the sequence of the pTet and pVir plasmids respectively. Based on the total number of base pairs sequenced (177.5 kb for the

chromosome, 98.8 kb for pVir, and 82.3 kb for pTet), the DNA sequences identified in this study represent about 2.5-times, 2.9-times, and 2.8-times coverage for the size of the pTet, pVir and chromosome of *C. jejuni* 81-176 respectively. The lower coverage obtained for the pTet plasmid is likely the result of an under-representation of its DNA sequences on the array. The genomic DNA used to construct the shotgun library was extracted from *C. jejuni* 81-176 grown in the absence of tetracycline. Given the observed high frequency (12%) of spontaneous loss of this plasmid in the absence of a selective pressure (tetracycline) (2), the DNA sequences from this plasmid is likely under-represented in our library and therefore on the array. Concomitant to this study, the complete sequence of the pTet plasmid was released by Batchelor *et al.* to the scientific community at the NCBI database (6). This plasmid was found to be 45.2 kb in length. Therefore, a 2.5-times coverage of the size of the pTet plasmid yielded 72% of its 45.2 kb, while a 2.9-times coverage of the size of the pVir plasmid yielded 92% of its 37.5 kb. Consequently, a 2.8-times coverage of the unique DNA fragments from the chromosome of *C. jejuni* 81-176 likely yielded between 72 and 92% of the total unique nucleotides.

Sequence and analysis of the DNA fragments unique to the chromosome of C. jejuni 81-176. As shown in Table 2, 87 ORFs unique to the chromosome of *C. jejuni* 81-176 were identified. The only criterion for including an ORF in this group was the presence of an uninterrupted coding frame of at least 120 nucleotides in length. Of the 87 ORFs, 69 were assigned a potential function based on BLASTP findings. The remaining 18 were unknown. The mean G+C content of the unique ORFs was found to be equal to 28.3%, slightly lower than the genome of *C. jejuni* NCTC 11168 (30.6%). These unique genes

can be grouped into 7 functional categories: (1) cell envelope and surface structure (mainly capsule and lipooligosaccharide), (2) restriction and modification systems, (3) small molecule metabolism, (4) transport, (5) antibiotic resistance, (6) others and hypothetical, and (7) unknown.

A large number of the unique genes encode putative proteins involved in the biosynthesis of the capsule and lipooligosaccharide. The entire capsular polysaccharide biosynthetic locus of *C. jejuni* 81-176 was previously identified and sequenced by Karlyshev *et al.* (GenBank accession number BX545858). The identification of genes encoding the proteins involved in its biosynthesis is in agreement with the difference in Penner serotype of the two *C. jejuni* strains. *C. jejuni* 81-176 is Penner serotype O:23/36, while NCTC 11168 is Penner serotype O:2. Noteworthy, the DNA sequence comparison of this locus with the unique genes identified here and the corresponding locus in *C. jejuni* NCTC 11168 revealed that the shotgun microarray approach employed in this study identified all of the expected unique genes, highlighting the power of the comparative shotgun microarray technology. The identification of *C. jejuni* 81-176 unique genes encoding putative proteins involved in the LOS biosynthesis is in agreement with the reported high divergence of this locus between strains of *C. jejuni* (12, 23, 24, 30, 32). In fact, the *C. jejuni* NCTC 11168 genes associated with the outer-core LOS biosynthesis were previously found to be absent or highly divergent in *C. jejuni* 81-176 by whole genome comparison (12, 23).

Another functional group of unique *C. jejuni* 81-176 ORFs encode several probable type I and type II restriction modification (R-M) systems (Cj81-032, Cj81-057, Cj81-060, Cj81-061, Cj81-087, Cj81-118, and Cj81-126). As seen in Table 2, Cj81-118 and Cj81-

126 exhibit identity with the same protein, a type I restriction-modification system methyltransferase subunit (HsdM) from *Trichodesmium erythraeum*, and thus they might be part of the same gene. Similarly, Cj81-060 and Cj81-061 might constitute a single gene encoding a protein with homology to a type I restriction-modification enzyme (HsdR) of *Vibrio cholerae*. A type I R-M system is constituted of 3 components: a restriction enzyme (HsdR), a methyltransferase (HsdM), and a site specific subunit (HsdS) (27). Therefore, Cj81-60/Cj81-61 and Cj81-118/Cj81-126, might be part of the same R-M system with the site specific subunit encoded by Cj81-032 or Cj81-057.

A noteworthy set of unique *C. jejuni* 81-176 ORFs encode proteins involved in bacterial respiratory metabolism, including several cytochrome C biogenesis proteins (Cj81-007, Cj81-097, Cj81-099, and Cj81-124), a cytochrome b subunit of formate dehydrogenase (Cj81-005), 2 components of a putative anaerobic dimethyl sulfoxide (DMSO) reductase (Cj81-083 and Cj81-105), and a nitrate/trimethylamine-N-oxide (TMAO) membrane-bound reductase (Cj81-098). These different proteins may allow *C. jejuni* to utilize alternative electron acceptors to oxygen under oxygen-restricted conditions, such as DMSO, TMAO, and nitrate. The DMSO reductase system is found in various bacteria, including *Escherichia coli* (7), *Actinobacillus pleuropneumonia* (3), *Rhodobacter sphaeroides* (17), *Rhodobacter capsulatus* (26) and *Haemophilus influenzae* (25). The *E. coli* anaerobic DMSO reductase is composed of three subunits, a catalytic subunit containing a molybdopterin cofactor (DsmA), a transmembrane electron carrier (DsmB), and a membrane anchor (DsmC) (8). Based on the BLASTP results, the *C. jejuni* 81-176 DsmA and DsmC proteins are encoded by Cj81-105 and Cj81-083, respectively. While a DmsB homolog was not recognized, several of the identified genes might encode a

protein with a DmsB-like function, such as the cytochrome C or the nitrate/TMAO reductase. DmsA requires molybdenum for its catalytic activity, however the *C. jejuni* NCTC 11168 genes (*modABC*) encoding the molybdenum transport apparatus were found to be missing in *C. jejuni* 81-176 (12). Interestingly, a gene encoding a probable molybdenum transport ATP-binding protein (a ModC homolog) was identified on the chromosome of *C. jejuni* 81-176 (Cj81-125), suggesting that this strain might be able to acquire molybdenum. However, homologs of ModA and ModC were not identified.

A notable ORF of *C. jejuni* 81-176 (Cj81-027) appears to be identical to a unique ORF from *C. jejuni* ATCC 43431 (Tgh073) (32). The sequences of these two ORFs overlap over a 78 bp nucleotide region with a 100% identity, suggesting they might be part of the same gene. In order to demonstrate that these 2 ORFs (Tgh073 and Cj81-027) are part of the same gene, a set of PCR primers was designed to amplify a DNA fragment encompassing both ORFs. As shown in Figure 1, the PCR amplification using either the chromosomal DNA of *C. jejuni* 81-176 (lane 1) or ATCC 43431 (lane 2) yielded a product of the expected size, demonstrating that both ORFs belong to the same gene. Interestingly, Tgh073 was shown to encode a homologue of TraG from *Neisseria gonorrhoeae* (11). In other organisms, TraG participates in the formation of a conjugative DNA transfer system and a type IV secretion apparatus (14, 22). In fact, TraG-like proteins have been proposed to play an essential role in the secretion of substrates through these apparatus, such as the cytotoxin CagA protein in *Helicobacter pylori* (34). Other unique *C. jejuni* 81-176 genes encode probable transporters, an efflux protein, and several hypothetical and unknown proteins.

Discussion

Despite the absence of clear evidence for the existence of *Campylobacter* pathotypes, epidemiological and phenotypical studies suggest that strains of *C. jejuni* vary in their colonization and invasion ability and thus likely in their virulence potential (13, 15, 16, 19, 28). Comparative genomic studies showed that there is substantial genetic diversity between isolates of *C. jejuni* (12, 23, 30). This diversity might enable *Campylobacter* to colonize various environmental niches and possibly contribute to its virulence. So far, these comparative genomic studies have been restricted to the investigation of the presence or absence of genes from the genome-sequenced strain *C. jejuni* NCTC 11168 in other *Campylobacter* isolates of various origins (12, 23, 30). Obviously, genes unique to non-sequenced strains cannot be identified by this classical approach. While the genomic subtractive hybridization procedure is the method of choice for the identification of strain specific DNA sequences, the generation of comprehensive coverage is labor intensive and tedious. As a result, this technology usually provides only a handful of new DNA fragments. Here, we describe and assess the use of a whole plasmid shotgun microarray technology to identify DNA fragments of *C. jejuni* 81-176 absent in *C. jejuni* NCTC 11168. This technology presents several advantages for identifying unique genes. Whole plasmid shotgun arrays are relatively simple and inexpensive to construct. These arrays are extremely versatile and the technology could be virtually adapted to any organism. Given the sequence of a reference strain, this approach enables the identification of novel and unique DNA sequences on other strains without the need of their whole genome sequences.

The results presented here demonstrate the efficiency of the whole plasmid shotgun microarray approach and add 87 novel chromosomal genes to the gene repertoire of *C.*

jejuni. These additional genes will allow the construction of a more comprehensive microarray of *C. jejuni*, which will likely facilitate the identification of genotypic markers and then improve the power of genotyping methods. The genome coverage of this technology was assessed by three parameters: (i) the number of times that the same DNA fragment was sequenced; (ii) the effectiveness of identifying genes previously known to be unique to *C. jejuni* 81-176; and (iii) the nucleotide coverage of the two plasmids pVir and pTet. Each chromosomal DNA fragment was sequenced on average 2.8 times, and among the 16 previously identified unique chromosomal *C. jejuni* 81-176 genes, only one was not detected by our study (NCBI accession number AAS46249). The failure to identify this gene is likely due to its relatively short length, 279 nucleotides, and the presence of flanking DNA sequences (in 5' and 3') with 100% identity to *C. jejuni* NCTC 11168. The two plasmids, pTet and pVir, were sequenced with a coverage of 2.5 and 2.9-fold respectively, representing 72 and 92% of their complete nucleotide sequences. All together, these results show that whole plasmid shotgun microarray provides an effective and comprehensive approach to identify novel genes without whole genome sequencing. Although this technology appears to be a powerful method for the identification of novel DNA fragments, this technique has its limitations. First, it is not possible to detect point mutations, gene location, or chromosomal rearrangements. Second, the boundaries of the novel genomic segments are not identified, probably due to a high level of cross-hybridization of these regions with the chromosomal DNA of *C. jejuni* NCTC 11168. Nevertheless, this technology allowed the comprehensive detection of the unique genes of *C. jejuni* 81-176 strain and thus holds promise as an efficient tool for comparative genomics .

It should be noticed that this study does not fully address the function of unique genes, and their possible role in virulence remains speculative. However, the in silico functional analysis of the unique *C. jejuni* 81-176 genes reveals several interesting traits of this isolate. The unique genes are characterized by three major functional groups: (1) capsule and LOS biosynthesis; (2) restriction and modification systems; and (3) respiratory metabolism. The *C. jejuni* NCTC 11168 capsule and LOS loci have been previously shown to be divergent in *C. jejuni* 81-176 (12, 23). This divergence has been attributed to the difference in Penner serotype between the two strains, O:23/36 and O:2 for *C. jejuni* 81-176 and NCTC 11168, respectively. Therefore, the identification of *C. jejuni* 81-176 unique genes encoding enzymes involved in the biosynthesis of these surface structures was expected and is in agreement with the diversity of these loci. Noteworthy, the unique genes involved in the capsule biosynthesis encode proteins that exhibit high identity with proteins of similar function in *C. jejuni* ATCC 43456 (Table 2), which belongs to the same Penner serotype as *C. jejuni* 81-176 (Penner serotype O:36), corroborating the role of the capsule as the serodeterminant in the Penner serotyping scheme (18). A number of unique genes encode proteins with a high BLASTP match with components of type I restriction and modification (R-M) systems. Interestingly, the presence of genes encoding type I R-M systems was found to correlate with the extent of the host gastric response to *H. pylori* (9). Consequently, these enzymes have been proposed to affect *Helicobacter pylori* pathogenesis by modulating its gene expression. Based on this observation, it is tempting to propose an essential role for these R-M systems in the ability of *Campylobacter* to induce gastroenteritis. Namely, the presence or absence of a particular R-M system might modulate the virulence potential of *Campylobacter* strains.

Another important group of genes identified in this study encode proteins involved in respiratory metabolism. *C. jejuni* is a microaerophilic bacteria with a growth requirement for oxygen. Recently, oxygen has been proposed to be essential for DNA synthesis due to the use of an oxygen-dependent class I-type ribonucleotide reductase (35). Nevertheless, *C. jejuni* NCTC 11168 has been shown to be able to carry out anaerobic-like respiration and use a wide range of alternative electron acceptors to oxygen, such as fumarate, nitrate, nitrite, TMAO, and DMSO (35). The genome of *C. jejuni* 81-176 has been shown to contain all of the NCTC 11168 genes that encode the predicted terminal reductases involved in these alternative respiratory pathways (12, 23). Therefore, the presence in *C. jejuni* 81-176 of additional genes encoding proteins with a potential function in electron transport suggests that this isolate might possess a capacity of adaptation to different ecological niches. In particular, the *C. jejuni* 81-176 genome contains an anaerobic dimethyl sulfoxide (DMSO) reductase. In *E. coli*, the DMSO reductase allows cell respiration under reduced-oxygen or anaerobic conditions by utilizing a broad range of various N-oxide and sulfoxide compounds as terminal electron acceptors (38). This wide range of substrate specificity may allow *C. jejuni* 81-176 to grow under low oxygen level on a diverse array of N- and S-oxide compounds, such as trimethylamine-N-oxide (TMAO) and DMSO. Interestingly, the DMSO reductase of *Actinobacillus pleuropneumoniae* has been shown to significantly contribute to the virulence of this pathogen (3). An *A. pleuropneumonia* mutant lacking the DMSO reductase was attenuated in acute disease using a pig infection model, suggesting a role for alternative respiratory pathways in necrotic lung tissue (3). However, the nature of the substrate for this enzyme in the lung tissue is unclear since DMSO and TMOA are essentially found in

marine organisms (such as fish and aquatic invertebrates) and algae (5). Consequently, this enzyme might allow *C. jejuni* 81-176 to colonize seafood, and/or to enhance gastrointestinal tract colonization if their hosts eat seafood.

Recently, we described the characterization of *C. jejuni* ATCC 43431 unique genes (32). The comparison of these unique genes with those of *C. jejuni* 81-176 uncovered some additional information. These two strains of *C. jejuni* appear to share only one unique gene. Interestingly this gene encodes a protein with identity to TraG from *Neisseria gonorrhoeae*. This TraG-like protein has been shown to be involved in mating and to also function as an essential component of the type IV secretion apparatus in several pathogens, including *H. pylori* (14, 22, 34). The function of this gene in *C. jejuni* and its presence in both strains are intriguing. In contrast to *C. jejuni* NCTC 11168, both of these strains are highly invasive into intestinal epithelial cells (2, 32). The invasion efficiency of *C. jejuni* 81-176 has been shown to be dependent on a type IV secretion system encoded by the plasmid pVir (2). However, the pVir plasmid was found to be insufficient for the high level of invasion observed. Furthermore, the transformation of the pVir plasmid into *C. jejuni* NCTC 11168 did not result in an increase of its level of invasion into INT407 cells (2). Therefore, although the pVir plasmid was required for a high level of invasion, chromosomally encoded protein(s) absent in NCTC 11168 was proposed to be also necessary (2). While *C. jejuni* ATCC 43431 does not harbor the pVir plasmid, its chromosome contains genes encoding components with homology to the type IV secretion system (32). Consequently, it is tempting to propose a role for the chromosomally encoded TraG protein in the formation of a type IV secretion system that might be involved in the invasion process into epithelial cells. We are currently

investigating this possibility. Notably, a TraG homologue was also found on the pTet plasmid (6). This plasmid has been shown to be self-mobilizable and to carry 10 genes encoding proteins involved in conjugation (6). While these genes also encode a type IV secretion system, this system seems to serve a function different than the one encoded by pVir and to be required for DNA transfer. In addition, this plasmid was found to be fully dispensable for invasion into epithelial cells (2). Because of the sufficiency of the pTet plasmid encoded DNA transfer functions, the chromosomally encoded TraG-like is unlikely to contribute to bacterial conjugation.

In conclusion, whole plasmid shotgun microarray is a powerful and versatile method that allows the comprehensive identification of new genomic DNA sequences from a non-sequenced bacterial strain by comparison with a related species or strain of known sequence. In this study, we identified the repertoire of genes unique to *C. jejuni* 81-176 by comparison to *C. jejuni* NCTC 11168. Interestingly, several of these unique genes might be associated with the *C. jejuni* 81-176 fitness and virulence. However, the true function of these genes need to be addressed by targeted mutagenesis and phenotypic analysis. The results presented here demonstrate that there is a substantial genetic diversity between isolates of *C. jejuni*. Therefore the study of additional strains of *C. jejuni* would be valuable to further increase the repertoire of genes. The identification of additional genes would allow the construction of a composite *Campylobacter* microarray. This microarray could then be used to identify unique and shared genes between strains of various origins. Ultimately, this type of comparative study may prove to be useful for the development of molecular typing methods. Finally, such microarrays could constitute a useful tool for epidemiological and evolutionary studies.

Acknowledgments

We would like to thank I. Turcot for her contribution to this manuscript. The authors are grateful to all the staff from OU Health Science Center Laboratory for Genomics and Bioinformatics and the University of Oklahoma's Microarray and Bioinformatics Core facilities. This work was supported by the National Institutes of Health grants RO1-AI055612 and RR15564.

