AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Nancy-Université
Institut National Polytechnique de Lorraine

Ecole :
Ecole Nationale Supérieure d'Agronomie et des Industries Alimentaires
Ecole doctorale :
Sciences et Ingénierie des Ressources Procédés Produits et Environnement

Thèse

présentée pour l'obtention du titre de

Docteur de l'Institut National Polytechnique de Lorraine

en Sciences Agronomiques

par **Delphine DUFOUR**

Recherche et caractérisation de déterminants génétiques permettant l'adaptation d'une souche d'*Escherichia coli* à la mamelle bovine

Soutenue publiquement le 24 octobre 2008 devant la commission d'examen :

Membres du jury :

Rapporteurs : **C. Le Bouguènec,** Chef de Laboratoire, Institut Pasteur, Paris
C. Burvenich, Professeur, Faculté de Médecine Vétérinaire, Gand, Belgique

Examinateurs : **P. Germon,** Chargé de recherche, Institut National de la Recherche Agronomique, Tours
Y. Le Roux, Maître de Conférence, Institut National Polytechnique de Lorraine, Nancy
A. Dary, Professeur, Université Henri Poincaré, Nancy 1

Unité de Recherche sur l'Animal et les Fonctionnalités des Produits Animaux
Faculté des Sciences – 54506 Vandoeuvre les Nancy

Remerciements

Merci Messieurs les Professeurs Jean-Luc Gaillard et Guido Rychen de m'avoir accueillie au sein de votre laboratoire et de la confiance que vous m'avez témoignée pour la réalisation de ce travail.

Merci Madame Annie Dary et Monsieur Yves Le Roux pour tous les savoirs que vous m'avez transmis, pour le suivi rigoureux et attentif de mes travaux et pour votre gentillesse de tous les instants.

Merci Madame Chantal Le Bouguènec et Monsieur Christian Burvenich d'avoir accepté d'être rapporteurs de ce travail.

Merci Monsieur Pierre Germon pour votre accueil lors de mon séjour à Tours, d'avoir accepté de juger ce travail et surtout pour le suivi, les conseils et l'aide considérable que vous m'avez si gentiment accordés pour sa réalisation.

Merci Monsieur Gérard Guédon pour vos conseils qui m'ont aidé à mieux comprendre et interpréter mes résultats.

Merci Emilie pour ta précieuse aide technique de haute qualité et nos bavardages de tout et de rien.

Merci Nawara et Wessam pour votre gentillesse, votre bonne humeur et les moments de franches rigolades que nous avons eu ensemble. Je vous souhaite une belle réussite pour la suite et d'être heureux comme vous le désirez.

Merci Monsieur Gérard Humbert de m'avoir accueillie pour la toute première fois au laboratoire et pour votre gentillesse.

Merci Monsieur Alain Driou pour votre gentillesse.

Merci Madame Catherine Corbier de la confiance que vous me témoignez pour m'avoir accordé des postes d'ATER à l'IUT.

Merci Chantal, Clarisse, Katy, Madame Lagrange, Laurent, Jean-Michel, Frank, Oukni, Maira, Aurélie, Emeline, Céline et Faiza.

Papa, je te dédie ce travail, maman, je t'embrasse très fort, et je vous remercie tous les deux d'être toujours là pour moi et avec moi car j'en ai toujours eu besoin et j'en aurai toujours besoin.

Mon amour et mon petit bébé, je vous aime et vous fais mille bisous tous doux.

Sommaire

Sommaire **5**

Abréviations **11**

Liste des figures **15**

Liste des tableaux **19**

Introduction **23**

Avant-propos **25**

1. Les éléments génétiques bactériens mobiles ou mobilisables par transfert génétique horizontal **28**

1.1. Présentation de différents types *29*
1.1.1. Les plasmides 29
1.1.2. Les transposons 31
1.1.3. Les intégrons 33
1.1.4. Les prophages 36
1.1.5. Les îlots génomiques 37

1.2. Mécanismes responsables de leur mobilité intercellulaire *39*
1.2.1. La transformation 40
1.2.2. La conjugaison 43
1.2.3. La transduction 47

1.3. Mécanismes responsables de leur insertion chromosomique *47*
1.3.1. La recombinaison homologue 48
1.3.2. La recombinaison site-spécifique 48
1.3.3. La recombinaison transpositionnelle 52

2. Les apports fonctionnels des éléments génétiques mobiles ou mobilisables bactériens **54**

2.1. La symbiose *55*

2.2. Le métabolisme *57*
2.2.1. Métabolisme des composés organiques “naturels” 57
2.2.2. Métabolisme des composés organiques “ non naturels” 58
2.2.3. Métabolisme des métaux lourds 60

2.3. La résistance aux antibiotiques *61*
2.3.1. L’élément SXT 62
2.3.2. L’élément SGI1 63
2.3.3. L’élément SCCmec 65
2.3.4. L’élément pNR1 66

2.4. La virulence *67*
2.4.1. Les plasmides porteurs de fonctions de virulence 68
2.4.2. Les prophages porteurs de fonctions de virulence 68
2.4.3. Les îlots génomiques porteurs de fonctions de virulence 70

2.5. Conclusion *74*

3. Les éléments génétiques mobiles ou mobilisables et leur impact sur la diversification de l’espèce *Escherichia coli* **75**

3.1. Situation de l’espèce E. coli *au sein de la famille* Enterobacteriaceae *76*
3.1.1. Proximité des genres *Escherichia*, *Shigella* et *Salmonella* 77
3.1.2. *Escherichia coli* au sein du genre *Escherichia* 78

3.2. Etude comparative des génomes de différentes souches de l’espèce E. coli *79*
3.2.1. Génome de la souche *E. coli* K12 MG1655 81
3.2.2. Génome de la souche *E. coli* O157:H7 Sakaï 82
3.2.3. Génome de la souche *E. coli* CFT073 85
3.2.4. Génome de la souche *E. coli* APEC01 87
3.2.5. Etude partielle des génomes des souches *E. coli* A0 34/86 et *E. coli* Nissle 1917 88
3.2.6. Conclusion 89

3.3. Description des différents pathotypes de l’espèce E. coli *89*

3.3.1. Pathotypes intestinaux ... 89
3.3.2. Pathotypes extra-intestinaux ... 91
3.3.3. Relation entre pathotypes et groupes phylogénétiques ... 92