References

1. **Bacon, D. J., R. A. Alm, D. H. Burr, L. Hu, D. J. Kopecko, C. P. Ewing, T. J. Trust, and P. Guerry.** 2000. Involvement of a plasmid in virulence of *Campylobacter jejuni* 81-176. *Infect Immun* **68**:4384-90.
2. **Bacon, D. J., R. A. Alm, L. Hu, T. E. Hickey, C. P. Ewing, R. A. Batchelor, T. J. Trust, and P. Guerry.** 2002. DNA sequence and mutational analyses of the pVir plasmid of *Campylobacter jejuni* 81-176. *Infect Immun* **70**:6242-50.
3. **Baltes, N., I. Hennig-Pauka, I. Jacobsen, A. D. Gruber, and G. F. Gerlach.** 2003. Identification of dimethyl sulfoxide reductase in *Actinobacillus pleuropneumoniae* and its role in infection. *Infect Immun* **71**:6784-92.
4. **Baqar, S., A. L. Bourgeois, L. A. Applebee, A. S. Mourad, M. T. Kleinosky, Z. Mohran, and J. R. Murphy.** 1996. Murine intranasal challenge model for the study of *Campylobacter* pathogenesis and immunity. *Infect Immun* **64**:4933-9.
5. **Barrett, E. L., and H. S. Kwan.** 1985. Bacterial reduction of trimethylamine oxide. *Annu Rev Microbiol* **39**:131-49.
6. **Batchelor, R. A., B. M. Pearson, L. M. Friis, P. Guerry, and J. M. Wells.** 2004. Nucleotide sequences and comparison of two large conjugative plasmids from different *Campylobacter* species. *Microbiology* **150**:3507-3517.
7. **Bilous, P. T., S. T. Cole, W. F. Anderson, and J. H. Weiner.** 1988. Nucleotide sequence of the *dmsABC* operon encoding the anaerobic dimethylsulphoxide reductase of *Escherichia coli*. *Mol Microbiol* **2**:785-95.
8. **Bilous, P. T., and J. H. Weiner.** 1988. Molecular cloning and expression of the *Escherichia coli* dimethyl sulfoxide reductase operon. *J Bacteriol* **170**:1511-8.

9. **Bjorkholm, B. M., J. L. Guruge, J. D. Oh, A. J. Syder, N. Salama, K. Guillemin, S. Falkow, C. Nilsson, P. G. Falk, L. Engstrand, and J. I. Gordon.** 2002. Colonization of germ-free transgenic mice with genotyped *Helicobacter pylori* strains from a case-control study of gastric cancer reveals a correlation between host responses and HsdS components of type I restriction-modification systems. *J Biol Chem* **277**:34191-7.
10. **Black, R. E., M. M. Levine, M. L. Clements, T. P. Hughes, and M. J. Blaser.** 1988. Experimental *Campylobacter jejuni* infection in humans. *J Infect Dis* **157**:472-9.
11. **Dillard, J. P., and H. S. Seifert.** 2001. A variable genetic island specific for *Neisseria gonorrhoeae* is involved in providing DNA for natural transformation and is found more often in disseminated infection isolates. *Mol Microbiol* **41**:263-77.
12. **Dorrell, N., J. A. Mangan, K. G. Laing, J. Hinds, D. Linton, H. Al-Ghusein, B. G. Barrell, J. Parkhill, N. G. Stoker, A. V. Karlyshev, P. D. Butcher, and B. W. Wren.** 2001. Whole genome comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity. *Genome Res* **11**:1706-15.
13. **Everest, P. H., H. Goossens, J. P. Butzler, D. Lloyd, S. Knutton, J. M. Ketley, and P. H. Williams.** 1992. Differentiated Caco-2 cells as a model for enteric invasion by *Campylobacter jejuni* and *C. coli*. *J Med Microbiol* **37**:319-25.

14. **Frost, L. S., K. Ippen-Ihler, and R. A. Skurray.** 1994. Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol Rev* **58**:162-210.
15. **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**:503-17.
16. **Hanel, I., J. Muller, W. Muller, and F. Schulze.** 2004. Correlation between invasion of Caco-2 eukaryotic cells and colonization ability in the chick gut in *Campylobacter jejuni*. *Vet Microbiol* **101**:75-82.
17. **Johnson, J. L., N. R. Bastian, and K. V. Rajagopalan.** 1990. Molybdopterin guanine dinucleotide: a modified form of molybdopterin identified in the molybdenum cofactor of dimethyl sulfoxide reductase from *Rhodobacter sphaeroides* forma specialis denitrificans. *Proc Natl Acad Sci U S A* **87**:3190-4.
18. **Karlyshev, A. V., D. Linton, N. A. Gregson, A. J. Lastovica, and B. W. Wren.** 2000. Genetic and biochemical evidence of a *Campylobacter jejuni* capsular polysaccharide that accounts for Penner serotype specificity. *Mol Microbiol* **35**:529-41.
19. **Konkel, M. E., and L. A. Joens.** 1989. Adhesion to and invasion of HEp-2 cells by *Campylobacter* spp. *Infect Immun* **57**:2984-90.
20. **Kopecko, D. J., L. Hu, and K. J. Zaal.** 2001. *Campylobacter jejuni*--microtubule-dependent invasion. *Trends Microbiol* **9**:389-96.

21. **Kuroki, S., T. Saida, M. Nukina, T. Haruta, M. Yoshioka, Y. Kobayashi, and H. Nakanishi.** 1993. *Campylobacter jejuni* strains from patients with Guillain-Barre syndrome belong mostly to Penner serogroup 19 and contain beta-N-acetylglucosamine residues. *Ann Neurol* **33**:243-7.
22. **Lawley, T. D., W. A. Klimke, M. J. Gubbins, and L. S. Frost.** 2003. F factor conjugation is a true type IV secretion system. *FEMS Microbiol Lett* **224**:1-15.
23. **Leonard, E. E., 2nd, T. Takata, M. J. Blaser, S. Falkow, L. S. Tompkins, and E. C. Gaynor.** 2003. Use of an open-reading frame-specific *Campylobacter jejuni* DNA microarray as a new genotyping tool for studying epidemiologically related isolates. *J Infect Dis* **187**:691-4.
24. **Leonard, E. E., 2nd, L. S. Tompkins, S. Falkow, and I. Nachamkin.** 2004. Comparison of *Campylobacter jejuni* isolates implicated in Guillain-Barre syndrome and strains that cause enteritis by a DNA microarray. *Infect Immun* **72**:1199-203.
25. **Loosmore, S. M., J. M. Shortreed, D. C. Coleman, D. M. England, and M. H. Klein.** 1996. Sequences of the genes encoding the A, B and C subunits of the *Haemophilus influenzae* dimethylsulfoxide reductase complex. *Gene* **169**:137-8.
26. **McEwan, A. G., S. J. Ferguson, and J. B. Jackson.** 1991. Purification and properties of dimethyl sulphoxide reductase from *Rhodobacter capsulatus*. A periplasmic molybdoenzyme. *Biochem J* **274 (Pt 1)**:305-7.
27. **Murray, N. E.** 2000. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol Rev* **64**:412-34.

28. **Newell, D. G., H. McBride, F. Saunders, Y. Dehele, and A. D. Pearson.** 1985. The virulence of clinical and environmental isolates of *Campylobacter jejuni*. J Hyg (Lond) **94**:45-54.
29. **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**:665-8.
30. **Pearson, B. M., C. Pin, J. Wright, K. I'Anson, T. Humphrey, and J. M. Wells.** 2003. Comparative genome analysis of *Campylobacter jejuni* using whole genome DNA microarrays. FEBS Lett **554**:224-30.
31. **Penner, J. L., and J. N. Hennessy.** 1980. Passive hemagglutination technique for serotyping *Campylobacter fetus* subsp. *jejuni* on the basis of soluble heat-stable antigens. J Clin Microbiol **12**:732-7.
32. **Poly, F., D. Threadgill, and A. Stintzi.** 2004. Identification of *Campylobacter jejuni* ATCC 43431-specific genes by whole microbial genome comparisons. J Bacteriol **186**:4781-95.
33. **Russell, R. G., M. J. Blaser, J. I. Sarmiento, and J. Fox.** 1989. Experimental *Campylobacter jejuni* infection in *Macaca nemestrina*. Infect Immun **57**:1438-44.
34. **Schroder, G., S. Krause, E. L. Zechner, B. Traxler, H. J. Yeo, R. Lurz, G. Waksman, and E. Lanka.** 2002. TraG-like proteins of DNA transfer systems and

of the *Helicobacter pylori* type IV secretion system: inner membrane gate for exported substrates? J Bacteriol **184**:2767-79.

35. **Sellars, M. J., S. J. Hall, and D. J. Kelly.** 2002. Growth of *Campylobacter jejuni* supported by respiration of fumarate, nitrate, nitrite, trimethylamine-N-oxide, or dimethyl sulfoxide requires oxygen. J Bacteriol **184**:4187-96.
36. **Skirrow, M. B., and M. J. Blaser.** 2000. Clinical aspects of *Campylobacter* infection, p. 69-88. In I. Nachamkin and M. J. Blaser (ed.), *Campylobacter*. 2nd edition. American Society for Microbiology, Washington.
37. **Wassenaar, T. M., and D. G. Newell.** 2000. Genotyping of *Campylobacter* spp. Appl Environ Microbiol **66**:1-9.
38. **Weiner, J. H., D. P. MacIsaac, R. E. Bishop, and P. T. Bilous.** 1988. Purification and properties of *Escherichia coli* dimethyl sulfoxide reductase, an iron-sulfur molybdoenzyme with broad substrate specificity. J Bacteriol **170**:1505-10.

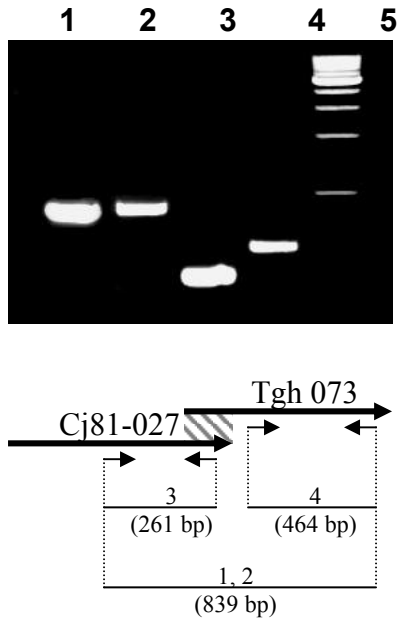


Figure 1: PCR amplification and schematic outline of the chromosomal TraG-like region identified in *C. jejuni* ATCC 43431 and *C. jejuni* 81-176. The ORFs are denoted by boxed arrows. The slashed box indicates the region with 100% identity between both ORFs. Predicted PCR fragments are shown at the bottom. Gel lanes match the PCR product label. Lane 5 shows the 1 kb DNA ladder. PCR products 1 and 2 were obtained by using as template the chromosomal DNA from *C. jejuni* ATCC 43431 and *C. jejuni* 81-176, respectively. PCR products 3 and 4 were obtained by using the chromosomal DNA of *C. jejuni* ATCC 43431 as template.

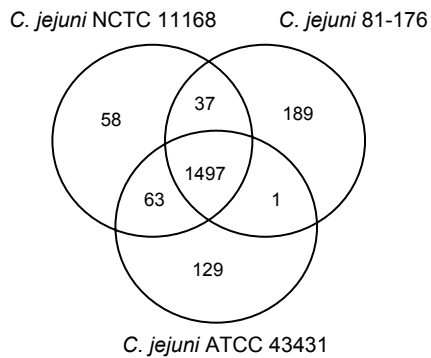


Figure 2: comparative gene content of strains *C. jejuni* NCTC 11168, ATCC 43431 and 81-176.

Size range (kb)	No. of contigs	Total DNA (kb)
Chromosome		
4 to 5	1	4.4
3 to 4	2	6.3
2 to 3	1	2.3
1 to 2	22	29.7
0.5 to 1	21	15.5
<0.5	10	4.1
Total	57	62.3
pTet		
>5	2	12
4 to 5	0	0
3 to 4	1	3.4
2 to 3	0	0
1 to 2	11	16
<1	2	1.2
Total	16	32.6
pVir		
>5	2	12.5
4 to 5	2	8.5
3 to 4	1	3.7
2 to 3	2	5.2
1 to 2	2	2.8
<1	4	1.8
Total	13	34.5
Total	86	130

Table 1: Size distribution of the assembled contigs for the unique regions from the chromosome, the pTet and the pVir plasmids.

ORF	Length ^a (bp)	%G+C content	Closest relationship ^b	Identity (%) ^c	Accession number
<u>Capsule</u>					
Cj81-045	1020 ^{3'}	26.6	Putative glycosyltransferase <i>Campylobacter jejuni</i> ATCC 43456 [AAR01891]	334/340 (98%)	AY681253
Cj81-046	319 ^{3'}	30.7	Putative dTDP-4-dehydro rhamnose 3,5-epimerase <i>Campylobacter jejuni</i> <i>Campylobacter jejuni</i> ATCC 43456 [AAR01889]	106/106 (100%)	AY681254
Cj81-047	564 ^{5'}	19.1	Putative glycosyltransferase <i>Campylobacter jejuni</i> ATCC 43456 [AAR01890]	139/177 (78%)	AY681254
Cj81-062	594 ^{3'}	24.7	Putative glycosyltransferase, <i>Campylobacter jejuni</i> ATCC 43456 [AAR01890]	198/198 (100%)	AY681260
Cj81-063	170 ^{5'}	30.4	Putative glycosyltransferase, <i>Campylobacter jejuni</i> ATCC 43456 [AAR01891]	44/56 (78%)	AY681260
Cj81-079	1058 ^{3'}	25.6	Putative glycosyltransferase, <i>Campylobacter jejuni</i> CCUG 10954 [AAR01916]	303/333 (90%)	AY681267
Cj81-080	1011 ^{5'}	21.9	Putative glycosyltransferase, <i>Campylobacter jejuni</i> ATCC 43456 [AAR01894]	547/603 (90%)	AY681267
Cj81-081	1613	21.9	Putative glycosyltransferase, <i>Campylobacter jejuni</i> ATCC 43456 [AAR01896]	535/536 (99%)	AY681267
Cj81-102	580 ^{5'3'}	21.6	Putative glycosyltransferase, <i>Campylobacter jejuni</i> ATCC 43456 [AAR01892]	176/192 (91%)	AY681278
Cj81-103	468	22.0	Putative glycosyltransferase, <i>Campylobacter jejuni</i> ATCC 43456 [AAR01881]	154/155 (99%)	AY681279
Cj81-128	531 ^{5'}	34.0	Putative GDP-mannoheptose-4,6 dehydratase <i>Campylobacter</i> <i>jejuni</i> ATCC 43456 [AAR01886]	142/143 (99%)	AY681295
Cj81-129	458 ^{5'}	25.6	Unknown <i>Campylobacter jejuni</i> ATCC 43456 [AAR01888]	151/151 (100%)	AY681296
Cj81-130	384 ^{3'}	31.2	Putative GDP-fucose synthase <i>Campylobacter jejuni</i> ATCC 43456 [AAR01887]	113/127 (88%)	AY681296

ORF	Length ^a (bp)	%G+C content	Closest relationship ^b	Identity (%) ^c	Accession number
<u>Lipooligosaccharide, cell envelop and surface structures</u>					
Cj81-003	194 ^{3'}	26.3	Putative acetyltransferase <i>Campylobacter jejuni</i> ATCC 43449 [AAL06009]	47/47 (100%)	AY681239
Cj81-008	603 ^{5'}	22.2	Beta-1,4-N- acetylgalactosaminyltransferase <i>Campylobacter jejuni</i> OH4384 [AAF31769]	200/200 (100%)	AY681241
Cj81-009	280 ^{5'}	20.3	Beta-1,3-galactosyltransferase <i>Campylobacter jejuni</i> OH4384 [AAF31770]	59/92 (64%)	AY681241
Cj81-010	360 ^{3'}	27.1	Beta-1,3-galactosyltransferase <i>Campylobacter jejuni</i> OH4384 [AAF31770]	120/120 (100%)	AY681242
Cj81-011	853 ^{3'}	25.8	Alpha-2,3-sialyltransferase <i>Campylobacter jejuni</i> OH4384 [AAF34137]	269/284 (94%)	AY681242
Cj81-030	182 ^{5'}	25.7	Putative amino acid activating enzyme Cj1307 <i>Campylobacter</i> <i>jejuni</i> NCTC 11168 [NP_282453]	48/60 (80%)	AY681244
Cj81-042	646 ^{5'3'}	24.5	Putative lipopolysaccharide A protein <i>Vibrio parahaemolyticus</i> RIMD 2210633 [NP_796570]	82/200 (41%)	AY681251
Cj81-043	839 ^{5'}	24.3	Putative glycosyltransferase <i>Campylobacter jejuni</i> OH4384 [AAF31768]	243/277 (87%)	AY681252
Cj81-044	167 ^{3'}	19.8	Beta-1,4-N- acetylgalactosaminyltransferase <i>Campylobacter jejuni</i> OH4384 [AAF31769]	46/50 (92%)	AY681252
Cj81-058	778 ^{3'}	31.7	Beta-1,4-N- acetylgalactosaminyltransferase <i>Campylobacter jejuni</i> ATCC 43456 [AAL05993]	259/259 (100%)	AY681257
Cj81-059	653 ^{5'}	26.4	NeuC1 <i>Campylobacter jejuni</i> 81-176 [AAL09370]	215/217 (99%)	AY681257
Cj81-085	1442	24.7	Penicillin-binding protein 1A <i>Oceanobacillus iheyensis</i> HTE831 [NP_692087]	53/137 (38%)	AY681270
Cj81-104	429 ^{5'3'}	34.0	Probable periplasmic protein Cj0737 <i>Campylobacter jejuni</i> NCTC 11168 [NP_281909]	84/131 (64%)	AY681280
Cj81-106	558	34.2	Sialic acid synthase <i>Campylobacter jejuni</i> ATCC 43449 [AAL06005]	185/185 (100%)	AY681282

ORF	Length ^a (bp)	%G+C content	Closest relationship ^b	Identity (%) ^c	Note
Cj81-119	853	33.0	FlgE <i>Campylobacter jejuni</i> 81-176 [AAP34261]	284/284 (100%)	AY681288
Cj81-121	644	28.2	Probable outer membrane protein Cj1721c <i>Campylobacter jejuni</i> NCTC 11168 [NP_282847]	140/214 (65%)	AY681290
Cj81-127	540 ^{3'}	28.6	a) Unknown <i>Campylobacter jejuni</i> 81-176 [AAR82939] b) Putative glycosyltransferase WavM <i>Vibrio cholerae</i> [AAL77351]	180/180 (100%) 67/187 (35%)	AY681294
<u>Restriction/modification, recombination and repair (DNA modification)</u>					
Cj81-032	299 ^{3'}	31.4	Type I site-specific restriction-modification system, <i>Trichodesmium erythraeum</i> IMS101 [ZP_00071585]	31/86 (36%)	AY681246
Cj81-057	1268	27.7	Type I restriction-modification system S subunit <i>Synechocystis</i> sp.PCC 6803 [BAD02014]	138/428 (32%)	AY681256
Cj81-060	690 ^{5'3'}	34.0	a) Type I site-specific restriction-modification system, <i>Haemophilus influenzae</i> [ZP_00321696] b) Type I restriction enzyme HsdR, putative <i>Vibrio cholerae</i> O1 biovareltor str. N16961 [NP_231400]	125/241 (51%) 122/244 (50%)	AY681258
Cj81-061	337 ^{5'3'}	30.6	Putative type I restriction-modification system (HsdR), <i>Acinetobacter</i> sp. ADP1 [YP_047908]	70/113 (61%)	AY681259
Cj81-087	419 ^{5'}	26.4	a) Probable type IIS restriction/modification enzyme, C-terminal half Cj0032 <i>Campylobacter jejuni</i> NCTC 11168 [NP_281254] b) Type I restriction enzyme HsdR, putative <i>Vibrio cholerae</i> O1 biovareltor str. N16961 [NP_231400]	89/141 (63%) 55/122 (45%)	AY681270
Cj81-118	941 ^{3'}	32.7	Type I restriction-modification system methyltransferase subunit <i>Trichodesmium erythraeum</i> IMS101 [ZP_00071587]	183/318 (57%)	AY681287

ORF	Length ^a (bp)	%G+C content	Closest relationship ^b	Identity (%) ^c	Note
Cj81-126	448 ^{5'3'}	30.3	a) Type I restriction-modification system specificity subunit, <i>Methanosarcina mazei</i> Go1 [NP_635005] b) Type I restriction-modification system methyltransferase subunit <i>Trichodesmium erythraeum</i> IMS101 [ZP_00071587]	69/128 (53%) 183/318 (57%)	AY681293
<u>Small molecule metabolism</u>					
Cj81-005	859 ^{3'}	34.8	Cytochrome c family protein, <i>Geobacter sulfurreducens</i> PCA [NP_951335]	128/281 (45%)	AY681240
Cj81-006	584	25.3	a) Conserved hypothetical protein <i>Wolinella succinogenes</i> DSMZ 1740 [NP_908252] b) Thiol-disulfide interchange protein HP0377 <i>Helicobacter pylori</i> 26695 [NP_207175]	66/184 (35%) 54/180 (30%)	AY681240
Cj81-007	492 ^{5'}	30.6	Cytochrome c biogenesis protein (YCF5) <i>Wolinella succinogenes</i> DSMZ 1740 [NP_908253]	70/152 (46%)	AY681240
Cj81-038	206 ^{5'}	25.1	Glutamate synthase (NADPH) large chain Cj0007 precursor, <i>Campylobacter jejuni</i> NCTC 11168 [NP_281229]	68/68 (100%)	AY681247
Cj81-041	700 ^{5'3'}	26.0	a) Hypothetical protein Cj1300 <i>Campylobacter jejuni</i> NCTC 11168 [NP_282446] b) 3-demethylubiquinone-9 3-methyltransferase <i>Bacteroides thetaiotaomicron</i> VPI-5482 [NP_811691]	126/229 (55%) 41/124 (33%)	AY681250
Cj81-074	513 ^{3'}	34.8	Predicted acyl esterases <i>Rhodobacter sphaeroides</i> [ZP_00007094]	92/151 (60%)	AY681266
Cj81-083	641 ^{5'3'}	30.9	A) Hypothetical protein WS1432 <i>Wolinella succinogenes</i> DSMZ 1740 [NP_907593] b) Anaerobic dimethyl sulfoxide reductase chain C <i>Haemophilus influenzae</i> Rd KW20 [NP_439204]	87/211 (41%) 46/185 (24%)	AY681268
Cj81-088	1167 ^{5'3'}	28.7	Cytochrome c biogenesis protein YCF5 <i>Wolinella succinogenes</i> DSMZ 1740 [NP_908253]	196/393 (49%)	AY681271

ORF	Length ^a (bp)	%G+C content	Closest relationship ^b	Identity (%) ^c	Accession number
Cj81-095	1053 ^{5'}	30.4	8-amino-7-oxononanoate synthase BioF, <i>Campylobacter jejuni</i> NCTC 11168 [CAB72773]	243/350 (69%)	AY681274
Cj81-097	579 ^{3'}	32.6	Cytochrome c family protein <i>Geobacter sulfurreducens</i> PCA [NP_951335]	60/204 (29%)	AY681275
Cj81-098	443	30.2	a) Cytochrome b subunit of formate dehydrogenase, <i>Dechloromonas aromatica</i> [ZP_00151826] b) Nitrate/TMAO reductases, membrane-bound tetraheme cytochrome c subunit, <i>Desulfitobacterium hafniense</i> taxon:49338 [ZP_00099300]	59/189 (31%) 48/133 (36%)	AY681275
Cj81-099	1141 ^{5'3'}	35.7	Cytochrome c, putative <i>Shewanella oneidensis</i> MR-1 [NP_716115]	235/376 (62%)	AY681276
Cj81-105	1627 ^{5'3'}	35.1	Anaerobic dimethyl sulfoxide reductase chain A (DmsA), <i>Wolinella succinogenes</i> DSM 1740 [NP_907591]	325/513 (63%)	AY681281
Cj81-109	893 ^{3'}	25.1	Acyl dehydratase <i>Azotobacter vinelandii</i> [ZP_00089607]	18/41 (43%)	AY681284
Cj81-110	227	23.2	Probable acyl carrier protein Cj1299 <i>Campylobacter jejuni</i> NCTC 11168 [NP_282445]	44/75 (58%)	AY681284
Cj81-111	758 ^{5'}	28.2	a) Hypothetical protein Cj1298 <i>Campylobacter jejuni</i> NCTC 11168 [NP_282444] b) Aminoglycoside N3'-acetyltransferase <i>Clostridium acetobutylicum</i> [NP_348814]	110/242 (45%) 85/250 (34%)	AY681284
Cj81-124	221 ^{5'}	30.2	Putative cytochrome C type biogenesis protein <i>Helicobacter pylori</i> J99 [NP_223720]	44/69 (63%)	AY681291
<u>Transport/binding proteins/Protein and peptide secretion</u>					
Cj81-027	822 ^{3'}	33.9	Hypothetical protein Tgh073, <i>Campylobacter jejuni</i> TGH 9011 [AAS99026]	26/26 (100%)	AY681243
Cj81-048	272	31.9	a) Conserved hypothetical protein <i>Clostridium perfringens</i> str. 13 [NP_561338] b) Putative transporter CgxT <i>Escherichia coli</i> taxon:562 [AAL61900]	38/72 (52%) 26/72 (36%)	AY681255

ORF	Length ^a (bp)	%G+C content	Closest relationship ^b	Identity (%) ^c	Accession number
Cj81-049	453	34.2	Preprotein translocase chain Cj1688c <i>Campylobacter jejuni</i> NCTC 11168 [NP_282814]	116/117 (99%)	AY681255
Cj81-073	297 ^{5'}	29.6	a) Permeases of the major facilitator superfamily, <i>Burkholderia cepacia</i> R18194 [ZP_00217922] b) Putative transporter CgxT <i>Escherichia coli</i> taxon:562 [AAL61900]	31/83 (37%) 23/77 (29%)	AY681266
Cj81-094	1103 ^{3'}	29.7	Similar to di-tripeptide ABC transporter PLU1813 <i>Photorhabdus luminescens</i> subsp. <i>Laumondii</i> [NP_929091]	148/366 (40%)	AY681273
Cj81-107	404 ^{5'}	29.3	Probable efflux protein Cj1687 <i>Campylobacter jejuni</i> NCTC 11168 [NP_282813]	88/129 (68%)	AY681283
Cj81-125	351 ^{5'3'}	32.2	Probable molybdenum transport ATP-binding protein Cj0300c <i>Campylobacter jejuni</i> NCTC 11168 [NP_281491]	77/117 (65%)	AY681292
<u>Hypothetical and Other</u>					
Cj81-026	452	32.9	a) Hypothetical protein Bucepa02004635, <i>Burkholderia</i> <i>cepacia</i> R1808 [ZP_00220764] b) Mating pair stabilization protein <i>Novosphingobium</i> <i>aromaticivorans</i> [NP_049155]	60/137 (43%) 60/140 (42%)	AY681243
Cj81-033	989	21.7	Hypothetical protein <i>Leptospira</i> <i>interrogans</i> serovar lai str.56601 [NP_711975]	117/309 (37%)	AY681246
Cj81-034	326 ^{5'}	24.7	Hypothetical protein STY3192 <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Typhi</i> CT18 [NP_457434]	42/58 (72%)	AY681246
Cj81-037	1074 ^{3'}	24.3	Hypothetical protein jhp0174 <i>Helicobacter pylori</i> strain J99 [NP_222895]	131/366 (35%)	AY681247
Cj81-068	847	23.5	Hypothetical protein <i>Bacillus</i> <i>cereus</i> ATCC 14579 [NP_831045]	52/172 (30%)	AY681264
Cj81-069	1319	24.5	ATP/GTP-binding protein <i>Bacillus cereus</i> ATCC 14579 [NP_831044]	153/456 (33%)	AY681264
Cj81-070	496 ^{5'3'}	24.2	Conserved hypothetical protein <i>Porphyromonas gingivalis</i> W83 [NP_905721]	60/143 (41%)	AY681265

ORF	Length ^a (bp)	%G+C content	Closest relationship ^b	Identity (%) ^c	Accession number
Cj81-070	496 ^{5'3'}	24.2	Conserved hypothetical protein <i>Porphyromonas gingivalis</i> W83 [NP_905721]	60/143 (41%)	AY681265
Cj81-086	894	23.8	Hypothetical protein <i>Enterococcus faecalis</i> V583 [NP_815563]	43/140 (30%)	AY681270
Cj81-092	420 ^{3'}	29.3	Hypothetical protein Cj0741 <i>Campylobacter jejuni</i> NCTC 11168 [NP_281913]	48/49 (97%)	AY681272
Cj81-096	611	27.3	Hypothetical protein Cj0305c <i>Campylobacter jejuni</i> NCTC 11168 [AL139074]	128/203 (63%)	AY681274
<u>Unknown</u>					
Cj81-004	123 ^{3'}	29.3	No hit	-	AY681239
Cj81-028	150 ^{3'}	30.7	No hit	-	AY681244
Cj81-029	725	24.2	No hit	-	AY681244
Cj81-040	1147 ^{5'}	29.4	No hit	-	AY681249
Cj81-064	298 ^{5'}	27.9	No hit	-	AY681261
Cj81-065	408 ^{3'}	31.0	No hit	-	AY681261
Cj81-084	275 ^{5'}	31.5	No hit	-	AY681269
Cj81-093	188 ^{5'}	32.8	No hit	-	AY681272
Cj81-100	172 ^{3'}	38.9	No hit	-	AY681277
Cj81-101	230 ^{5'}	37.7	No hit	-	AY681277
Cj81-108	446 ^{5'}	32.9	No hit	-	AY681283
Cj81-115	319 ^{5'}	38.0	No hit	-	AY681285
Cj81-116	884 ^{3'}	34.1	No hit	-	AY681285
Cj81-117	767 ^{5'}	26.8	No hit	-	AY681286

Table 2: Chromosomal DNA fragments present in *C. jejuni* 81-176 but not in *C. jejuni* NCTC 11168.

^aThe superscript number denotes a partial ORF with the 3' and/or 5' region missing.

^bDatabase hits, a) first hit; b) other relevant hit.

^cAmino-acid identity in the region of homology identify by BLASTP.

3 Résultats additionnels

Introduction

Au cours de l'analyse d'hybridation sur la puce à ADN de la souche *C. jejuni* 81-176, nous avons été en mesure d'isoler les séquences chromosomiques et plasmidiques uniques à la souche 81-176 (par hybridation compétitive entre les ADN des souches *C. jejuni* NCTC 11168 et 81-176). La souche 81-176 contient deux larges plasmides pVir et pTet. La séquence du plasmide pVir, présentant une taille de 37,5 kb, a été publiée en 2002 (17). Au moment de nos propres expériences, la séquence du plasmide pTet n'avait pas été encore publiée. La taille de ce plasmide a été estimée être approximativement de la même taille que pVir. A la suite de notre expérience d'hybridation nous avons pu isoler 15 contigs représentant 31,4 kb de séquences correspondantes à pTet. Nous avons décidé de compléter la séquence de ce plasmide par la création d'amorces PCR complémentaires aux deux extrémités des contigs trouvés. Ces amorces nous ont permis de séquencer les produits PCR correspondants et de compléter ainsi la séquence de ce plasmide.

Résultats

Il est à noter que ces résultats n'ont pas été publiés dans le papier « 3-1 Diversity in *Campylobacter jejuni*: Identification of *C. jejuni* 81-176 Specific Genes » car malheureusement nous avons été devancés par Batchelor et coll.(19)

Après séquençage et annotation, nous avons obtenu pour ce plasmide une taille totale de 45 212 pb codant pour 45 ORFs nommés *pTet01* à *pTet45*. L'attribution du nom pTet01 a été faite de façon aléatoire. Nous avons identifié 68,78% de la séquence totale de pTet par utilisation de notre puce à ADN. La couverture de ce plasmide est faible comparée à la couverture du plasmide pVir (92%). Ce faible pourcentage peut être expliqué par le fait que l'ADN ayant servi à la construction de la puce de 81-176 a été extrait d'une culture ayant poussé en absence de tétracycline. Il a été montré que sans cet antibiotique pTet était instable et spontanément perdu (16). Cette raison pourrait donc expliquer la faible couverture de la séquence pTet.

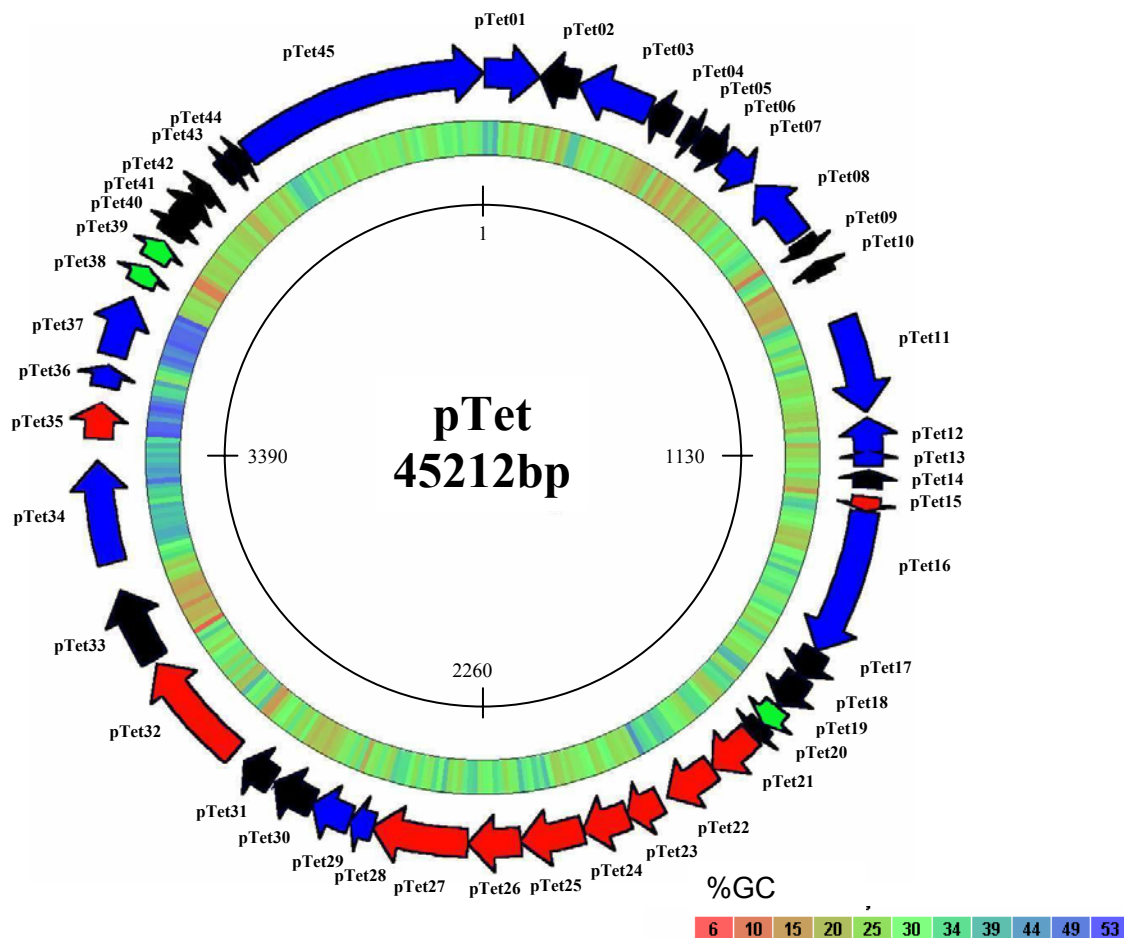


Figure 3 : Représentation schématique du plasmide pTet de la souche *C. jejuni* 81-176. Le premier cercle au centre représente la position sur le plasmide en paires de bases. Le second cercle représente le pourcentage G-C. Chaque trait correspond à 113 pb. Chaque ORF identifiée est représentée par une flèche de couleur correspondant à une fonction : Bleu, ORFs codant des protéines ayant des similarités avec des protéines bactériennes n'ayant pas de fonctions attribuées ; Noire, ORFs codant des protéines de fonctions inconnues ; Rouge, ORFs codant des protéines ayant des similarités avec des protéines impliquées dans le type IV de sécrétion ; Vert, ORFs codant des protéines présentant des similarités avec des protéines de pVir.

pTet, données physiques (Figure 3)

Le pourcentage G+C du plasmide pTet est de 29,14%, ce qui est supérieur au pourcentage GC de pVir (26%) (17), mais proche du pourcentage GC des séquences uniques à 81-176 (28,15%) et du chromosome de NCTC 11168 (30,6%). Il est à noter une région présentant un haut pourcentage G+C (*pTet34* à *pTet37*) qui atteint presque 50% pour le gène *pTet35*. Les gènes compris dans cette région présentent des homologies avec la protéine impliquée dans la résistance à la tétracycline (*pTet34*), avec VirD2

(pTet35), une hypothétique protéine de *T. denticola* (pTet36), et une protéine de réplication Rep (pTet37).