4. Objectif du travail : caractérisation d'une souche *Escherichia coli* inductrice de mammites bovines ... 94

4.1. Présentation des mammites bovines ... 94
4.1.1. Définition ... 94
4.1.2. Agents inducteurs ... 95
4.1.3. Conséquences ... 96

4.2. Connaissances actuelles des souches E. coli *inductrices de mammites bovines ... 97*
4.2.1. Variabilités génotypique et pathologique ... 97
4.2.2. Constance du lipopolysaccharide ... 99
4.2.3. Apparition de souches persistantes et invasives ... 100

4.3. Problématique abordée ... 101

Résultats ... 103

1. Etude de la croissance d'*E. coli* P4 dans un lait issu de mammite bovine ... 105

2. Typage moléculaire d'*E. coli* P4 ... 107

2.1. Détermination du groupe phylogénétique et séquençage d'un ITS 16S-23S ... 107

2.2. Séquençage de multiples loci (MLST) ... 108

3. Etude du rôle d'*E. coli* P4 dans la caséinolyse du lait observée lors d'une mammite ... 109

3.1. Mise en évidence d'une sécrétion de protéases extracellulaires par la technique de zymographie ... 110

*3.2. Evaluation de l'activité caséinolytique d'*E. coli *P4 et des quatre protéases extracellulaires mises en évidence ... 113*

3.3. Essai de purification des protéases extracellulaires mises en évidence ... 116

3.4. Publication ... 120

4. Recherche de séquences nucléotidiques particulières à *E. coli* P4 ... 120

4.1. Recherche de facteurs de virulence ... 132

4.2. Recherche d'îlots génomiques ... 134
4.2.1. Criblage des loci d'ARNt ... 134
4.2.2. Caractérisation partielle des régions aval des loci *asnT*, *leuX*, *pheV* et *thrW* grâce à une technique de PCR arbitraire ... 136
4.2.2.1. Cas du locus *asnT* ... 138
4.2.2.2. Cas du locus *leuX* ... 138
4.2.2.3. Cas du locus *pheV* ... 140
4.2.2.4. Cas du locus *thrW* ... 142
4.2.3. Caractérisation partielle de la région aval au locus *serU* grâce au séquençage partiel d'un cosmide recombinant ... 142
4.2.4. Caractérisation complète de la région aval au locus *argW* ... 143
4.2.4.1. Contenu en ORFs ... 144
4.2.4.2. Analyse de l'ORF 4-*argW* ... 148
4.2.5. Conclusion ... 149

4.3. Affiche ... 150

4.4. Publication ... 151

Discussion et perspectives ... 181

1. *E. coli* P4 appartient au groupe phylogénétique A de l'espèce *E. coli* ... 185

2. Implication d'*E. coli* P4 dans la caséinolyse observée lors d'une mammite bovine ... 188

2.1. Mise en doute du rôle caséinolytique des quatre protéases extracellulaires mises en évidence ... 188

2.2. Rôle des protéases extracellulaires sécrétées par les bactéries E. coli *pathogènes ... 190*

3. Recherche, au sein du génome d'*E. coli* P4, de gènes codant des facteurs de virulence connus chez les autres pathotypes de l'espèce *E. coli* **192**

3.1. Pertinence des gènes de virulence recherchés *192*
3.1.1. Gènes bactériens de résistance aux défenses immunitaires hôtes 192
3.1.1.1. Gènes *neuA-neuC* 194
3.1.1.2. Gène *traT* 194
3.1.1.3. Gène *iucD* 195
3.1.2. Gènes bactériens d'adhérence et d'invasion aux cellules hôtes 197

3.2. Hypothèses expliquant les résultats obtenus *200*

4. *E. coli* P4 renferme dans son génome plusieurs îlots génomiques potentiels à proximité de différents loci d'ARNt **203**

4.1. Analyse des loci d'ARNt altérés partiellement caractérisés *203*
4.1.1. Cas du locus *asnT* 203
4.1.2. Cas du locus *leuX* 204
4.1.3. Cas du locus *pheV* 205
4.1.4. Cas du locus *thrW* 207
4.1.5. Cas du locus *serU* 207

4.2. Analyse de l'îlot génomique argW *entièrement caractérisé* *208*
4.2.1. Caractéristiques générales 208
4.2.2. Hypothèses sur le rôle de la protéine codée par l'ORF 4-*argW* 213

4.3. Conclusion *216*

Références bibliographiques **217**

Résumé **245**

Annexe **249**

Abréviations

Da : Dalton
DAEC : *Escherichia coli* à adhésion diffuse (Diffusely Adherent *Escherichia coli*)
Db : Double brin
EAEC : *Escherichia coli* entéroagrégative (Enteroaggregative *Escherichia coli*)
EHEC : *Escherichia coli* entérohémolytique (Enterohemorrhagic *Escherichia coli*)
EIEC : *Escherichia coli* entéroinvasive (Enteroinvasive *Escherichia coli*)
EPEC : *Escherichia coli* entéropathogène (Enteropathogenic *Escherichia coli*)
ERIC : Séquence consensus intergénique répétée des entérobactéries (Enterobacterial Repetitive Intergenic Consensus)
ETEC : *Escherichia coli* entérotoxinogène (Enterotoxigenic *Escherichia coli*)
ExPEC : *Escherichia coli* pathogène extra-intestinale (Extraintestinal Pathogenic *Escherichia coli*)
IS : Séquence d'insertion (Insertion Sequence)
ITS : Séquence interne transcrite (Internal Transcribed Spacer)
kb : Kilobases
kDa : Kilodaltons
LDR : Longue répétition directe (Long Direct Repeat)
LPS : Lipopolysaccharide
Mb : Mégabases
MLST : Typage par séquençage de multiples loci (Multi Locus Sequence Typing)
MNEC : *Escherichia coli* à l'origine de méningites néonatales (Neonatal Meningitis *Escherichia coli*)
MPEC : *Escherichia coli* mammopathogène (Mammopathogenic *Escherichia coli*)
ORF : Phase ouverte de lecture (Open Reading Frame)
PAGE : Electrophorèse sur gel de polyacrylamide (Polyacrylamide Gel Electrophoresis)
PCR : Réaction de polymerisation en chaîne (Polymerase Chain Reaction)
PFGE : Electrophorèse en champs pulsés (Pulsed Field Gel Electrophoresis)
RBS : Site de fixation des ribosomes (Ribosome Binding Site)
REP : Palindrome répété extragénique (Repetitive Extragenic Palindromic)
RSA : Séquence répétée A (Repetitive Sequence A)
Sb : Simple brin
SDS : Dodécyl sulphate de sodium (Sodium Dodecyl Sulfate)
SPATE : Protéase à serine capable de s'auto-transporter et spécifique des Entérobactéries (Serin Protease Autotransporter of *Enterobacteriaceae)*
ufc : Unité formant des colonies
UPEC : *Escherichia coli* uropathogène (Uropathogenic *Escherichia coli*)