Comme pVir, pTet présente de larges zones non codantes (entre *pTet10* et *pTet11*, 910 pb ; entre *pTet33* et *pTet34*, 423 pb ; entre *pTet34* et *pTet35*, 423 pb et entre *pTet35* et *pTet36*, 422 pb). La zone comprise entre *pTet10* et *pTet11* présente des répétitions directes (DR) de 15 pb (TTACATTAAAGTCAA). Il a été suggéré que les origines de réplication de petits plasmides cryptiques de *C. jejuni* sp. présentent des séquences répétées (DR), précédées par une région riche en nucléotides A-T (103). Après analyse, il s'avère que la région suivant pTet10 est précédée d'une région riche en AT. Cette zone pourrait donc représenter l'origine de réplication de pTet.

De nombreuses DR ou des séquences inverses répétées (IR) sont présentes sur la séquence de pTet. Ces séquences pourraient être de potentiels vestiges d'éléments transposables. Sur les 45 ORFs caractérisées, il a été possible de caractériser pour 15 d'entre elles un possible site de fixation ribosomal (RBS, séquence consensus : AAGGA-N7-, avec N variant de 5 à 9).

Annotation.

Nous avons recherché les ORFs potentielles de plus de 80 acides aminés (aa) via le programme Artemis. Ces ORFs sont référencées dans le tableau 3.

Dix huit protéines ne présentent aucune homologie significative avec la banque protéique totale. Trois ORFs de pTet codent pour des protéines présentant des homologies avec des protéines encodées par le plasmide pVir (*pTet19* avec *Cjp13*, *pTet 38* avec *Cjp20* et *pTet39* avec *Cjp35*). Aucune des ORFs identifiées ne présente d'homologie avec les ORFs de NCTC 11168.

Un total de 10 protéines exprimées par pTet présentent des homologies avec des protéines impliquées dans le type IV de sécrétion. La protéine codée par *pTet21* présente 21% d'identité avec VirB5, une sous unité périplasmique du pilus, pTet27 présente 45% d'identité avec VirD4, une protéine impliquée dans le recrutement d'ADN ou une protéine à sécréter secrétée, de sécrétion ????(31), la protéine codée par *pTet25* présente 42% d'identité avec VirB10, une protéine faisant partie du pore créant un pont entre les membranes externe et interne (31). La protéine pTet26 présente 99% d'identité

avec VirB11, une ATPase nécessaire à la translocation du substrat (31). Les protéines codées par *pTet22*, *23*, *24* présentent respectivement une identité avec VirB6 (22%), VirB8 (86%), et VirB9 (97%) et la protéine codée par *pTet35* présente 31% d'identité avec VirD2, protéine qui attache l'ADN simple brin au cours de son transfert (31). Deux autres ORFs, *pTet15* et *pTet32* codent pour des protéines présentant des homologies avec respectivement TraC (35%) et TraE (37%). La protéine TraC appartient à la famille VirB4 qui représente des ATPases nécessaires à la translocation du substrat (même rôle que VirB11) (31). TraE est une topoisomérase III (31). L'ORF pTet 35, non trouvée dans l'analyse de Batchelor *et al.*, code pour une protéine présentant des homologies avec VirD2, protéine clé impliquée dans le transfert et l'intégration du plasmide T de *A. tumefaciens* dans le génome de la cellule hôte. L'ORF *cmgB7* (selon l'annotation de Batchelor et coll. (19)) du fait de sa petite taille, n'a pas pu être identifié dans notre annotation. Ce gène code pour une protéine ne présentant aucune homologie significative avec une protéine bactérienne. Mais tout comme VirB7 d'*A. tumefaciens*, CmgB7 présente un signal lipoprotéique et un site conservé d'attachement à un lipide. Des protéines similaires à VirB5, B8, B10, B11 et VirD4 sont localisées de façon contigue sur pTet, zone présentant une similarité avec le plasmide pCjA13 (souche *C. jejuni* CjA13). Les protéines VirB8, B9, B10, B11 et D4 sont présentes sur les deux plasmides pTet et pVir.

Selon Batchelor et coll. (19), la plupart des gènes impliqués dans ce type IV de sécrétion présentent de fortes homologies avec des protéines du plasmide pVT745 d'*Actinobacillus actinomycetemcomitans*, impliquées dans le processus de conjugaison. Pour cette raison Batchelor *et al* suggèrent l'implication de ces gènes dans le transfert plasmidique plutôt que dans un mécanisme de pathogénèse. Ce rôle serait conféré par la présence en partie du plasmide pVir chez la souche 81-176.

L'analyse de mutants d'ORF de pVir démontre des résultats hétérogènes concernant la capacité d'invasion *in vivo*. Les mutants pVir correspondant aux protéines VirB8 et SSB n'ont aucun effet sur le pouvoir de colonisation. A la lumière des ORFs détectées sur pTet, ces résultats peuvent être potentiellement expliqués par la présence de fonctions similaires codées par pTet. Les mutants VirB9, B10 et B11 de pVir démontrent la

réduction maximum d'invasion in vivo, malgré la présence de fonctions similaires sur pTet (16). Dans ces cas, une complémentation en trans n'est donc pas observée.

Comparaison de notre séquence du plasmide pTet avec celle publiée par Batchelor *et al.* (19)

Le plasmide pTet séquencé par Batchelor et coll. (19) a été déterminé comme étant d'une taille de 45205pb, soit 7 paires de bases plus petite que notre propre séquence. Après comparaison de ces 2 séquences il a pu être déterminé que ces plasmides présentent en tout et pour tout 24 différences comprenant: 17 nucléotides manquants dans l'une des 2 séquences (12 nucléotides de notre séquence dans l'analyse de Batchelor et coll. (19) et 5 nucléotides de la séquence de Batchelor et coll. non présents dans notre séquence), 5 transversions et deux inversions de nucléotides. Ces différences peuvent être le résultat de mutations ayant eu cours de façon indépendante au niveau des souches utilisées pour le séquençage, ou plus probablement à des erreurs de séquençage.

Après annotation nous avons déterminé 45 ORFs (pTet01 à pTet45). L'annotation de Batchelor et coll quant à elle, a permis d'identifier un total de 49 ORFs. Cette différence provient du fait que nous avons annoté seulement les ORFs supérieures à 240 nucléotides, alors que Batchelor et coll (19) ont annoté toutes les ORFs supérieures à 150 nt . Ainsi, dans notre annotation, nous omettons l' ORF *cyp2* (57 aa) codant pour une protéine homologue à une protéine codée par un transposon conjugatif, l'ORF *cyp6* (58 aa) homologue de la protéine Cjp19 de fonction inconnue du plasmide pVir de 81-176, celles de *cyp24* (72 aa) et *cyp25* (67 aa) codant pour des protéines ayant des homologies avec des protéines putatives de *H. pylori*, *cmgB7* (55aa) codant une protéine pouvant avoir des homologies avec VirB7, et *cyp51* (59aa) homologue de la protéine Cjp20 de fonction inconnue du plasmide pVir de 81-176 .

Deux ORFs que nous avons identifiées sur notre séquence ne sont pas répertoriées dans l'analyse de Batchelor et coll (19). Ces ORFs sont *pTet10* et *pTet35*. Les protéines codées par ces gènes n'ont aucun homologue protéique permettant de leur attribuer une fonction putative. Les ORFs *pTet01* et *pTet45* codent respectivement pour une protéine homologue hypothétique de *H. pylori* et une méthylase d'ADN de *B. cepacia*. Elles sont déduites comme étant une seule ORFs (*cyp14*) dans l'analyse de Batchelor et coll (19)

Certaines ORFs diffèrent légèrement de celles de Batchelor coll. (19) Ces variations d'annotation sont dues aux différences nucléotidiques rencontrées sur ces séquences ainsi qu'aux algorithmes utilisés au cours de l'annotation.

Gène	Nom putatif	Brin	Start-Stop	Identité	Relation significative
pTet01		+	29-1063	114/372 (30%)	Hypotétique protéine; OrfQ, <i>Helicobacter pylori</i> [AAM03043]
pTet02		-	1085-1789	N/A	Inconnue
pTet03	nickase	-	1815-3203	108/303 (35%)	Nickase, <i>Actinobacillus actinomycetemcomitans</i> [NP_067550]
pTet04		-	3203-3754	N/A	Inconnue
pTet05		+	3955-4236	N/A	Inconnue
pTet06		+	4265-4876	N/A	Inconnue
pTet07		+	4880-5533	55/112 (49%)	Hypotétique protéine HH0278, <i>Helicobacter hepaticus</i> , [NP_859809]
pTet08	SogL	+	5652-6878	118/389 (30%)	ADN primase SogL, <i>Escherichia coli</i> , [AAQ17618]
pTet09		+	6986-7249	N/A	Inconnue
pTet10		-	7471-7728	N/A	Inconnue
pTet11	RadA	+	8638-10431	46/170 (27%)	Protéine de réparation de l'ADN RadA, <i>Caulobacter crescentus</i> , [NP_420470]
pTet12		-	10554-11168	67/180 (37%)	Putative Site-spécifique recombinase <i>W. succinogenes</i> , [NP_908035]
pTet13		-	11165-11440	N/A	Inconnue
pTet14		-	11533-11856	N/A	Inconnue
pTet15	TraC	+	12027-12290	29/82 (35%)	TraC protéine, Plasmide pIPO2T, [NP_444519]
pTet16		+	12303-15071	366/921 (39%)	ATPase, <i>Actinobacillus actinomycetemcomitans</i> , [NP_067581]
pTet17		+	15082-15648	82/183 (44%)	Inconnue <i>Campylobacter jejuni</i> pCjA13, [AAO65161]
pTet18		+	15645-16310	N/A	Inconnue

Gène	Nom putatif	Brin	Start-Stop	Identité	Relation significative
pTet19	Ssb	+	16344-16769	56/128 (43%)	Single-stranded binding protéine Cjp13, <i>Campylobacter jejuni</i> , [NP_863310]
pTet20		+	16785-17060	N/A	Inconnue
pTet21	VirB5	+	17064-18035	67/214 (31%)	Type IV de secretion. Composant VirB5, <i>Yersinia pestis biovar Mediaevalis</i> , [NP_995424]
pTet22	VirB6	+	18032-19030	33/146 (22%)	Type IV de secretion. Composant VirB6, <i>Yersinia pestis biovar Mediaevalis</i> , [NP_995426]
pTet23	VirB8	+	19177-19839	190/220 (86%)	Type IV de secretion. Composant VirB8, <i>Campylobacter jejuni</i> , [AAO65157]
pTet24	VirB9	+	19836-20723	287/295 (97%)	Type IV de secretion. Composant VirB9, <i>Campylobacter jejuni</i> , [AAO65158]
pTet25	VirB10	+	20729-21925	142/142 (100%) 88/207 (42%)	Inconnue <i>Campylobacter jejuni</i> pCjA13, [AAO65159] Type IV de secretion. Composant VirB10, <i>Brucella melitensis</i> , [AAD56620]
pTet26	VirB11	+	21906-22898	328/329 (99%)	Type IV de secretion. Composant VirB11, <i>Campylobacter jejuni</i> , [AAO65160]
pTet27	VirD4/TraN	+	22899-24710	240/531 (45%) 211/490 (43%)	TraN protéine, Plasmid pIPO2T, [NP_444530] Type IV de secretion. Composant VirD4
pTet28		+	24691-25128	35/113 (30%)	Protéine Cag ilot de pathogénécité <i>Helicobacter pylori</i> ,
pTet29	YggA	+	25140-25904	65/189 (34%)	Protéine d'exportation de la Lysine
pTet30		+	25907-26704	N/A	Inconnue
pTet31		+	26775-27395	N/A	Inconnue
pTet32	TraE	+	27659-29851	280/740 (37%)	ADN topoisomerase III
pTet33		+	29957-31378	N/A	Inconnue
pTet34	TetO	+	31936-33855	627/639 (98%)	Gène de resistance à la Tetracycline
pTet35	VirD2	+	34278-34946	38/119 (31%)	Type IV de secretion. Composant VirD2 <i>Clostridium thermocellum</i> ATCC 27405
pTet36		+	35243-35665	40/131 (30%)	Protéine hypothétique TDE1306 <i>Treponema denticola</i>
pTet37	Rep	+	35858-37006	133/356 (37%)	Protéine de replication
pTet38		+	37283-37633	23/27 (85%)	Inconnue Cjp20 <i>Campylobacter jejuni</i>

Gène	Nom putatif	Brin	Start-Stop	Identité	Relation significative
pTet39		+	37807-38205	38/103 (36%)	Inconnue Cjp35 <i>Campylobacter jejuni</i>
pTet40		+	38356-38715	N/A	Inconnue
pTet41		+	38712-39269	N/A	Inconnue
pTet42		+	39274-39540	N/A	Inconnue
pTet43		+	39879-40160	N/A	Inconnue
pTet44		+	40163-40429	N/A	Inconnue
pTet45		+	40476-45212	555/1370 (40%)	ADN methylase <i>Burkholderia cepacia</i>

Tableau 3: Liste des ORFs identifiées au cours de l'annotation de notre séquence du plasmide pTet.

Chapitre V

Utilisation de la puce à ADN de *C. jejuni* TGH
9011 pour la caractérisation des gènes régulés
par Fur

Le fer est un élément essentiel pour la croissance des microorganismes (33). Son importance réside dans ses propriétés d'oxydoréduction puisqu'il peut exister sous deux états d'oxydation, fer ferreux (Fe^{2+}) et fer ferrique (Fe^{3+}). De par ses propriétés chimiques, le fer intervient au sein du site actif de nombreux enzymes dans des processus biochimiques d'oxydo-réduction (114), : le fer est lié aux hémoprotéines (hémoglobine, myoglobine), aux cytochromes, et à des enzymes telles que nitrogénases, ferredoxines, catalases, peroxidases, ou encore ribonucléotide réductases (13).

Dans la nature, le fer est surtout présent sous forme de complexes, en association avec des molécules de minéraux silicatés. En milieu réducteur et anaérobie, l'altération de ces minéraux conduit à la libération de fer ferreux, soluble, et donc facilement assimilable par les microorganismes (49). En aérobiose, par contre, le fer ferreux est spontanément oxydé en fer ferrique qui, en particulier aux pHs physiologiques, s'hydrolyse pour donner des hydroxydes ferriques (13). Au niveau de l'organisme hôte, le fer est présent sous forme complexée à des molécules tel que l'hémine, l'hémoglobine ou la transferrine ou la lactoferrine au niveau extracellulaire (53). Ainsi, nombreux sont les microorganismes qui ont développé des systèmes de solubilisation et de transport du Fe^{3+} faisant intervenir des molécules particulières : les sidérophores (114). Les propriétés chimiques du fer sont également à l'origine d'un stress oxydatif dû à la formation de radicaux hydroxyl hautement réactifs et néfastes pour la cellule (154). Pour cette raison le taux de fer intracellulaire est fortement régulé. Chez les bacteries Gram- telles que celles des genres *Vibrio*, *Yersinia*, *Salmonella*, *Pseudomonas*, ou *Helicobacter* (52), et chez certaines Gram+ comme *Bacillus*, *Staphylococcus* ou encore chez les cyanobactéries (52), cette régulation est entre autres exercée par la protéine Fur.

Fur est une protéine monomérique d'environ 17 kDa présentant un domaine N-terminal ayant une affinité envers l'ADN au niveau d'une séquence consensus, dite Fur-box, ainsi qu'un domaine d'homodimérisation et de fixation au fer ferreux au niveau C-terminal (37). La protéine Fur présenterait également un domaine lui permettant une interaction protéique via un domaine autre que celui impliqué dans l'homodimérisation

Fur est connu comme étant un répresseur de transcription des gènes situés en aval de la séquence Fur-box. En présence de fer, Fur s'attache à l'ion Fe^{2+} , résultant en une configuration lui permettant de s'attacher aux Fur-box, situées au niveau des promoteurs

des gènes régulés par le fer. Cette fixation empêche la transcription des gènes situés en aval par encombrement stérique du promoteur. En condition de faible concentration intracellulaire en fer, la protéine Fur ne présente plus d'affinité envers l'ADN, permettant l'accession et la transcription des gènes situés en aval des Fur-box.

Plus récemment, Fur a été également montré comme régulant positivement la transcription de certains gènes tels que le gène de superoxide dismutase (*sodB*) d'*E. coli* (47,79), le gène de catalase (*katA*) de *Staphylococcus aureus* (79), le gène codant pour une protéine non héminique (*Pfr*) d'*Helicobacter pylori* (41), ou les gènes *panI*, *norB* et *nuoABCDE* codant pour des métallo-protéines contenant du fer impliquées dans les mécanismes de respiration anaérobie et aérobie.

Fur a été également montré comme régulant positivement l'expression de gènes bactériens de façon indirecte au niveau post-traductionnel. Ce processus est créé par la répression d'un répresseur de transcription, un petit ARN régulateur (ARNs) appelée RyhB (105).

En présence de fer, Fur réprime la transcription de *ryhB* permettant la transcription des gènes réprimés par cette ARNs. RyhB est connu comme régulant au moins 6 protéines bactériennes impliquées dans la chélation du fer intracellulaire (105). Chez *Pseudomonas aeruginosa* un mécanisme similaire de régulation positive par Fur a été montré via les ARNs PrrF1 and PrrF2 (168).

Fur est également associé à la régulation de processus cellulaires autre que celui du métabolisme du fer comme par exemple la réponse au stress acide (22,70), le chimiotactisme (87), la production de toxines (136) et d'autres facteurs de virulence (102).

L'espèce *C. jejuni* contient également un homologue du gène *fur*, Cj0400 chez la souche séquencée NCTC 11168 (123). Le profil protéique in vitro d'un mutant de ce gène a été étudié chez cette souche. L'étude a permis la caractérisation de 3 protéines situées dans la membrane externe, 4 protéines du périplasme et de la membrane cytoplasmique et deux protéines cytoplasmiques caractérisées comme étant impliquées dans la réponse au stress oxydatif (157). Un mutant *fur* de *C. jejuni* NCTC 11168 a également été utilisé pour des

expériences de mesure d'expression par puces à ADN (121). 53 gènes au total ont été recensés comme étant régulés par Fur de façon directe ou indirecte.

Il est à noter que le mutant *fur* est affecté dans la colonisation des intestins de poulet (121). La colonisation de ce mutant est dramatiquement réduite comparée à celle de la souche sauvage, ces résultats suggérant l'importance des gènes régulés par la protéine Fur dans le processus de colonisation (121).

Un homologue de la séquence *fur* a été également caractérisé chez la souche *C. jejuni* TGH 9011 (33) par séquençage du gène situé en amont du gène *lys*. Ce gène code pour une protéine de 157 aa (18.1kDa). Le gène présente au niveau de sa séquence promotrice quatre motifs « Fur-box », démontrés comme étant accessibles à Fur et suggérant ainsi une autorégulation (33).

Connaissant l'implication des gènes régulés par le fer dans les mécanismes d'invasion de *C. jejuni*, nous avons voulu utiliser la propriété unique de la puce à ADN de TGH 9011 pour l'identification des gènes directement régulés par Fur chez TGH 9011. Contrairement aux puces à ADN construites à partir de produits PCR ou d'oligonucléotides, notre puce présente l'ensemble des séquences génomiques et donc les séquences promotrices de gènes. Les expériences n'étant pas terminées, les résultats présentés restent préliminaires.

1 Obtention de la protéine Fur

Après les différentes étapes d'amplification PCR du gène *fur* de *C. jejuni* TGH 9011, de son insertion dans le vecteur d'expression pASK-IBA7 et électroporation dans *E. coli* DH5 α , nous avons sélectionné deux clones, CII et CVI. Ces 2 clones après linéarisation avec *EcoRI* présentent sur gel d'agarose une taille correspondant à un clone positif (3780bp, Figure 1)

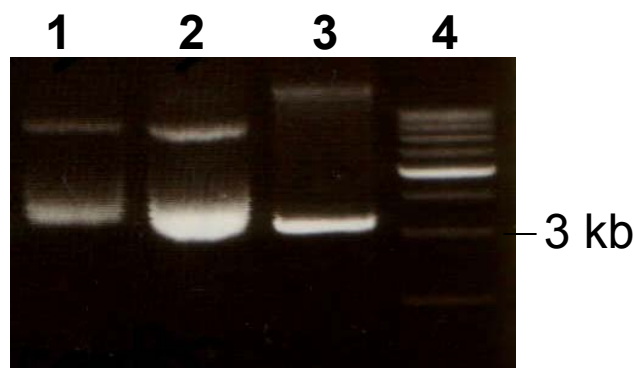


Figure 1: Contrôle des clones fur II et VI par digestion avec *EcoRI*. Colonne 1 : Clone II coupé par *EcoRI*, colonne 2 : Clone VI coupé par *EcoRI*, colonne 3 : vecteur pASK-IBA7 sans insert (contrôle), colonne 4 : marqueurs de taille.

Ces 2 clones ont été utilisés tout d'abord comme source d'ADN pour une PCR via l'utilisation des amorces spécifiques au gène *fur*. Une bande attendue de 505pb a été obtenue pour ces deux clones (Figure 2).

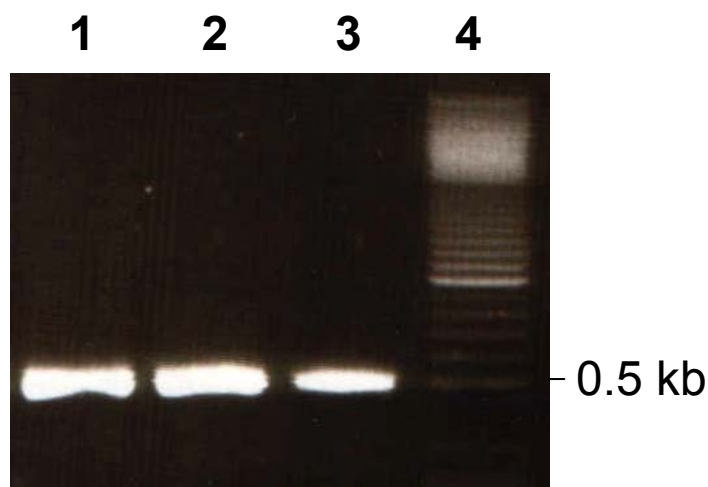


Figure 2 : Contrôle des clones fur II et VI par PCR. Colonne 1 : PCR du clone II coupé, colonne 2 : PCR du clone II, colonne 3 : PCR de l'ADN génomique de TGH 9011 (contrôle), colonne 4 : marqueurs de taille.

Ces produits PCR, ainsi que les plasmides correspondants, ont été ensuite confirmés comme étant positifs par profil de digestion. Pour cela nous avons utilisé les enzymes *BclI* et *Sau3I*. Les profils de restriction décrits sont présentés Fig. 3. A la suite de ces

confirmations, le plasmide issu du clone II est séquencé pour vérification. La séquence obtenue est 100% identique au gène fur. Aucune erreur n'a été introduite au cours de l'amplification PCR.

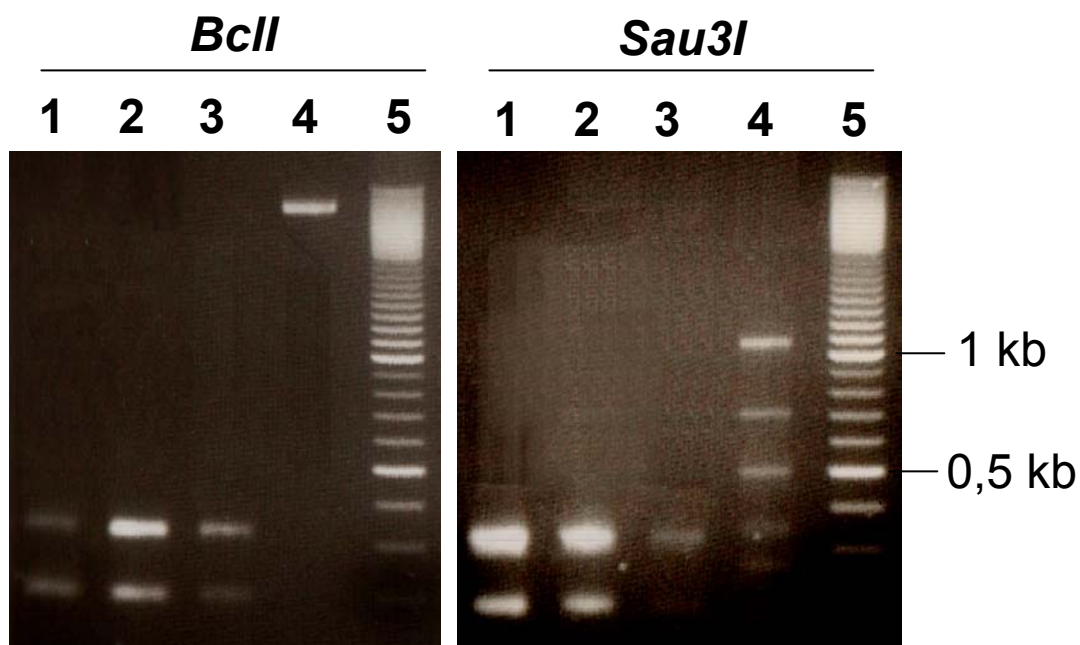


Figure 3 : Digestion contrôlée des clones fur II et VI par les enzymes de restriction BclI et Sau3I. Colonne 1 : PCR de l'insert du clone II digéré, colonne 2, PCR de l'insert du clone VI digéré, colonne 3 : PCR de l'insert génomique de TGH 9011 digéré (contrôle), colonne 4 : digestion du plasmide correspondant au clone II, colonne 5 : marqueurs de taille.

Après purification sur colonne d'affinité de la protéine Fur, issue de 100ml de culture du clone II, l'éluat est déposé sur gel de polyacrylamide 15% en condition dénaturante. Selon des paramètres qui restent à éclaircir, nous obtenons 1 ou 2 bandes après coloration au bleu de Coomassie. Ces bandes correspondent à une protéine de 18kb et à une bande supplémentaire d'environ 22 kb ([Figure 4](#)).

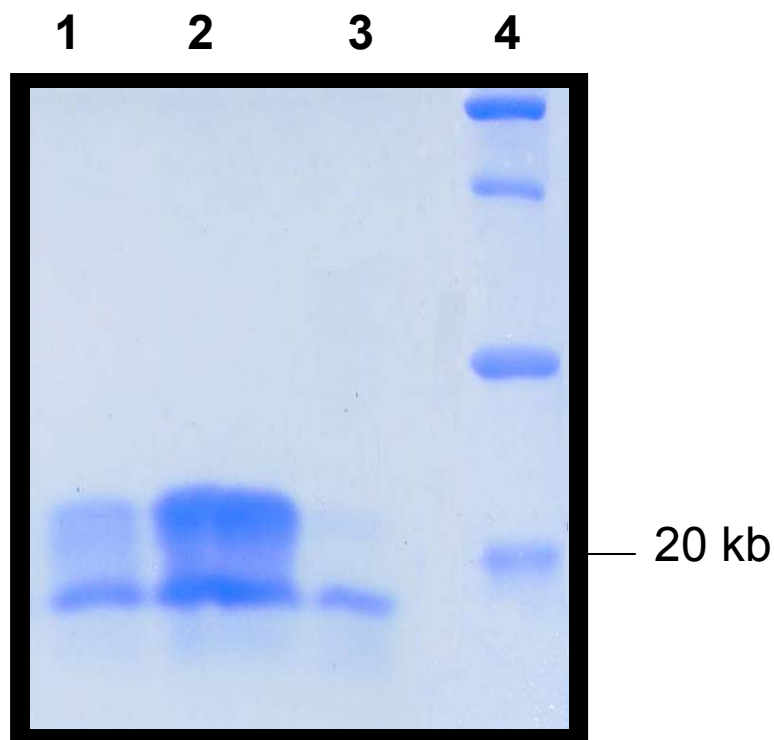


Figure 4 : dépôt de l'éluat contenant la protéine Fur recombinante sur gel de polyacrylamide. Colonne 1 : fraction d'élution de colonne d'affinité numéro 2, Colonne 2 : fraction d'élution de colonne d'affinité numéro 3, Colonne 3 : fraction d'élution de colonne d'affinité numéro 4, colonne 4 : marqueurs de taille.

Ces 2 bandes ont été extraites du gel pour être analysées par spectrophotométrie de masse (MALDI-TOF, Oklahoma State University core facility). Il s'avère que ces bandes correspondent toutes deux à la protéine Fur (Figure 4). A la suite de ces différents tests le clone II est conservé en vue de la production de Fur.

La production de la protéine recombinante est faite dans un volume de 100ml contenu en Erlenmeyer de 750ml de capacité. Ce volume de culture permet en moyenne de récupérer 0.2 a 0.4 mg de protéine recombinante.

2 Production des Anticorps

Trois miligrammes de protéine Fur ont été fournis à la compagnie Invitrogen pour l'obtention d'anticorps anti-Fur. Un total de 24,7 mg d'anticorps ont été obtenus

A la réception des anticorps anti-Fur de lapin, nous avons testé la bonne marche du processus d'immunisation et de purification. Pour cela nous avons réalisé une immunoprécipitation de la protéine Fur purifiée dans le tampon de lyse, suivie d'une séparation par des billes magnétiques (MagnaBind). Le surnageant d'élution a été ensuite analysé sur gel de SDS-polyacrylamide 15%. Les résultats sont présentés Fig. 5. Les résultats montrent la présence d'une bande d'environ 18 kDa dans les puits correspondants à la protéine Fur purifiée par l'anticorps.

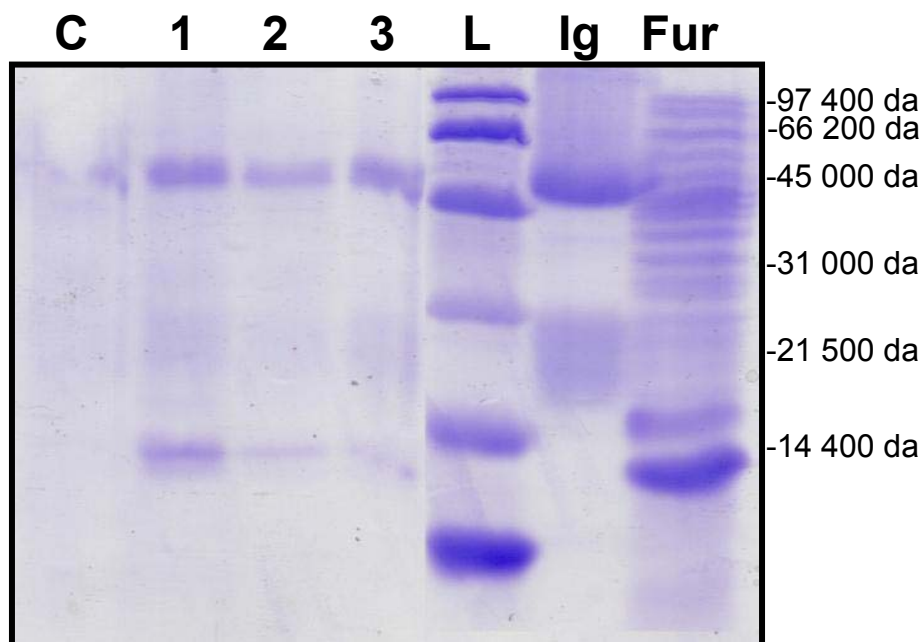


Figure 5: Surnageant d'immunoprécipitation purifiés par billes MagnaBind et déposés sur gel de SDS polyacrylamide 15%. C : contrôle d'immunoprécipitation sans Anticorps. Puits 1 : 100 µg d'anticorps et 100 µg de protéine Fur. Puits 2 : 100 µg d'anticorps et 50 µg de protéine Fur. Puits 3 : 100 µg d'anticorps et 10 µg de protéine Fur. L : marqueurs de taille. Ig : dépôt de 200 µg d'anticorps. Fur : dépôts de 200 µg de protéine Fur.

3 Immunoprécipitation

3-1 Principe :

Cette technique va nous permettre d'isoler les séquences nucléotidiques reconnues sur lesquelles Fur va se fixer, et ainsi détecter les gènes régulés par cette protéine.

Une culture bactérienne en phase exponentielle de croissance est additionnée d'un volume final de 1% de formamide. Ce composé attache tous les complexes protéine/protéine et protéine/acide nucléique de façon covalente. Après ce traitement les cellules sont lysées et l'ADN cassé par ultrasons en vue de l'obtention de fragments de taille moyenne d'environ 700 pb. Le surnageant de lyse est additionné de l'anticorps anti-Fur et mis à incuber 90 min à température ambiante. Le complexe reconnu (ADN-Fur-anticorps anti Fur) est purifié à l'aide de billes magnétiques recouvertes de protéine-A attachant spécifiquement la partie conservée des anticorps. Après purification le complexe Fur/ADN est dissocié par chauffage. L'ADN est ensuite purifié et additionné d'amorces permettant une future amplification par PCR. Les produits resultants sont marqués et hybridés sur la puce à ADN de TGH 9011.

3-2 Calculs théoriques de la quantité d'anticorps et de billes MagnaBind à utiliser pour l'immunoprécipitation.

Masse moléculaire des anticorps (IgG)= 150 000 g/mol

Masse moléculaire de la protéine Fur= 18 000 g/mol

Masse moléculaire moyenne d'une paire de base= 660 g/mol

Capacité de fixation des IgG sur les billes magnétiques= 200 µg/ml

Nombre de bactéries pour une DO_{600} de 1= $7,09 \cdot 10^9$ bactéries/ml

Pour un millilitre d'une culture ayant atteint une DO_{600} de 0,8 nous avons donc $5,7 \cdot 10^9$ bactéries. Si Fur est capable de fixer 30 différents sites sur le chromosome bactérien nous devrions obtenir après sonication 30 fragments de 700 pb. Un fragment d'ADN double brin de 700 pb pèse en moyenne 854 000 g/mol. Pour un millilitre de culture nous

devrions donc obtenir $30 \times 5,7 \cdot 10^9$ copies de morceaux d'ADN soit $1,71 \cdot 10^{11}$ molécules ou encore $2,84 \cdot 10^{-13}$ mol, représentant un poids de $2,62 \cdot 10^{-7}$ g d'ADN pour 1 ml de culture.

Sachant que deux molécules Fur se fixent sur la séquence Fur-Box et qu'un anticorps est capable de fixer 2 protéines Fur, il faut $2,84 \cdot 10^{-13}$ mol anticorps/ml de culture soit $150\,000 \times 2,84 \cdot 10^{-13} = 4,26 \cdot 10^{-8}$ g. Nous avons une solution d'anticorps concentrée à 1,9 mg/ml. L'addition de 22 nl de cette solution est en théorie suffisante à la chélation de toutes les protéines Fur fixées à l'ADN correspondant à un millilitre de culture bactérienne.

Avec une capacité de fixation de 200 μ g d'anticorps par millilitre, il faut environ 0,20 μ l de billes magnétiques pour la fixation des $4,26 \cdot 10^{-8}$ g d'anticorps

En conclusion pour 1ml d'une culture bactérienne de $DO_{600}=0,8$ il faut ajouter 22 nl d'anticorps et 0,2 μ l de billes magnétiques pour récupérer un maximum de 240 ng d'ADN.

Pour nos expériences nous allons utiliser 10ml de culture bactérienne avec 22 μ l d'anticorps (100 fois plus qu'en théorie pour saturer les protéines Fur libres) et 30 μ l de billes magnétiques pour récupérer 2.4 μ g d'ADN.