Genres bactériens :

E. : *Escherichia*
P. : *Pseudomonas*
S. : *Streptococcus*
Sh. : *Shigella*
St. : *Staphylococcus*

Liste des figures

Figure 1……p.31
Figure 2……p.34
Figure 3……p.41
Figure 4……p.41
Figure 5……p.44
Figure 6……p.45
Figure 7……p.46
Figure 8……p.50
Figure 9……p.51
Figure 10……p.53
Figure 11……p.56
Figure 12……p.59
Figure 13……p.60
Figure 14……p.61
Figure 15……p.63
Figure 16……p.64
Figure 17……p.65
Figure 18……p.67
Figure 19……p.93
Figure 20……p.106
Figure 21……p.111
Figure 22……p.112
Figure 23……p.113
Figure 24……p.114
Figure 25……p.115
Figure 26……p.116
Figure 27……p.117
Figure 28……p.119
Figure 29……p.135
Figure 30……p.137
Figure 31……p.139
Figure 32……p.141
Figure 33……p.143
Figure 34……p.145
Figure 35……p.146
Figure 36……p.147
Figure 37……p.149
Figure 38……p.186
Figure 39……p.187
Figure 40……p.193
Figure 41……p.196
Figure 42……p.201
Figure 43……p.209
Figure 44……p.211
Figure 45……p.212

Liste des tableaux

Tableau 1…… …….p.55
Tableau 2…… …….p.56
Tableau 3…… …….p.59
Tableau 4…… …….p.66
Tableau 5…… …….p.69
Tableau 6…… …….p.71
Tableau 7…… …….p.76
Tableau 8…… …….p.79
Tableau 9…… …….p.79
Tableau 10…… …….p.80
Tableau 11…… …….p.82
Tableau 12…… …….p.84
Tableau 13…… …….p.87
Tableau 14…… …….p.93
Tableau 15…… …….p.98
Tableau 16…… …….p.105
Tableau 17…… …….p.133
Tableau 18…… …….p.135
Tableau 19…… …….p.136
Tableau 20…… …….p.138
Tableau 21…… …….p.139
Tableau 22…… …….p.141
Tableau 23…… …….p.142
Tableau 24…… …….p.143
Tableau 25…… …….p.145
Tableau 26…… …….p.191
Tableau 27…… …….p.198
Tableau 28…… …….p.201
Tableau 29…… …….p.202
Tableau 30…… …….p.209

Introduction

Avant-propos

Notre planète se caractérise par une grande biodiversité : 1,75 millions d'espèces différentes sont actuellement inventoriées par les taxonomistes et entre 6 et 14 millions seraient encore à décrire. Depuis toujours, cette biodiversité, qui se reflète au niveau des structures cellulaires, des morphologies, des habitats, des modes de croissance, des modes de reproduction, des ressources énergétiques et des voies métaboliques, gouverne le travail des biologistes qui, l'observant et la découvrant sans cesse, s'efforcent de l'ordonner, de la classer pour mieux la comprendre. Aussi, tout en effectuant de manière progressive cette classification du vivant, ces mêmes biologistes se sont rendus compte de son évidente unité. Tous les organismes vivants sont en effet faits de la même chimie, ne renfermant qu'un petit nombre de macromolécules de types différents (les protéines, les acides nucléiques et les lipides) ; tous utilisent la même manière de coder les protéines et la même machinerie de synthèse protéique ; et tous sont séparés du milieu extérieur par une membrane et sont en relation avec lui par des échanges de matière liquide, solide ou gazeuse. Ceci suggère qu'ils dériveraient tous d'un ancêtre commun, nommé dans le milieu des années 90 LUCA (Last Universal Common Ancestor) (Forterre *et al*., 2005). Comment expliquer alors l'apparition d'une telle diversité à partir d'une seule et même forme de vie ancestrale ? L'évolution, décrivant les modifications des êtres vivants au cours du temps, prend une part importante ici dans la compréhension de ce phénomène. Depuis les travaux de Lamarck (XVIII^{ème} siècle) et Darwin (XIX^{ème} siècle), l'idée de créationnisme et de fixisme, issue de l'antiquité et confortée par le dogme religieux, a fait progressivement place à plusieurs courants évolutionnistes qui ont été au centre de nombreuses controverses philosophiques, politiques et religieuses, car ils remettaient en cause l'origine de l'humanité. Au milieu du XIX^{ème} siècle, Darwin propose que la sélection naturelle est le principal moteur à l'évolution des espèces. Il suggère ainsi que les individus subissent des variations aléatoires et que seules celles qui se révèlent favorables à leur survie, car elles améliorent l'adaptation des organismes les possédant à leur environnement, sont sélectionnées par le milieu et transmises, tandis que celles qui se révèlent défavorables sont éliminées. Au XX^{ème} siècle, le développement considérable de la génétique a apporté un fondement moléculaire à la théorie de Darwin, établissant que l'évolution résulte de la modification des gènes des êtres vivants. Est née la théorie actuelle de l'évolution, « le Néodarwinisme », selon laquelle de nouvelles formes de vie apparaissent après l'avènement de modifications génétiques, aléatoires et porteuses d'une valeur sélective (fitness) favorable ou défavorable, soumises ensuite à la sélection naturelle, c'est-à-dire à la pression de l'environnement. Ces modifications génétiques sources de la diversité, de l'adaptation et de l'évolution des organismes vivants sont appelées mutations. Trois catégories de mutations peuvent être distinguées. La première, qualifiée de mutations géniques et considérée comme le processus

évolutif le plus lent et le plus graduel, implique des changements locaux dans la séquence d'un gène telles que des substitutions, des insertions ou des délétions d'un ou plusieurs nucléotides. La deuxième, qualifiée de mutations chromosomiques, crée des changements de structure d'un chromosome, incluant des duplications, des amplifications, des inversions, des délétions ou des insertions de segments chromosomiques. Enfin, la troisième, qualifiée de mutations génomiques, repose sur des changements dans le nombre de chromosomes.