Chapitre VI

Conclusions et Perspectives

Dans cette étude nous avons construit la première puce à ADN de type « shotgun » correspondant à des clones plasmidiques d'une banque génomique totale de deux souches de *C. jejuni*, *C. jejuni* TGH 9011 et 81-176. Nous avons démontré l'efficacité de son utilisation pour l'identification rapide de la diversité génomique de ces souches bactériennes, par comparaison avec la souche séquencée NCTC 11168.

Cette technologie nous a permis d'identifier chez la souche *C. jejuni* TGH 9011 84 kb d'ADN original, représentant au total 130 ORFs. Ces ORFs codent pour des protéines impliquées dans des mécanismes de pathogénicité reconnus de *C. jejuni*, tels que la biosynthèse de la capsule ou du lipooligosaccharide. Il est remarquable que 28 ORFs présentent des homologies avec des gènes appartenant à un PAI d'*H. hepaticus*. Etant donné que les produits de ces gènes n'ont pas d'homologie avec des protéines de fonction connues, ils pourraient ainsi être impliqués dans la pathogénicité de *C. jejuni* TGH 9011. Ces 28 gènes sont de bons candidats à une étude plus approfondie par mutagenèse de leur implication potentielle dans la pathogénicité de cette souche.

La mise en œuvre de la puce à ADN de la souche 81-176 a, de même, été entreprise dans l'objectif de caractériser les séquences chromosomiques impliquées dans la pathogénicité de cette souche. Ces études nous ont permis d'identifier 86 ORFs qui pourraient être impliqués. Au cours de cette étude nous avons également été amené à séquencer en intégralité le plasmide pTet, mettant en évidence un grand nombre de gènes impliqués dans le mécanisme de sécrétion de type IV. La comparaison des gènes identifiés comme uniques, nous a permis d'identifier un seul gène commun à ces deux souches. Ce gène présente des homologies avec un gène impliqué dans le système de sécrétion de type IV, *traG*. Celui-ci représente un très bon candidat à l'analyse phénotypique de son mutant dans le but d'identifier son rôle dans les mécanismes d'invasion des cellules humaines ou de colonisation d'intestin dans les modèles animaux.

L'identification de ces nouveaux gènes va également servir à la construction d'une puce à ADN de type composite, comprenant l'ensemble des séquences détectées au cours de ces analyses ainsi que les séquences de NCTC 11168. Cette puce à ADN pourra s'avérer utile pour des expériences de typage génomique utile à la classification de souches de *C. jejuni*, ainsi qu'en épidémiologie pour le typage rapide de souches émergentes impliquées

par exemple dans des épidémies. L'analyse des différentes souches caractérisées comme présentant des phénotypes distincts très marqués pourrait permettre par recoupement, de discerner les gènes responsables de ces phénotypes.

La technologie des puces à ADN de type « shotgun », représente une efficace alternative aux techniques utilisées jusqu'à présent pour le typage génomique, tels que SSH (mets le nom en entier) ou le séquençage entier de génome bactérien. Effectivement, même si l'avancement technologique des méthodes de séquençage d'ADN a évolué de façon très rapide au cours des dernières années, le séquençage de génomes de souches proches phylogénétiquement entraîne une inévitable redondance pour de nombreux gènes. De plus, si le temps mis pour un séquençage entier de génome a été considérablement réduit, il subsiste que l'annotation reste encore une étape limitante au rendement de génome séquencés.

La technique de puces à ADN de type « shotgun » présente cependant quelques désavantages comparée au séquençage systématique tels que l'absence de localisation physique, l'absence d'information sur des mutations ponctuelles, sur de réarrangements chromosomiques ou de diversités nucléotidiques inférieurs à 700 pb. Après analyse des séquences uniques détectées au cours de ce travail, il a pu être constaté en effet que des fragments d'ADN uniques à la souche testée et inférieurs à environ 600pb, sont très difficilement détectés. Cela pourrait être une conséquence de la taille des fragments d'ADN déposés sur la puce. Effectivement, il paraît difficile de détecter des fragments d'ADN présentant plus de 20% de séquences communes entre les deux souches. Notre sélection et/ou la sensibilité de la technique, ne permet pas de détecter ces petits fragments.

En complément à ces travaux génomiques, une étude portant sur le métabolisme du fer chez *C. jejuni*, a été entreprise.

Au cours de cette étude nous avons voulu prendre avantage des particularités de la puce à ADN de la souche TGH 9011. Effectivement, comparée à une puce à ADN classique construite par dépôt d'oligonucléotides ou de produits PCR, notre puce devrait contenir (pourquoi pas contient ??) l'ensemble du génome comprenant ainsi les régions intergéniques et donc les promoteurs des gènes. Nous avons voulu tirer partie de cette spécificité pour l'identification des gènes directement régulés par le facteur de

transcription Fur. Cette protéine est impliquée dans la régulation de transcription des gènes bactériens impliqués dans le métabolisme du fer. Fur en présence de fer présente une affinité envers une séquence d'ADN conservé appelée Fur-box. Ces séquences se trouvent au niveau des promoteurs des gènes régulés par cette protéine. En présence de fer, la liaison de Fur au niveau de la séquence Fur-box empêche la transcription de ces gènes. Au cours de notre étude nous avons débuté le processus d'identification de ces gènes par utilisation de notre puce à ADN et d'une immunoprécipitation du complexe ADN-Fur. Les expériences en cours devraient nous permettre de sélectionner les gènes situés en amont des séquences reconnues par Fur et les gènes impliqués dans ce mécanisme. En théorie le même type d'expérience peut être effectué avec n'importe quel facteur de transcription.

Reference List

1. 2003. Preliminary FoodNet data on the incidence of foodborne illnesses--selected sites, United States, 2002. *MMWR Morb. Mortal. Wkly. Rep.* **52**:340-343.
2. 2004. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food--selected sites, United States, 2003. *MMWR Morb. Mortal. Wkly. Rep.* **53**:338-343.
3. 2002. Outbreak of *Campylobacter jejuni* infections associated with drinking unpasteurized milk procured through a cow-leasing program--Wisconsin, 2001. *MMWR Morb. Mortal. Wkly. Rep.* **51**:548-549.
4. Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, P. W. Li, R. A. Hoskins, R. F. Galle, R. A. George, S. E. Lewis, S. Richards, M. Ashburner, S. N. Henderson, G. G. Sutton, J. R. Wortman, M. D. Yandell, Q. Zhang, L. X. Chen, R. C. Brandon, Y. H. Rogers, R. G. Blazej, M. Champe, B. D. Pfeiffer, K. H. Wan, C. Doyle, E. G. Baxter, G. Helt, C. R. Nelson, G. L. Gabor, J. F. Abril, A. Agbayani, H. J. An, C. Andrews-Pfannkoch, D. Baldwin, R. M. Ballew, A. Basu, J. Baxendale, L. Bayraktaroglu, E. M. Beasley, K. Y. Beeson, P. V. Benos, B. P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M. R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K. C. Burtis, D. A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J. M. Cherry, S. Cawley, C. Dahlke, L. B. Davenport, P. Davies, P. B. de, A. Delcher, Z. Deng, A. D. Mays, I. Dew, S. M. Dietz, K. Dodson, L. E. Doup, M. Downes, S. Dugan-Rocha, B. C. Dunkov, P. Dunn, K. J. Durbin, C. C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A. E. Gabrielian, N. S. Garg, W. M. Gelbart, K. Glasser, A. Glodek, F. Gong, J. H. Gorrell, Z. Gu, P. Guan, M. Harris, N. L. Harris, D. Harvey, T. J. Heiman, J. R. Hernandez, J. Houck, D. Hostin, K. A. Houston, T. J. Howland, M. H. Wei, C. Ibegwam, M. Jalali, F. Kalush, G. H. Karpen, Z. Ke, J. A. Kennison, K. A. Ketchum, B. E. Kimmel, C. D. Kodira, C. Kraft, S. Kravitz, D. Kulp, Z. Lai, P. Lasko, Y. Lei, A. A. Levitsky, J. Li, Z. Li, Y. Liang, X. Lin, X. Liu, B. Mattei, T. C. McIntosh, M. P. McLeod, D. McPherson, G. Merkulov, N. V. Milshina, C. Mobarry, J. Morris, A. Moshrefi, S. M. Mount, M. Moy, B. Murphy, L. Murphy, D. M. Muzny, D. L. Nelson, D. R. Nelson, K. A. Nelson, K. Nixon, D. R. Nusskern, J. M. Pacle, M. Palazzolo, G. S. Pittman, S. Pan, J. Pollard, V. Puri, M. G. Reese, K. Reinert, K. Remington, R. D. Saunders, F. Scheeler, H. Shen, B. C. Shue, I. Siden-Kiamos, M. Simpson, M. P. Skupski, T. Smith, E. Spier, A. C. Spradling, M. Stapleton, R. Strong, E. Sun, R. Svirskas, C. Tector, R. Turner, E. Venter, A. H. Wang, X. Wang, Z. Y. Wang, D. A. Wassarman, G. M. Weinstock, J. Weissenbach, S. M. Williams, Woodage, T. K. C. Worley, D.

- Wu, S. Yang, Q. A. Yao, J. Ye, R. F. Yeh, J. S. Zaveri, M. Zhan, G. Zhang, Q. Zhao, L. Zheng, X. H. Zheng, F. N. Zhong, W. Zhong, X. Zhou, S. Zhu, X. Zhu, H. O. Smith, R. A. Gibbs, E. W. Myers, G. M. Rubin, and J. C. Venter. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**:2185-2195.
5. Agron, P. G., M. Macht, L. Radnedge, E. W. Skowronski, W. Miller, and G. L. Andersen. 2002. Use of subtractive hybridization for comprehensive surveys of prokaryotic genome differences. *FEMS Microbiol. Lett.* **211**:175-182.
 6. Ahmed, I. H., G. Manning, T. M. Wassenaar, S. Cawthraw, and D. G. Newell. 2002. Identification of genetic differences between two *Campylobacter jejuni* strains with different colonization potentials. *Microbiology* **148**:1203-1212.
 7. Akopyants, N. S., A. Fradkov, L. Diatchenko, J. E. Hill, P. D. Siebert, S. A. Lukyanov, E. D. Sverdlov, and D. E. Berg. 1998. PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. *Proc. Natl. Acad. Sci. U. S. A* **95**:13108-13113.
 8. Allos, B. M. 2001. *Campylobacter jejuni* Infections: update on emerging issues and trends. *Clin. Infect. Dis.* **32**:1201-1206.
 9. Allos, B. M. and M. J. Blaser. 1995. *Campylobacter jejuni* and the expanding spectrum of related infections. *Clin. Infect. Dis.* **20**:1092-1099.
 10. Alm, R. A., L. S. Ling, D. T. Moir, B. L. King, E. D. Brown, P. C. Doig, D. R. Smith, B. Noonan, B. C. Guild, B. L. deJonge, G. Carmel, P. J. Tummino, A. Caruso, M. Uria-Nickelsen, D. M. Mills, C. Ives, R. Gibson, D. Merberg, S. D. Mills, Q. Jiang, D. E. Taylor, G. F. Vovis, and T. J. Trust. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397**:176-180.
 11. Alm, R. A. and T. J. Trust. 1999. Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes. *J. Mol. Med.* **77**:834-846.
 12. Altekruze, S. F., N. J. Stern, P. I. Fields, and D. L. Swerdlow. 1999. *Campylobacter jejuni*--an emerging foodborne pathogen. *Emerg. Infect. Dis.* **5**:28-35.
 13. Andrews, S. C., A. K. Robinson, and F. Rodriguez-Quinones. 2003. Bacterial iron homeostasis. *FEMS Microbiol. Rev.* **27**:215-237.
 14. Arber, W. 2000. Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiol. Rev.* **24**:1-7.
 15. Aspinall, G. O., C. M. Lynch, H. Pang, R. T. Shaver, and A. P. Moran. 1995. Chemical structures of the core region of *Campylobacter jejuni* O:3 lipopolysaccharide and an associated polysaccharide. *Eur. J. Biochem.* **231**:570-578.

16. **Bacon, D. J., R. A. Alm, D. H. Burr, L. Hu, D. J. Kopecko, C. P. Ewing, T. J. Trust, and P. Guerry.** 2000. Involvement of a plasmid in virulence of *Campylobacter jejuni* 81-176. *Infect. Immun.* **68**:4384-4390.
17. **Bacon, D. J., R. A. Alm, L. Hu, T. E. Hickey, C. P. Ewing, R. A. Batchelor, T. J. Trust, and P. Guerry.** 2002. DNA sequence and mutational analyses of the pVir plasmid of *Campylobacter jejuni* 81-176. *Infect. Immun.* **70**:6242-6250.
18. **Bacon, D. J., C. M. Szymanski, D. H. Burr, R. P. Silver, R. A. Alm, and P. Guerry.** 2001. A phase-variable capsule is involved in virulence of *Campylobacter jejuni* 81-176. *Mol. Microbiol.* **40**:769-777.
19. **Batchelor, R. A., B. M. Pearson, L. M. Friis, P. Guerry, and J. M. Wells.** 2004. Nucleotide sequences and comparison of two large conjugative plasmids from different *Campylobacter* species. *Microbiology* **150**:3507-3517.
20. **Beltran, P., J. M. Musser, R. Helmuth, J. J. Farmer, III, W. M. Frerichs, I. K. Wachsmuth, K. Ferris, A. C. McWhorter, J. G. Wells, and A. Cravioto.** 1988. Toward a population genetic analysis of *Salmonella*: genetic diversity and relationships among strains of serotypes *S. choleraesuis*, *S. derby*, *S. dublin*, *S. enteritidis*, *S. heidelberg*, *S. infantis*, *S. newport*, and *S. typhimurium*. *Proc. Natl. Acad. Sci. U. S. A* **85**:7753-7757.
21. **Bereswill, S. and M. Kist.** 2003. Recent developments in *Campylobacter* pathogenesis. *Curr. Opin. Infect. Dis.* **16**:487-491.
22. **Bijlsma, J. J., B. Waidner, A. H. Vliet, N. J. Hughes, S. Hag, S. Bereswill, D. J. Kelly, C. M. Vandenbroucke-Grauls, M. Kist, and J. G. Kusters.** 2002. The *Helicobacter pylori* homologue of the ferric uptake regulator is involved in acid resistance. *Infect. Immun.* **70**:606-611.
23. **Black, R. E., M. M. Levine, M. L. Clements, T. P. Hughes, and M. J. Blaser.** 1988. Experimental *Campylobacter jejuni* infection in humans. *J. Infect. Dis.* **157**:472-479.
24. **Blattner, F. R., G. Plunkett, III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**:1453-1474.
25. **Boneca, I. G., R. H. de, J. C. Epinat, M. Pupin, A. Labigne, and I. Moszer.** 2003. A revised annotation and comparative analysis of *Helicobacter pylori* genomes. *Nucleic Acids Res.* **31**:1704-1714.
26. **Braun, V., R. Schonherr, and S. Hobbie.** 1993. Enterobacterial hemolysins: activation, secretion and pore formation. *Trends Microbiol.* **1**:211-216.

27. **Brenner, D. J., G. R. Fanning, A. G. Steigerwalt, I. Orskov, and F. Orskov.** 1972. Polynucleotide sequence relatedness among three groups of pathogenic *Escherichia coli* strains. *Infect. Immun.* **6**:308-315.
28. **Butzler, J. P.** 2004. *Campylobacter*, from obscurity to celebrity. *Clin. Microbiol. Infect.* **10**:868-876.
29. **Carniel, E.** 2001. The *Yersinia* high-pathogenicity island: an iron-uptake island. *Microbes. Infect.* **3**:561-569.
30. **Carrillo, C. D., E. Taboada, J. H. Nash, P. Lanthier, J. Kelly, P. C. Lau, R. Verhulp, O. Mykytczuk, 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.
31. **Cascales, E. and P. J. Christie.** 2003. The versatile bacterial type IV secretion systems. *Nat. Rev. Microbiol.* **1**:137-149.
32. **Cawthraw, S. A., T. M. Wassenaar, R. Ayling, and D. G. Newell.** 1996. Increased colonization potential of *Campylobacter jejuni* strain 81116 after passage through chickens and its implication on the rate of transmission within flocks. *Epidemiol. Infect.* **117**:213-215.
33. **Chan, V. L., H. Louie, and H. L. Bingham.** 1995. Cloning and transcription regulation of the ferric uptake regulatory gene of *Campylobacter jejuni* TGH9011. *Gene* **164**:25-31.
34. **Chen, T. and S. S. Skiena.** 2000. A case study in genome-level fragment assembly. *Bioinformatics.* **16**:494-500.
35. **Colwell, R. R.** 1970. Polyphasic taxonomy of the genus *Vibrio*: numerical taxonomy of *Vibrio cholerae*, *Vibrio parahaemolyticus*, and related *Vibrio* species. *J. Bacteriol.* **104**:410-433.
36. **Cornelis, G. R., T. Biot, R. C. Lambert de, T. Michiels, B. Mulder, C. Sluiter, M. P. Sory, B. M. Van, and J. C. Vanooteghem.** 1989. The *Yersinia* *yop* regulon. *Mol. Microbiol.* **3**:1455-1459.
37. **Coy, M. and J. B. Neilands.** 1991. Structural dynamics and functional domains of the *fur* protein. *Biochemistry* **30**:8201-8210.
38. **Dasika, G. K., S. C. Lin, S. Zhao, P. Sung, A. Tomkinson, and E. Y. Lee.** 1999. DNA damage-induced cell cycle checkpoints and DNA strand break repair in development and tumorigenesis. *Oncogene* **18**:7883-7899.