L'évolution et donc la diversité des organismes vivants appartenant au règne des bactéries sont particulièrement intéressantes et importantes à étudier pour au moins trois raisons. La première est le rôle considérable de ces individus dans le recyclage des éléments essentiels de la biosphère et de l'atmosphère. La Terre est en effet un système clos pourvu de quantités limitées de carbone, d'oxygène et d'azote disponibles pour toutes les formes de vie. Ces éléments essentiels doivent être convertis d'une forme à une autre et partagés par tous les organismes vivants. De par leur faculté à utiliser différentes ressources énergétiques, leurs riches et diverses voies métaboliques et leur présence dans tous les types de biotopes rencontrés sur la Terre, ce sont les bactéries qui jouent ce rôle crucial de transformation et de recyclage de ces éléments. La deuxième raison est l'importance démographique et la diversité moléculaire de ces organismes. En effet, des estimations indiquent que 12.10^{28} cellules bactériennes vivraient dans les habitats aquatiques, 355.10^{28} cellules dans les souterrains océaniques, 26.10^{28} cellules dans les sols et 25 à 250.10^{28} cellules dans les souterrains terrestres (Whitman *et al.*, 1998). Environ 4000 espèces bactériennes sont décrites à l'heure actuelle, mais des données moléculaires basées sur les ARNr 16S de bactéries non cultivables suggèrent que ce chiffre représenterait seulement 1 % du nombre total d'espèces bactériennes réellement présentes sur la Terre (Rosselló-Mora et Amann, 2001). Enfin, la troisième raison est que les bactéries, comme tous les microorganismes en général (procaryotes ou eucaryotes), peuvent évoluer très rapidement, étant donné leur temps de génération généralement court comparé à celui d'autres organismes (au maximum quelques dizaines d'heures pour beaucoup de bactéries, contre quelques dizaines d'années pour l'homme par exemple). L'homme peut ainsi suivre l'évolution des bactéries de ses propres yeux et de son vivant. Les mécanismes reconnus comme sources les plus actives et les plus rapides de la diversité génétique, donc de l'évolution, des bactéries sont les mutations chromosomiques de type insertion, reposant sur des échanges entre individus, pas forcément de la même espèce, de matériels génétiques mobiles ou mobilisables et aboutissant à l'acquisition de nouvelles informations génétiques (Lawrence et Roth, 1999). Ces échanges sont également appelés transferts génétiques horizontaux

La description des éléments génétiques bactériens mobiles ou mobilisables par transfert génétique horizontal (leurs structures, leurs voies de déplacement, leurs modes d'insertion chromosomique), de leurs divers apports évolutifs recensés dans le monde bactérien et de leur impact

tout particulier au sein de l'espèce *Escherichia coli* va permettre d'introduire l'objectif du travail présenté dans ce mémoire. La thématique étudiée concerne principalement l'adaptation d'une souche *E. coli* à la glande mammaire bovine. Cette adaptation se traduit par une capacité à y induire des inflammations appelées mammites, considérées comme la pathologie la plus fréquente chez les vaches laitières à travers le monde. Les souches *E. coli* responsables de mammites bovines et leur mode de virulence restent, cependant, encore à l'heure actuelle peu étudiés.

1. Les éléments génétiques bactériens mobiles ou mobilisables par transfert génétique horizontal

Au sein des génomes bactériens, deux types d'entités peuvent être distingués : le (ou les) chromosome en lui-même, perçu comme entité constante, permanente, au sein d'une espèce ; et un ensemble d'éléments extra-chromosomiques ou présents sur le chromosome, identifiés au contraire comme entités instables et variables, à la fois entre espèces différentes et au sein d'une même espèce.

Le chromosome est chez la plupart des bactéries unique et circulaire. Il existe toutefois de rares exemples de bactéries où deux chromosomes sont présents (comme chez *Vibrio cholerae*) (Heidelberg *et al*., 2000) et où le chromosome est linéaire (comme chez les Spirochètes et les *Streptomyces*) (Hinnebusch et Tilly, 1993). La taille du chromosome peut être très variable selon les espèces de bactéries étudiées. Les plus petits sont ceux des parasites intracellulaires (comme celui de *Mycoplasma genitalium*, parasite du tractus génital des primates, de 580 kb) (Fraser *et al*., 1995) ou des symbiotes intracellulaires (comme celui de *Candidatus Carsonella ruddii*, symbiote d'insectes, de 160 kb) (Nakabachi *et al*., 2006). Les plus grands appartiennent au contraire aux bactéries présentant des différenciations cellulaires (comme celui de *Sorrangium cellulosum*, bactérie du sol, de 12200 kb) (Pradella *et al*., 2002). Une dernière caractéristique des chromosomes bactériens est leur pourcentage en paires de bases guanine et cytosine, également très variable entre les espèces : une gamme allant de 16,6 % (pour le chromosome de *Candidatus Carsonella ruddii*) (Nakabachi *et al*., 2006) à 74,9 % (pour le chromosome de *Anaeromyxobacter dehalogenans* 2CP-C, myxobactérie anaérobie facultative) (non publié) peut ainsi être observée.