39. **Daubin, V., N. A. Moran, and H. Ochman.** 2003. Phylogenetics and the cohesion of bacterial genomes. *Science* **301**:829-832.
40. **Dekeyser, P., M. Gossuin-Detrain, J. P. Butzler, and J. Sternon.** 1972. Acute enteritis due to related vibrio: first positive stool cultures. *J. Infect. Dis.* **125**:390-392.
41. **Delany, I., G. Spohn, R. Rappuoli, and V. Scarlato.** 2001. The Fur repressor controls transcription of iron-activated and -repressed genes in *Helicobacter pylori*. *Mol. Microbiol.* **42**:1297-1309.
42. **Deng, W., V. Burland, G. Plunkett, III, A. Boutin, G. F. Mayhew, P. Liss, N. T. Perna, D. J. Rose, B. Mau, S. Zhou, D. C. Schwartz, J. D. Fetherston, L. E. Lindler, R. R. Brubaker, G. V. Plano, S. C. Straley, K. A. McDonough, M. L. Nilles, J. S. Matson, F. R. Blattner, and R. D. Perry.** 2002. Genome sequence of *Yersinia pestis* KIM. *J. Bacteriol.* **184**:4601-4611.
43. **Deng, W., S. R. Liou, G. Plunkett, III, G. F. Mayhew, D. J. Rose, V. Burland, V. Kodoyianni, D. C. Schwartz, and F. R. Blattner.** 2003. Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18. *J. Bacteriol.* **185**:2330-2337.
44. **Diatchenko, L., Y. F. Lau, A. P. Campbell, A. Chenchik, F. Moqadam, B. Huang, S. Lukyanov, K. Lukyanov, N. Gurskaya, E. D. Sverdlov, and P. D. Siebert.** 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. U. S. A* **93**:6025-6030.
45. **Doig, P., N. Kinsella, P. Guerry, and T. J. Trust.** 1996. Characterization of a post-translational modification of *Campylobacter* flagellin: identification of a sero-specific glycosyl moiety. *Mol. Microbiol.* **19**:379-387.
46. **Dorrell, N., J. A. Mangan, K. G. Laing, J. Hinds, D. Linton, H. Al-Ghusein, B. G. Barrell, J. Parkhill, N. G. Stoker, A. V. Karlyshev, P. D. Butcher, and B. W. Wren.** 2001. Whole genome comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity. *Genome Res.* **11**:1706-1715.
47. **Dubrac, S. and D. Touati.** 2000. Fur positive regulation of iron superoxide dismutase in *Escherichia coli*: functional analysis of the *sodB* promoter. *J. Bacteriol.* **182**:3802-3808.
48. **Dziejman, M., E. Balon, D. Boyd, C. M. Fraser, J. F. Heidelberg, and J. J. Mekalanos.** 2002. Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease. *Proc. Natl. Acad. Sci. U. S. A* **99**:1556-1561.

49. **Earhart and C.F.** 1996. Uptake and metabolism of iron and molybdenum, p. 1075-1090. *In* R. C. I. J. L. I. E. C. C. L. K. B. L. B. M. W. S. R. M. R. M. S. a. H. E. U. *In* F.C.Neidhardt (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. American Society for Microbiology, Washington, D.C.
50. **Elliott, S. J., V. Sperandio, J. A. Giron, S. Shin, J. L. Mellies, L. Wainwright, S. W. Hutcheson, T. K. McDaniel, and J. B. Kaper.** 2000. The locus of enterocyte effacement (LEE)-encoded regulator controls expression of both LEE- and non-LEE-encoded virulence factors in enteropathogenic and enterohemorrhagic *Escherichia coli*. *Infect. Immun.* **68**:6115-6126.
51. **Endtz, H. P., J. S. Vliegenthart, P. Vandamme, H. W. Weverink, N. P. van den Braak, H. A. Verbrugh, and B. A. van.** 1997. Genotypic diversity of *Campylobacter lari* isolated from mussels and oysters in The Netherlands. *Int. J. Food Microbiol.* **34**:79-88.
52. **Escolar, L., J. Perez-Martin, and L. de, V.** 1999. Opening the iron box: transcriptional metalloregulation by the Fur protein. *J. Bacteriol.* **181**:6223-6229.
53. **Evans, R. W., J. B. Crawley, R. C. Garratt, J. G. Grossmann, M. Neu, A. Aitken, K. J. Patel, A. Meilak, C. Wong, J. Singh, and .** 1994. Characterization and structural analysis of a functional human serum transferrin variant and implications for receptor recognition. *Biochemistry* **33**:12512-12520.
54. **Everest, P., J. Wain, M. Roberts, G. Rook, and G. Dougan.** 2001. The molecular mechanisms of severe typhoid fever. *Trends Microbiol.* **9**:316-320.
55. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, and J. M. Merrick.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**:496-512.
56. **Fouts, D. E., E. F. Mongodin, R. E. Mandrell, W. G. Miller, D. A. Rasko, J. Ravel, L. M. Brinkac, R. T. Deboy, C. T. Parker, S. C. Daugherty, R. J. Dodson, A. S. Durkin, R. Madupu, S. A. Sullivan, J. U. Shetty, M. A. Ayodeji, A. Shvartsbeyn, M. C. Schatz, J. H. Badger, C. M. Fraser, and K. E. Nelson.** 2005. Major structural differences and novel potential virulence mechanisms from the genomes of multiple campylobacter species. *PLoS. Biol.* **3**:e15.
57. **Fraser, C. M. and R. D. Fleischmann.** 1997. Strategies for whole microbial genome sequencing and analysis. *Electrophoresis* **18**:1207-1216.
58. **Frederick M.Ausubel, R. B. R. E. K. D. D. M. J. G. S. K. S.** 2005. **Current Protocols in Molecular Biology**, *In* . John Wiley and Sons Inc.

59. **Fry, B. N., S. Feng, Y. Y. Chen, D. G. Newell, P. J. Coloe, and V. Korolik.** 2000. The *galE* gene of *Campylobacter jejuni* is involved in lipopolysaccharide synthesis and virulence. *Infect. Immun.* **68**:2594-2601.
60. **Fry, B. N., V. Korolik, J. A. ten Brinke, M. T. Pennings, R. Zalm, B. J. Teunis, P. J. Coloe, and B. A. van der Zeijst.** 1998. The lipopolysaccharide biosynthesis locus of *Campylobacter jejuni* 81116. *Microbiology* **144** (Pt 8):2049-2061.
61. **Gilbert, M., M. F. Karwaski, S. Bernatchez, N. M. Young, E. Taboada, J. Michniewicz, A. M. Cunningham, and W. W. Wakarchuk.** 2002. The genetic bases for the variation in the lipo-oligosaccharide of the mucosal pathogen, *Campylobacter jejuni*. Biosynthesis of sialylated ganglioside mimics in the core oligosaccharide. *J. Biol. Chem.* **277**:327-337.
62. **Godschalk, P. C., A. P. Heikema, M. Gilbert, T. Komagamine, C. W. Ang, J. Glerum, D. Brochu, J. Li, N. Yuki, B. C. Jacobs, B. A. van, and H. P. Endtz.** 2004. The crucial role of *Campylobacter jejuni* genes in anti-ganglioside antibody induction in Guillain-Barre syndrome. *J. Clin. Invest* **114**:1659-1665.
63. **Griffiths, P. L. and R. W. Park.** 1990. *Campylobacters* associated with human diarrhoeal disease. *J. Appl. Bacteriol.* **69**:281-301.
64. **Guerry, P., C. P. Ewing, T. E. Hickey, M. M. Prendergast, and A. P. Moran.** 2000. Sialylation of lipooligosaccharide cores affects immunogenicity and serum resistance of *Campylobacter jejuni*. *Infect. Immun.* **68**:6656-6662.
65. **Guerry, P., C. M. Szymanski, M. M. Prendergast, T. E. Hickey, C. P. Ewing, D. L. Pattarini, and A. P. Moran.** 2002. Phase variation of *Campylobacter jejuni* 81-176 lipooligosaccharide affects ganglioside mimicry and invasiveness in vitro. *Infect. Immun.* **70**:787-793.
66. **Guiyoule, A., F. Grimont, I. Iteman, P. A. Grimont, M. Lefevre, and E. Carniel.** 1994. Plague pandemics investigated by ribotyping of *Yersinia pestis* strains. *J. Clin. Microbiol.* **32**:634-641.
67. **Hacker, J., G. Blum-Oehler, I. Muhldorfer, and H. Tschape.** 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol. Microbiol.* **23**:1089-1097.
68. **Hacker, J. and E. Carniel.** 2001. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep.* **2**:376-381.
69. **Hald, B. and M. Madsen.** 1997. Healthy puppies and kittens as carriers of *Campylobacter* spp., with special reference to *Campylobacter upsaliensis*. *J. Clin. Microbiol.* **35**:3351-3352.

70. **Hall, H. K. and J. W. Foster.** 1996. The role of fur in the acid tolerance response of *Salmonella typhimurium* is physiologically and genetically separable from its role in iron acquisition. *J. Bacteriol.* **178**:5683-5691.
71. **Han, F. C., M. Gong, H. C. Ng, and B. Ho.** 2003. Identification of *H. pylori* strain specific DNA sequences between two clinical isolates from NUD and gastric ulcer by SSH. *World J. Gastroenterol.* **9**:1747-1751.
72. **Hannenhalli, S. S., W. S. Hayes, A. G. Hatzigeorgiou, and J. W. Fickett.** 1999. Bacterial start site prediction. *Nucleic Acids Res.* **27**:3577-3582.
73. **Hayward, R. E., J. L. DeRisi, S. Alfadhli, D. C. Kaslow, P. O. Brown, and P. K. Rathod.** 2000. Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum* malaria. *Mol. Microbiol.* **35**:6-14.
74. **Heidelberg, J. F., J. A. Eisen, W. C. Nelson, R. A. Clayton, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, L. Umayam, S. R. Gill, K. E. Nelson, T. D. Read, H. Tettelin, D. Richardson, M. D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, I. Dragoi, P. Sellers, L. McDonald, T. Utterback, R. D. Fleishmann, W. C. Nierman, O. White, S. L. Salzberg, H. O. Smith, R. R. Colwell, J. J. Mekalanos, J. C. Venter, and C. M. Fraser.** 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**:477-483.
75. **Heidelberg, J. F., J. A. Eisen, W. C. Nelson, R. A. Clayton, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, L. Umayam, S. R. Gill, K. E. Nelson, T. D. Read, H. Tettelin, D. Richardson, M. D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, I. Dragoi, P. Sellers, L. McDonald, T. Utterback, R. D. Fleishmann, W. C. Nierman, O. White, S. L. Salzberg, H. O. Smith, R. R. Colwell, J. J. Mekalanos, J. C. Venter, and C. M. Fraser.** 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**:477-483.
76. **Hinnebusch, B. J., E. R. Fischer, and T. G. Schwan.** 1998. Evaluation of the role of the *Yersinia pestis* plasminogen activator and other plasmid-encoded factors in temperature-dependent blockage of the flea. *J. Infect. Dis.* **178**:1406-1415.
77. **Hinnebusch, B. J., R. D. Perry, and T. G. Schwan.** 1996. Role of the *Yersinia pestis* hemin storage (hms) locus in the transmission of plague by fleas. *Science* **273**:367-370.
78. **Hood, D. W., M. E. Deadman, M. P. Jennings, M. Bisercic, R. D. Fleischmann, J. C. Venter, and E. R. Moxon.** 1996. DNA repeats identify novel virulence genes in *Haemophilus influenzae*. *Proc. Natl. Acad. Sci. U. S. A* **93**:11121-11125.

79. **Horsburgh, M. J., M. O. Clements, H. Crossley, E. Ingham, and S. J. Foster.** 2001. PerR controls oxidative stress resistance and iron storage proteins and is required for virulence in *Staphylococcus aureus*. *Infect. Immun.* **69**:3744-3754.
80. **Hu, L. and D. J. Kopecko.** 1999. *Campylobacter jejuni* 81-176 associates with microtubules and dynein during invasion of human intestinal cells. *Infect. Immun.* **67**:4171-4182.
81. **Israel, D. A., N. Salama, C. N. Arnold, S. F. Moss, T. Ando, H. P. Wirth, K. T. Tham, M. Camorlinga, M. J. Blaser, S. Falkow, and R. M. Peek, Jr.** 2001. *Helicobacter pylori* strain-specific differences in genetic content, identified by microarray, influence host inflammatory responses. *J. Clin. Invest* **107**:611-620.
82. **Janssen, P. J., B. Audit, and C. A. Ouzounis.** 2001. Strain-specific genes of *Helicobacter pylori*: distribution, function and dynamics. *Nucleic Acids Res.* **29**:4395-4404.
83. **Jin, Q., Z. Yuan, J. Xu, Y. Wang, Y. Shen, W. Lu, J. Wang, H. Liu, J. Yang, F. Yang, X. Zhang, J. Zhang, G. Yang, H. Wu, D. Qu, J. Dong, L. Sun, Y. Xue, A. Zhao, Y. Gao, J. Zhu, B. Kan, K. Ding, S. Chen, H. Cheng, Z. Yao, B. He, R. Chen, D. Ma, B. Qiang, Y. Wen, Y. Hou, and J. Yu.** 2002. Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. *Nucleic Acids Res.* **30**:4432-4441.
84. **Johnson, W. M. and H. Lior.** 1988. A new heat-labile cytolethal distending toxin (CLDT) produced by *Campylobacter* spp. *Microb. Pathog.* **4**:115-126.
85. **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.
86. **Joseph, S. W., R. R. Colwell, and J. B. Kaper.** 1982. *Vibrio parahaemolyticus* and related halophilic *Vibrios*. *Crit Rev. Microbiol.* **10**:77-124.
87. **Karjalainen, T. K., D. G. Evans, D. J. Evans, Jr., D. Y. Graham, and C. H. Lee.** 1991. Iron represses the expression of CFA/I fimbriae of enterotoxigenic *E. coli*. *Microb. Pathog.* **11**:317-323.
88. **Karlyshev, A. V., P. Everest, D. Linton, S. Cawthraw, D. G. Newell, and B. W. Wren.** 2004. The *Campylobacter jejuni* general glycosylation system is important for attachment to human epithelial cells and in the colonization of chicks. *Microbiology* **150**:1957-1964.
89. **Karlyshev, A. V., D. Linton, N. A. Gregson, A. J. Lastovica, and B. W. Wren.** 2000. Genetic and biochemical evidence of a *Campylobacter jejuni* capsular

- polysaccharide that accounts for Penner serotype specificity. *Mol. Microbiol.* **35**:529-541.
90. **Karlyshev, A. V., D. Linton, N. A. Gregson, and B. W. Wren.** 2002. A novel paralogous gene family involved in phase-variable flagella-mediated motility in *Campylobacter jejuni*. *Microbiology* **148**:473-480.
 91. **Karlyshev, A. V. and B. W. Wren.** 2001. Detection and initial characterization of novel capsular polysaccharide among diverse *Campylobacter jejuni* strains using alcian blue dye. *J. Clin. Microbiol.* **39**:279-284.
 92. **Ketley, J. M.** 1997. Pathogenesis of enteric infection by *Campylobacter*. *Microbiology* **143** (Pt 1):5-21.
 93. **Korolik, V., B. N. Fry, M. R. Alderton, B. A. van der Zeijst, and P. J. Coloe.** 1997. Expression of *Campylobacter hyoilei* lipo-oligosaccharide (LOS) antigens in *Escherichia coli*. *Microbiology* **143** (Pt 11):3481-3489.
 94. **Kotloff, K. L., J. P. Winickoff, B. Ivanoff, J. D. Clemens, D. L. Swerdlow, P. J. Sansonetti, G. K. Adak, and M. M. Levine.** 1999. Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. *Bull. World Health Organ* **77**:651-666.
 95. **Lara-Tejero, M. and J. E. Galan.** 2002. Cytolethal distending toxin: limited damage as a strategy to modulate cellular functions. *Trends Microbiol.* **10**:147-152.
 96. **Lara-Tejero, M. and J. E. Galan.** 2000. A bacterial toxin that controls cell cycle progression as a deoxyribonuclease I-like protein. *Science* **290**:354-357.
 97. **Lara-Tejero, M. and J. E. Galan.** 2001. CdtA, CdtB, and CdtC form a tripartite complex that is required for cytolethal distending toxin activity. *Infect. Immun.* **69**:4358-4365.
 98. **Leonard, E. E., T. Takata, M. J. Blaser, S. Falkow, L. S. Tompkins, and E. C. Gaynor.** 2003. Use of an open-reading frame-specific *Campylobacter jejuni* DNA microarray as a new genotyping tool for studying epidemiologically related isolates. *J. Infect. Dis.* **187**:691-694.
 99. **Leonard, E. E., L. S. Tompkins, S. Falkow, and I. Nachamkin.** 2004. Comparison of *Campylobacter jejuni* isolates implicated in Guillain-Barre syndrome and strains that cause enteritis by a DNA microarray. *Infect. Immun.* **72**:1199-1203.
 100. **Linton, D., M. Gilbert, P. G. Hitchen, A. Dell, H. R. Morris, W. W. Wakarchuk, N. A. Gregson, and B. W. Wren.** 2000. Phase variation of a beta-1,3 galactosyltransferase involved in generation of the ganglioside GM1-like lipo-oligosaccharide of *Campylobacter jejuni*. *Mol. Microbiol.* **37**:501-514.