Les éléments génétiques qualifiés d'instables (car pouvant être perdu à des fréquences plus ou moins importantes) et de variables (car de diverses natures, structures et tailles) sont soit cryptiques, soit porteurs de fonctions enzymatiques nombreuses et variées. Selon leur nature, ils sont capables pour certains d'entre eux de se déplacer du cytoplasme d'une cellule vivante à celui d'une autre cellule et pour d'autres de se déplacer au sein d'une même molécule d'ADN et/ou entre deux molécules d'ADN différentes portées ou non par la même cellule. S'ils portent leurs propres fonctions de mobilité ou s'ils en sont dépourvus et utilisent les fonctions de mobilité d'autres éléments, cette mobilité qualifie ces éléments génétiques instables et variables d'éléments mobiles ou mobilisables, respectivement, capables de se transférer de manière horizontale entre bactéries, c'est-à-dire pas uniquement de la cellule mère aux cellules filles mais également entre cellules n'ayant aucun lien de parenté. Ces éléments mobiles ou mobilisables ne sont pas à négliger au sein des génomes bactériens : ils constitueraient par exemple 15 % du génome de l'espèce *E. coli* (Lawrence et Ochman, 1997) et 25 % de celui de l'espèce *Thermotoga maritima* (Nelson *et al*., 1999).

1.1. Présentation de différents types

Cinq types d'éléments génétiques mobiles ou mobilisables vont être ici décrits d'un point de vue structural. Cette description ne devra pas être considérée comme finie, étant donné que de nouveaux types sont continuellement découverts par le séquençage de plus en plus massif des génomes bactériens (550 séquences complètes sont actuellement disponibles).

1.1.1. Les plasmides

Les plasmides sont très largement répandus dans le monde bactérien puisqu'ils ont pu être identifiés dans pratiquement toutes les espèces où ils ont été recherchés. Ils sont habituellement perdus à une fréquence de 10^{-6} à 10^{-3}, fréquence un peu plus élevée que celle des mutations inactivant les gènes, qui est souvent de l'ordre de 10^{-9} à 10^{-7} par génération. Tous présentent deux caractéristiques communes. La première est d'être des molécules d'ADN double brin. La seconde est de contrôler leur propre réplication et donc leur nombre de copies présentes dans la cellule : généralement faible, de 1 à 5, pour les plasmides de grande taille (supérieure à 70 kb) et élevé, de 20 à 100, pour ceux de petite taille (inférieure à 70 kb). Certains plasmides peuvent présenter une troisième caractéristique, une fonction de partition nommée *par*, contrôlant leur transmission à la descendance lors de la division cellulaire, généralement selon un mécanisme d'équipartition (égale répartition du nombre de plasmides dans les cellules filles). Une grande variabilité de systèmes génétiques intervenant dans la régulation de la réplication et dans la répartition lors de la division cellulaire des plasmides peut être observée. Ceci conduit à classer les plasmides en groupe d'incompatibilité (Novick, 1987). Les plasmides d'un même groupe possèdent des systèmes apparentés de régulation de la réplication et/ou des systèmes apparentés de répartition des copies dans les cellules filles (lorsqu'ils existent) et ne peuvent coexister de façon stable dans la même bactérie. Ainsi, près de trente groupes différents d'incompatibilité sont identifiés chez les entérobactéries et sept chez les staphylocoques (Datta, 1979).

Les autres propriétés des plasmides peuvent être très variables. Leur spectre d'hôte peut être large ou étroit selon la nature des gènes qu'ils renferment. Leur structure est dans la plupart des cas circulaire. Des exemples de linéarité ont été cependant identifiés chez des bactéries du genre *Streptomyces*, comme le plasmide pSLA2-L de la souche *Streptomyces rochei* 7434AN4, productrice d'antibiotiques (telles que la lankacidine et la lankamycine) (Mochizuki *et al*., 2003) et chez des bactéries du genre *Borrelia*, comme le plasmide de 44 kb de la souche *Borrelia duttonii* Ly, agent pathogène à l'origine de fièvres récurrentes chez les vertébrés (Tabuchi *et al*., 2002). Habituellement présent sous forme extra-chromosomique, un ADN plasmidique peut s'intégrer à un ADN chromosomique à la suite d'un évènement de recombinaison homologue entre deux séquences identiques ou de recombinaison site-spécifique entre deux séquences presque identiques, portées l'une

par le chromosome et l'autre par le plasmide (l'intégration par recombinaison transpositionnelle reste plus accidentelle). Un tel plasmide intégré dans le chromosome est appelé épisome et se réplique avec le chromosome. La taille des plasmides peut aller de 0,85 kb, comme le plasmide pRKU1 de la bactérie thermophile anaérobie *Thermotoga petrophila* RKU-1 (Nesbo *et al*., 2006), à 1,68 Mb, comme le plasmide pSymB de la bactérie *Sinorhizobium meliloti* 1021, symbiote des légumes (Finan *et al*., 2001). Les plasmides peuvent ainsi représenter une part importante du génome de l'hôte. *Sinorhizobium meliloti* 1020 en est un exemple type : son chromosome présente une taille de 3,65 Mb, soit 54,6 % de son génome et ses deux plasmides pSymA et pSymB des tailles de 1,35 Mb et 1,68 Mb, soit 20,2 % et 25,1 %, respectivement, de son génome (Capela *et al*., 2001). Les bactéries peuvent posséder plusieurs plasmides de nature différente à la fois. Ainsi, chez la plus utilisée et la mieux connue des bactéries lactiques utilisées dans les fermentations, *Lactococcus lactis*, toutes les souches présentent de deux à quatorze plasmides différents, le plus souvent entre quatre et sept pouvant aller jusqu'à 150 kb (Mills *et al*., 2006).

En tant qu'éléments génétiques mobiles, les plasmides peuvent être disséminés par différents modes de transfert horizontaux (dont les mécanismes seront expliqués en détail dans le paragraphe 1.2 de ce chapitre) : soit par transduction ou transformation et ceci de manière indépendante des fonctions qu'ils codent ; soit par conjugaison et ceci en relation avec les fonctions qu'ils codent. Dans ce dernier cas, il est nécessaire de distinguer deux catégories de plasmides : les plasmides conjugatifs et les plasmides non conjugatifs. Seuls les plasmides dits conjugatifs portent des fonctions assurant une mobilité par conjugaison. Les non conjugatifs en sont dépourvus mais peuvent toutefois être transférés par ce phénomène s'ils coexistent dans une même cellule avec un plasmide conjugatif, empruntant ainsi le système conjugatif de ce dernier (phénomène de mobilisation ou de co-transfert). Des plasmides conjugatifs sont identifiés chez des bactéries aussi bien à Gram négatif qu'à Gram positif. Le premier plasmide conjugatif découvert est le plasmide F de la bactérie à Gram négatif *E. coli* K12 (Willetts et Skurray, 1987). Considéré comme modèle pour la compréhension de la conjugaison, ce plasmide a une taille de 94,5 kb. Ses fonctions de mobilité sont codées au niveau d'un unique opéron, encore appelé région de transfert, d'une taille de 34 kb environ et composé de 39 gènes : 26 gènes codent des protéines Tra (TraA à TraZ), 10 gènes codent des protéines Trb (TrbA à TrbJ), un gène code une protéine FinO et deux gènes codent des ARNs antisens FinP et ArtA (figure 1) (Frost *et al*., 1994). Cinq fonctions essentielles sont assurées par ces 39 acteurs : le métabolisme du transfert, la régulation du transfert, la synthèse et l'assemblage des pili, la stabilisation de l'agrégation (agrégation formée pendant le phénomène de conjugaison entre une bactérie porteuse d'un plasmide conjugatif et une bactérie qui en est dépourvue) et l'exclusion de surface (empêchant deux bactéries porteuses de plasmides conjugatifs d'établir un contact entre elles). Le transfert débute au niveau d'une origine de transfert appelée *ori*T, à partir de laquelle le désenroulement simple brin du plasmide commence avant le transfert vers la bactérie réceptrice.

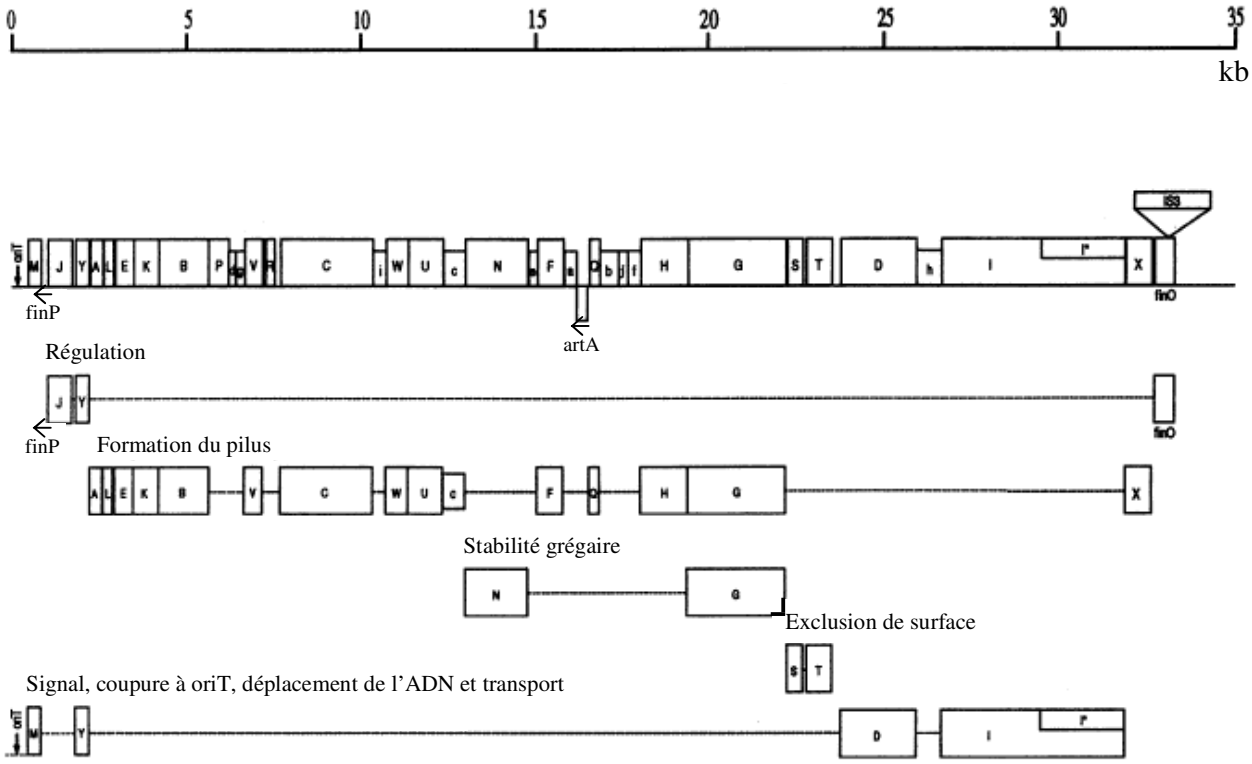


Figure 1 : Carte du module de transfert du plasmide F, d'après Frost *et al.*, 1994.

La première ligne correspond à une échelle en kilobases. La seconde ligne indique les gènes du module de transfert du plasmide F. oriT : origine de transfert ; lettres en majuscules : gènes *tra* ; lettres en minuscules : gènes *trb* ; flèches finP et artA : ARN antisens codés par ces deux gènes. Les cinq dernières lignes indiquent les fonctions identifiées pour les différents gènes du module.

1.1.2. Les transposons

Les transposons, ou éléments génétiques transposables, sont des séquences d'ADN double brin linéaires, n'apparaissant jamais à l'état libre, n'ayant pas d'existence autonome stable et devant s'intégrer dans un réplicon pour être dupliqués avec lui et exprimés. Ils représentent une famille très hétérogène du point de vue de leur structure et quatre classes d'éléments transposables peuvent être ainsi différenciées.

Les séquences d'insertion, ou ISs, sont les éléments transposables les plus simples, retrouvés dans presque tous les génomes bactériens (et ceci en nombreuses copies et de différentes natures), de petite taille (de 750 à 2500 pb) et présentant une organisation très compacte. Ils ne codent généralement que les informations nécessaires à leur mobilité. Ces dernières comportent des séquences d'ADN répétées inversées délimitant les extrémités des éléments et une enzyme, la transposase, reconnaissant ces extrémités et catalysant la mobilité des éléments (Mahillon et Chandler, 1998).