101. **Linton, D., A. V. Karlyshev, and B. W. Wren.** 2001. Deciphering *Campylobacter jejuni* cell surface interactions from the genome sequence. *Curr. Opin. Microbiol.* **4**:35-40.
102. **Litwin, C. M. and S. B. Calderwood.** 1993. Role of iron in regulation of virulence genes. *Clin. Microbiol. Rev.* **6**:137-149.
103. **Luo, N. and Q. Zhang.** 2001. Molecular characterization of a cryptic plasmid from *Campylobacter jejuni*. *Plasmid* **45**:127-133.
104. **Makino, K., K. Oshima, K. Kurokawa, K. Yokoyama, T. Uda, K. Tagomori, Y. Iijima, M. Najima, M. Nakano, A. Yamashita, Y. Kubota, S. Kimura, T. Yasunaga, T. Honda, H. Shinagawa, M. Hattori, and T. Iida.** 2003. Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. *Lancet* **361**:743-749.
105. **Masse, E. and S. Gottesman.** 2002. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A* **99**:4620-4625.
106. **Miller WG, M. R.** 2004. *Campylobacter* in the food and water supply: Prevalence, outbreaks, isolation, and detection., p. 109-163. *In* . Horizon Scientific Press.
107. **Miller, W. G., A. H. Bates, S. T. Horn, M. T. Brandl, M. R. Wachtel, and R. E. Mandrell.** 2000. Detection on surfaces and in Caco-2 cells of *Campylobacter jejuni* cells transformed with new gfp, yfp, and cfp marker plasmids. *Appl. Environ. Microbiol.* **66**:5426-5436.
108. **Mishu, B., A. A. Ilyas, C. L. Koski, F. Vriesendorp, S. D. Cook, F. A. Mithen, and M. J. Blaser.** 1993. Serologic evidence of previous *Campylobacter jejuni* infection in patients with the Guillain-Barre syndrome. *Ann. Intern. Med.* **118**:947-953.
109. **Moens, S. and J. Vanderleyden.** 1997. Glycoproteins in prokaryotes. *Arch. Microbiol.* **168**:169-175.
110. **Moran, A. P., M. M. Prendergast, and B. J. Appelmelk.** 1996. Molecular mimicry of host structures by bacterial lipopolysaccharides and its contribution to disease. *FEMS Immunol. Med. Microbiol.* **16**:105-115.
111. **Morooka, T., A. Umeda, and K. Amako.** 1985. Motility as an intestinal colonization factor for *Campylobacter jejuni*. *J. Gen. Microbiol.* **131**:1973-1980.
112. **Munson, R. S., Jr., A. Harrison, A. Gillasp, W. C. Ray, M. Carson, D. Armbruster, J. Gipson, M. Gipson, L. Johnson, L. Lewis, D. W. Dyer, and L. O. Bakaletz.** 2004. Partial analysis of the genomes of two nontypeable *Haemophilus influenzae* otitis media isolates. *Infect. Immun.* **72**:3002-3010.

113. **Nakata, N., T. Tobe, I. Fukuda, T. Suzuki, K. Komatsu, M. Yoshikawa, and C. Sasakawa.** 1993. The absence of a surface protease, OmpT, determines the intercellular spreading ability of *Shigella*: the relationship between the ompT and kcpA loci. *Mol. Microbiol.* **9**:459-468.
114. **Neilands, J. B.** 1973. Chemistry of iron in biological systems. *Adv. Exp. Med. Biol.* **40**:13-42.
115. **Newell, D. G., H. McBride, and J. M. Dolby.** 1985. Investigations on the role of flagella in the colonization of infant mice with *Campylobacter jejuni* and attachment of *Campylobacter jejuni* to human epithelial cell lines. *J. Hyg. (Lond)* **95**:217-227.
116. **Nishikubo, S., M. Ohara, Y. Ueno, M. Ikura, H. Kurihara, H. Komatsuzawa, E. Oswald, and M. Sugai.** 2003. An N-terminal segment of the active component of the bacterial genotoxin cytolethal distending toxin B (CDTB) directs CDTB into the nucleus. *J. Biol. Chem.* **278**:50671-50681.
117. **Nishikubo, S., M. Ohara, Y. Ueno, M. Ikura, H. Kurihara, H. Komatsuzawa, E. Oswald, and M. Sugai.** 2003. An N-terminal segment of the active component of the bacterial genotoxin cytolethal distending toxin B (CDTB) directs CDTB into the nucleus. *J. Biol. Chem.* **278**:50671-50681.
118. **Ochman, H., J. G. Lawrence, and E. A. Groisman.** 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**:299-304.
119. **Ochman, H. and N. A. Moran.** 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* **292**:1096-1099.
120. **Orskov, F. and I. Orskov.** 1992. *Escherichia coli* serotyping and disease in man and animals. *Can. J. Microbiol.* **38**:699-704.
121. **Palyada, K., D. Threadgill, and A. Stintzi.** 2004. Iron acquisition and regulation in *Campylobacter jejuni*. *J. Bacteriol.* **186**:4714-4729.
122. **Parkhill, J., G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. Holden, M. Sebaihia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connerton, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell.** 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413**:848-852.
123. **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.
124. **Parkhill, J., B. W. Wren, N. R. Thomson, R. W. Titball, M. T. Holden, M. B. Prentice, M. Sebaihia, K. D. James, C. Churcher, K. L. Mungall, S. Baker, D. Basham, S. D. Bentley, K. Brooks, A. M. Cerdeno-Tarraga, T. Chillingworth, A. Cronin, R. M. Davies, P. Davis, G. Dougan, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Leather, S. Moule, P. C. Oyston, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell.** 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**:523-527.
 125. **Pearson, B. M., C. Pin, J. Wright, K. I'Anson, T. Humphrey, and J. M. Wells.** 2003. Comparative genome analysis of *Campylobacter jejuni* using whole genome DNA microarrays. *FEBS Lett.* **554**:224-230.
 126. **Perez Perez, G. I. and M. J. Blaser.** 1985. Lipopolysaccharide characteristics of pathogenic campylobacters. *Infect. Immun.* **47**:353-359.
 127. **Perna, N. T., G. Plunkett, III, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamosis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner.** 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **409**:529-533.
 128. **Perry, R. D. and J. D. Fetherston.** 1997. *Yersinia pestis*--etiologic agent of plague. *Clin. Microbiol. Rev.* **10**:35-66.
 129. **Pickett, C. L., E. C. Pesci, D. L. Cottle, G. Russell, A. N. Erdem, and H. Zeytin.** 1996. Prevalence of cytolethal distending toxin production in *Campylobacter jejuni* and relatedness of *Campylobacter* sp. *cdtB* gene. *Infect. Immun.* **64**:2070-2078.
 130. **Poly, F., D. Threadgill, and A. Stintzi.** 2004. Identification of *Campylobacter jejuni* ATCC 43431-specific genes by whole microbial genome comparisons. *J. Bacteriol.* **186**:4781-4795.
 131. **Porwollik, S., R. M. Wong, and M. McClelland.** 2002. Evolutionary genomics of *Salmonella*: gene acquisitions revealed by microarray analysis. *Proc. Natl. Acad. Sci. U. S. A* **99**:8956-8961.
 132. **Power, P. M. and M. P. Jennings.** 2003. The genetics of glycosylation in Gram-negative bacteria. *FEMS Microbiol. Lett.* **218**:211-222.

133. **Prendergast, M. M. and A. P. Moran.** 2000. Lipopolysaccharides in the development of the Guillain-Barre syndrome and Miller Fisher syndrome forms of acute inflammatory peripheral neuropathies. *J. Endotoxin. Res.* **6**:341-359.
134. **Prendergast, M. M., D. R. Tribble, S. Baqar, D. A. Scott, J. A. Ferris, R. I. Walker, and A. P. Moran.** 2004. In vivo phase variation and serologic response to lipooligosaccharide of *Campylobacter jejuni* in experimental human infection. *Infect. Immun.* **72**:916-922.
135. **Preston, G. M., B. Haubold, and P. B. Rainey.** 1998. Bacterial genomics and adaptation to life on plants: implications for the evolution of pathogenicity and symbiosis. *Curr. Opin. Microbiol.* **1**:589-597.
136. **Prince, R. W., C. D. Cox, and M. L. Vasil.** 1993. Coordinate regulation of siderophore and exotoxin A production: molecular cloning and sequencing of the *Pseudomonas aeruginosa* *fur* gene. *J. Bacteriol.* **175**:2589-2598.
137. **Pumbwe, L. and L. J. Piddock.** 2002. Identification and molecular characterisation of CmeB, a *Campylobacter jejuni* multidrug efflux pump. *FEMS Microbiol. Lett.* **206**:185-189.
138. **Reid, S. D., C. J. Herbelin, A. C. Bumbaugh, R. K. Selander, and T. S. Whittam.** 2000. Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* **406**:64-67.
139. **Riley, L. W., R. S. Remis, S. D. Helgerson, H. B. McGee, J. G. Wells, B. R. Davis, R. J. Hebert, E. S. Olcott, L. M. Johnson, N. T. Hargrett, P. A. Blake, and M. L. Cohen.** 1983. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N. Engl. J. Med.* **308**:681-685.
140. **Ruiz-Palacios, G. M., J. Torres, N. I. Torres, E. Escamilla, B. R. Ruiz-Palacios, and J. Tamayo.** 1983. Cholera-like enterotoxin produced by *Campylobacter jejuni*. Characterisation and clinical significance. *Lancet* **2**:250-253.
141. **Russell, R. G., M. J. Blaser, J. I. Sarmiento, and J. Fox.** 1989. Experimental *Campylobacter jejuni* infection in *Macaca nemestrina*. *Infect. Immun.* **57**:1438-1444.
142. **Russell, R. G., M. O'Donnoghue, D. C. Blake, Jr., J. Zulty, and L. J. DeTolla.** 1993. Early colonic damage and invasion of *Campylobacter jejuni* in experimentally challenged infant *Macaca mulatta*. *J. Infect. Dis.* **168**:210-215.
143. **Salzberg, S. L., A. L. Delcher, S. Kasif, and O. White.** 1998. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* **26**:544-548.
144. **Schmidt, H. and M. Hensel.** 2004. Pathogenicity islands in bacterial pathogenesis. *Clin. Microbiol. Rev.* **17**:14-56.

145. **SEBALD, M. and M. V'ERON.** 1963. [BASE DNA CONTENT AND CLASSIFICATION OF VIBRIOS]. *Ann. Inst. Pasteur (Paris)* **105**:897-910.
146. **Sodeinde, O. A., A. K. Sample, R. R. Brubaker, and J. D. Goguen.** 1988. Plasminogen activator/coagulase gene of *Yersinia pestis* is responsible for degradation of plasmid-encoded outer membrane proteins. *Infect. Immun.* **56**:2749-2752.
147. **St, M. F., C. M. Szymanski, J. Li, K. H. Chan, N. H. Khieu, S. Larocque, W. W. Wakarchuk, J. R. Brisson, and M. A. Monteiro.** 2002. The structures of the lipooligosaccharide and capsule polysaccharide of *Campylobacter jejuni* genome sequenced strain NCTC 11168. *Eur. J. Biochem.* **269**:5119-5136.
148. **Stathopoulos, C., D. R. Hendrixson, D. G. Thanassi, S. J. Hultgren, G. J. St, III, and R. Curtiss, III.** 2000. Secretion of virulence determinants by the general secretory pathway in gram-negative pathogens: an evolving story. *Microbes. Infect.* **2**:1061-1072.
149. **Stocki, S. L., L. A. Babiuk, N. A. Rawlyk, A. A. Potter, and B. J. Allan.** 2002. Identification of genomic differences between *Escherichia coli* strains pathogenic for poultry and *E. coli* K-12 MG1655 using suppression subtractive hybridization analysis. *Microb. Pathog.* **33**:289-298.
150. **Szymanski, C. M., S. M. Logan, D. Linton, and B. W. Wren.** 2003. *Campylobacter*--a tale of two protein glycosylation systems. *Trends Microbiol.* **11**:233-238.
151. **Szymanski, C. M., R. Yao, C. P. Ewing, T. J. Trust, and P. Guerry.** 1999. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol. Microbiol.* **32**:1022-1030.
152. **Tauxe, R. V., C. M. Patton, P. Edmonds, T. J. Barrett, D. J. Brenner, and P. A. Blake.** 1985. Illness associated with *Campylobacter laridis*, a newly recognized *Campylobacter* species. *J. Clin. Microbiol.* **21**:222-225.
153. **Tomb, J. F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzgerald, N. Lee, M. D. Adams, and J. C. Venter.** 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**:539-547.
154. **Touati, D.** 2000. Iron and oxidative stress in bacteria. *Arch. Biochem. Biophys.* **373**:1-6.

155. **Tran Van, N. G., R. Bourdet-Sicard, G. Dumenil, A. Blocker, and P. J. Sansonetti.** 2000. Bacterial signals and cell responses during *Shigella* entry into epithelial cells. *Cell Microbiol.* **2**:187-193.
156. **Trulzsch, K., T. Sporleder, E. I. Igwe, H. Russmann, and J. Heesemann.** 2004. Contribution of the major secreted yops of *Yersinia enterocolitica* O:8 to pathogenicity in the mouse infection model. *Infect. Immun.* **72**:5227-5234.
157. **van Vliet, A. H., K. G. Wooldridge, and J. M. Ketley.** 1998. Iron-responsive gene regulation in a campylobacter jejuni fur mutant. *J. Bacteriol.* **180**:5291-5298.
158. **Vandamme, P., B. Pot, M. Gillis, V. P. De, K. Kersters, and J. Swings.** 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* **60**:407-438.
159. **Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. bu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, F. Di, V. P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan,**

- N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, and M. Nodell. 2001. The sequence of the human genome. *Science* **291**:1304-1351.
160. **Walker, J. C. and N. K. Verma.** 2002. Identification of a putative pathogenicity island in *Shigella flexneri* using subtractive hybridisation of the *S. flexneri* and *Escherichia coli* genomes. *FEMS Microbiol. Lett.* **213**:257-264.
 161. **Walker, R. I., M. B. Caldwell, E. C. Lee, P. Guerry, T. J. Trust, and G. M. Ruiz-Palacios.** 1986. Pathophysiology of *Campylobacter* enteritis. *Microbiol. Rev.* **50**:81-94.
 162. **Wassenaar, T. M.** 1997. Toxin production by *Campylobacter* spp. *Clin. Microbiol. Rev.* **10**:466-476.
 163. **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 (Pt 6)**:1171-1175.
 164. **Wei, J., M. B. Goldberg, V. Burland, M. M. Venkatesan, W. Deng, G. Fournier, G. F. Mayhew, G. Plunkett, III, D. J. Rose, A. Darling, B. Mau, N. T. Perna, S. M. Payne, L. J. Runyen-Janecky, S. Zhou, D. C. Schwartz, and F. R. Blattner.** 2003. Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. *Infect. Immun.* **71**:2775-2786.
 165. **Weiser, J. N., A. A. Lindberg, E. J. Manning, E. J. Hansen, and E. R. Moxon.** 1989. Identification of a chromosomal locus for expression of lipopolysaccharide epitopes in *Haemophilus influenzae*. *Infect. Immun.* **57**:3045-3052.
 166. **Welch, R. A.** 1991. Pore-forming cytolysins of gram-negative bacteria. *Mol. Microbiol.* **5**:521-528.
 167. **Whitfield, C. and I. S. Roberts.** 1999. Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Mol. Microbiol.* **31**:1307-1319.
 168. **Wilderman, P. J., N. A. Sowa, D. J. FitzGerald, P. C. FitzGerald, S. Gottesman, U. A. Ochsner, and M. L. Vasil.** 2004. Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. *Proc. Natl. Acad. Sci. U. S. A* **101**:9792-9797.

169. **Wren, B. W.** 2003. The yersiniae--a model genus to study the rapid evolution of bacterial pathogens. *Nat. Rev. Microbiol.* **1**:55-64.
170. **Young, N. M., J. R. Brisson, J. Kelly, D. C. Watson, L. Tessier, P. H. Lanthier, H. C. Jarrell, N. Cadotte, M. F. St, E. Aberg, and C. M. Szymanski.** 2002. Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. *J. Biol. Chem.* **277**:42530-42539.
171. **Yuki, N., K. Susuki, M. Koga, Y. Nishimoto, M. Odaka, K. Hirata, K. Taguchi, T. Miyatake, K. Furukawa, T. Kobata, and M. Yamada.** 2004. Carbohydrate mimicry between human ganglioside GM1 and *Campylobacter jejuni* lipooligosaccharide causes Guillain-Barre syndrome. *Proc. Natl. Acad. Sci. U. S. A* **101**:11404-11409.
172. **Zaigler, A., S. C. Schuster, and J. Soppa.** 2003. Construction and usage of a onefold-coverage shotgun DNA microarray to characterize the metabolism of the archaeon *Haloferax volcanii*. *Mol. Microbiol.* **48**:1089-1105.
173. **Zhang, L., U. Srinivasan, C. F. Marrs, D. Ghosh, J. R. Gilsdorf, and B. Foxman.** 2004. Library on a slide for bacterial comparative genomics. *BMC. Microbiol.* **4**:12.

Liste des mots clefs

- Campylobacter jejuni*
- pathogénicité bactérienne
- Puces à ADN
- génomique
- Ferric Uptake Regulator

Key words

- Campylobacter jejuni*
- Bacterial pathogenesis
- Microarray
- Genomotyping
- Ferric Uptake Regulator

Etude de la diversité génétique de l'espèce *Campylobacter jejuni* par l'utilisation de puces à ADN

**Study of *Campylobacter jejuni* genomic
diversity by microarray**

Résumé

Cette étude a conduit à l'identification des séquences nucléotidiques uniques de deux souches non séquencées de *Campylobacter jejuni*, bactérie à l'origine d'affections entériques chez l'homme par comparaison avec la souche séquencée NCTC 11168. Nous avons étudié les souches *C. jejuni* TGH 9011 et 81-176, toutes deux présentant des phénotypes de virulence plus accentués que la souche séquencée. Nous avons mis au point une puce à ADN de type « shotgun », générée par dépôt de plasmides représentatifs d'une banque génomique totale de chacune de ces deux souches. L'étude a permis l'identification de 130 ORFs spécifiques par rapport à la souche-type pour TGH 9011 et 86 ORFs ainsi que la séquence du plasmide pTet pour 81-176. Les gènes identifiés chez ces souches sont impliqués principalement dans des mécanismes démontrés comme étant liés à la pathogenèse de *C. jejuni*. La puce à ADN de TGH 9011 a également été utilisée de façon préliminaire à l'identification des gènes régulés par la protéine Fur (Ferric Uptake Regulator). Ces travaux démontrent l'efficacité de l'utilisation de puce à ADN de type « shotgun » pour l'identification de la diversité génétique bactérienne.

Abstract

This study represents the identification of unique sequences of two non-sequenced, human-enteropathogenic *Campylobacter jejuni* strains, , by comparison with the sequenced strain NCTC 11168. We choose to study TGH 9011 and 81-176 strains for

their pathogenicity characters. We constructed a shotgun microarray made by the printing of plasmids issued from a whole representative genomic library of the two studied strains. After analysis, we were able to identify 130 ORFs specific to TGH 9011 compare to NCTC 11168 and 86 ORFs and the sequence of the plasmid pTet of 81-176. Identified genes are mostly involved in demonstrated pathogenesis mechanisms of *C. jejuni*. The microarray of TGH 9011 has been also used for the study of genes regulated by the Ferric Uptake regulator. In summary, these studies demonstrate the power of the use of shotgun microarray for bacterial genomic comparisons.