Les répétitions terminales inversées, ou IRs, sont constituées de 10 à 40 pb formant deux domaines fonctionnels, les domaines I et II. Le domaine II, positionné à l'intérieur des IRs, lie la transposase ; le domaine I, incluant les 2 ou 3 paires de bases terminales des IRs, est impliqué dans les réactions de clivage et de transfert de brin lors de la mobilisation des ISs. L'IR située en amont du gène codant la transposase comporte le promoteur permettant l'expression de cette dernière et est nommée par convention IR gauche (ou IRL) ; celle située en aval du gène est appelée IR droite (ou IRR). Quelques ISs ne présentent aucune séquence répétée inversée terminale, comme l'IS*91*, l'IS*110* et l'IS*200/605*, identifiées dans les génomes des souches *E. coli* EC185, *Streptomyces coelicolor* A3 et *Salmonella typhimurium* TR6238 et LT2, respectivement.

La transposase est codée généralement par une seule ORF, rarement par deux ORF chevauchantes. Dans ce dernier cas, un évènement de décodage non conventionnel, le décalage de phase en -1, permet tout de même d'associer les deux ORFs et d'aboutir ainsi à la création d'une seule protéine de fusion. C'est le cas par exemple des transposases de l'IS*1* et de certains membres de la famille IS*3*. Deux domaines définissent la transposase : un domaine N-terminal portant l'activité de liaison à une séquence spécifique d'ADN (le domaine II des IRs) et un domaine C-terminal portant l'activité catalytique de mobilisation. La plupart des transposases ont la capacité de générer des formes multimériques essentielles à leur activité (Polard et Chandler, 1995).

Une autre propriété des ISs est que leur insertion génère, dans la plupart des cas, la duplication de quelques paires de bases (de 2 à 14 pb) du site cible, formant des séquences répétées directes (DRs) flanquant l'IS insérée.

Une très grande diversité de ces trois caractéristiques, nature et structure des IRs, nature et structure de la transposase et devenir du site cible d'insertion, est observée. Ceci a amené à établir une classification des séquences d'insertion ; 17 familles sont ainsi reconnues : IS*1*, IS*3*, IS*4*, IS*5*, IS*6*, IS2*1*, IS*30*, IS*66*, IS*91*, IS*110*, IS*200/605*, IS*256*, IS*630*, IS*982*, IS*1380*, ISAs*1* et ISL3 (Mahillon et Chandler, 1998).

Les transposons composites, ou Tn composites, sont des structures plus importantes (quelques milliers de paires de bases) dans lesquelles un ou plusieurs gènes sont encadrés par deux séquences d'insertion identiques disposées dans la même direction ou en direction opposée, collaborant entre elles lors de leur transposition et mobilisant ainsi toute la séquence d'ADN qui les sépare. Les transposons composites sont très nombreux chez les bactéries à Gram négatif. Les plus connus sont : l'élément Tn*5* de 5700 pb portant trois gènes de résistance à la kanamycine, à la bléomycine et à la streptomycine, l'élément Tn*9* de 2638 pb portant un gène de résistance au chloramphénicol et l'élément Tn*10* de 9300 pb portant un gène de résistance à la tétracycline (Snyder, 2007).

Les transposons non composites, encore appelés transposons unitaires, ne sont pas associés à des séquences d'insertion. Ils possèdent des séquences flanquantes répétées inversées (d'environ 40 pb), l'information génétique nécessaire à leur mobilisation et divers autres gènes. Contrairement aux séquences d'insertion, les transposons unitaires forment une famille très homogène, tant du point de vue structural que du point de vue fonctionnel. Leur mode de transposition implique une transposase mais également une autre enzyme, une résolvase, qui permet de terminer la réaction en catalysant une étape de recombinaison à un site spécifique (*res*) porté par le transposon. Les transposons unitaires se subdivisent en deux groupes selon l'orientation relative du cadre de lecture de la résolvase par rapport à celui de la transposase. Ils sont soit consécutifs et non interrompus par le site *res* où agit la résolvase, soit divergeant et séparés par le site *res*. Leur transposition conduit invariablement à une duplication de 5 pb du site cible. De nombreux transposons unitaires ont été identifiés dans le génome des bactéries aussi bien à Gram négatif qu'à Gram positif. Les exemples les plus connus sont : l'élément Tn*3* de 5000 pb, portant un gène de résistance à l'ampicilline et identifié dans le génome de bactéries à Gram négatif et l'élément Tn*501* de 5200 pb, portant un gène de résistance au mercure et identifié pour la première fois sur un plasmide d'une souche du genre *Pseudomonas* (Snyder, 2007).

Enfin, les transposons conjugatifs ne sont également pas associés à des séquences d'insertion et constituent la classe des plus grands transposons, les plus importants pouvant atteindre 100 kb. Ils possèdent des séquences flanquantes répétées inversées (de 20 à 30 pb), des informations génétiques nécessaires à leur mobilisation par transposition, des informations génétiques nécessaires à leur mobilisation par conjugaison et divers autres gènes (Salyers *et al*., 1995). Les fonctions assurant leur mobilisation par transposition sont portées par deux gènes codant une recombinase spécifique de site et une excisionase. Les fonctions catalysant la conjugaison sont très similaires à celles portées par les plasmides conjugatifs d'*E. coli*. Les prototypes des transposons conjugatifs ont été détectés dans le génome de bactérie à Gram positif. Il s'agit de l'élément Tn*916* de 16 kb, portant un gène de résistance à la tétracycline et identifié dans le génome des bactéries *Enterococcus faecalis* et de l'élément Tn*1545* de 25,5 kb, portant trois gènes de résistance à l'érythromycine, à la tétracycline et à la kanamycine et identifié dans le génome des bactéries *Streptococcus pneumoniae* (Clewell *et al*., 1995).

1.1.3. Les intégrons

Les intégrons constituent des systèmes génétiques de capture, d'intégration et d'expression de gènes qualifiés de "gènes cassettes". Incapables d'autoréplication, ces systèmes sont obligatoirement portés par un réplicon. Ils sont également très souvent retrouvés au sein de transposons, décrits précédemment.

Les composants essentiels des intégrons se situent dans une région 5' conservée d'environ 1,4 kb et incluent : un gène codant une intégrase, *int*I, responsable de l'intégration des gènes cassettes, un site d'attachement, *att*I, cible de l'intégration et un promoteur permettant l'expression des gènes cassettes intégrés (figure 2A).

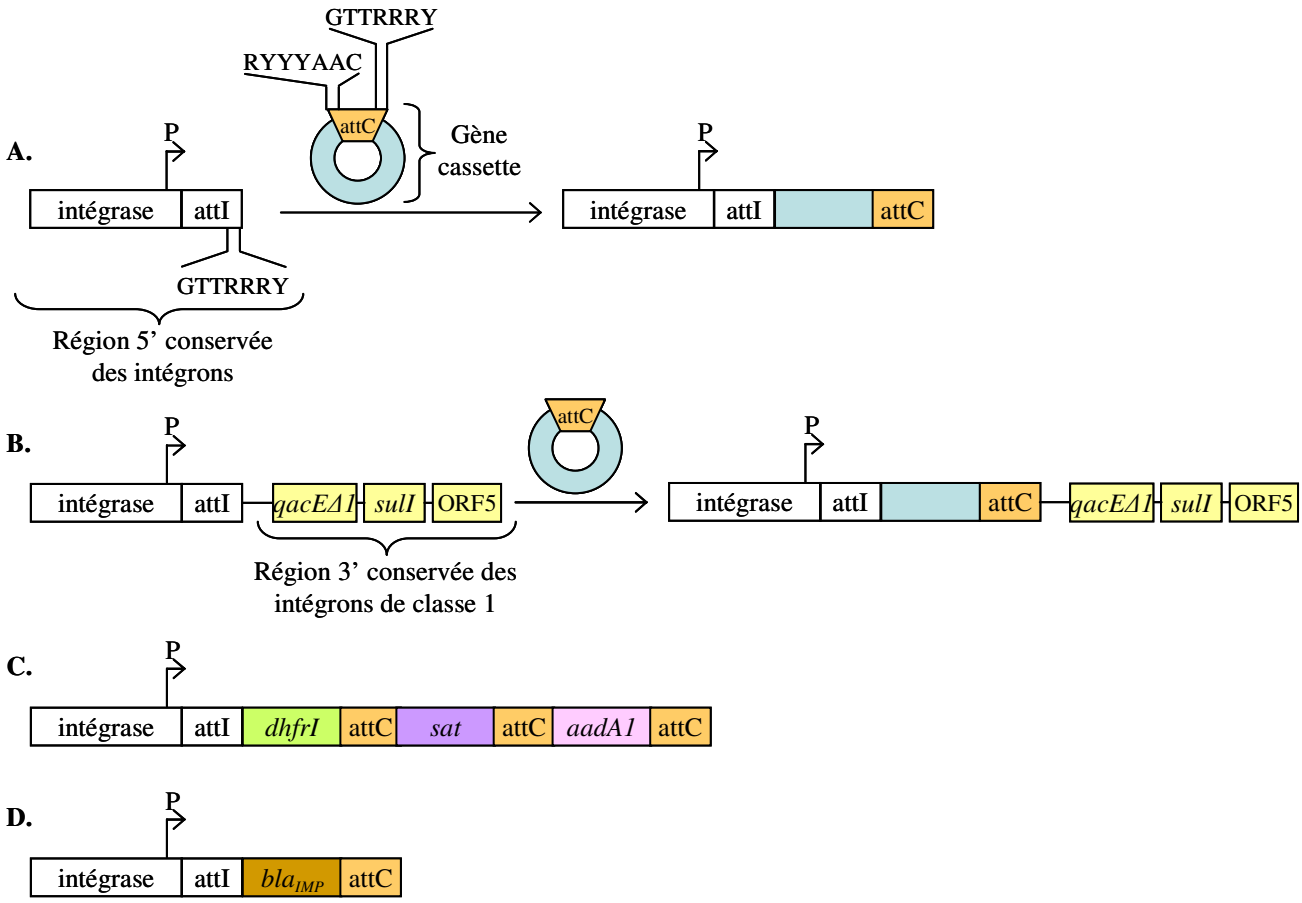


Figure 2 : Structure des intégrons.

A. Accueil d'un gène cassette par la région 5' conservée des intégrons, d'après Carattoli, 2001. P : promoteur ; att : site d'attachement ; R : purine ; Y : pyrymidine.

B. Accueil d'un gène cassette par un intégron de classe 1, d'après Carattoli, 2001. *qacEΔ1* : gène de résistance aux composés de type ammonium quaternaire partiellement délété ; *sul1* : gène de résistance aux sulfonamides ; ORF5 : ORF de fonction inconnue.

C. Intégron de classe 2 du transposon non composite Tn*7*, d'après Radström *et al.*, 1994. *dhfrI* : gène de résisitance à la triméthoprime ; *sat* : gène de résistance à la streptothricine ; *aadA1* : gène de résistance à la streptomycine / spectinomycine.

D. Intégron de classe 3 présent sur un plasmide de la souche *Serratia marcescens* AK9373, d'après Arakawa *et al.*, 1995. bla_{IMP} : gène codant une metallo-β-lactamase.

Selon l'homologie de la structure primaire des intégrases rencontrées, quatre classes d'intégrons ont été définies (Caratolli, 2001). Les sites *att*I de ces différentes classes n'ont pas de séquence commune, excepté un motif GTTRRRY conservé en 3' (R, purine ; Y, pyrimidine).

La classe 1 héberge les structures les plus étudiées et les plus communes, dont les toutes premières décrites comme intégrons (Stokes et Hall, 1989). En plus de la région 5' conservée, les intégrons de

classe 1 présentent une région 3' conservée d'environ 2 kb contenant un gène *qacEΔ1* de résistance aux antiseptiques (composés de type ammonium quaternaire) partiellement délété, un gène *sulI* de résistance aux sulfonamides et une phase ouverte de lecture ORF5 de fonction inconnue (figure 2B). Les gènes cassettes viennent s'insérer entre les régions 5' et 3' conservées. Il existe toutefois des intégrons de classe 1 pour lesquels aucun gène cassette inséré n'est observé, laissant les régions conservées adjacentes. C'est le cas de l'intégron In0 situé sur le plasmide pVS1 de la souche *P. aeruginosa* PAO25, pathogène opportuniste (Bissonnette et Roy, 1992). Quelques intégrons de classe 1 possèdent également deux promoteurs fonctionnels dans leur région 5' conservée.

Les intégrons de classe 2 sont situés au sein de transposons de la famille de Tn*7*. L'intégron du transposon non composite Tn*7* contient trois gènes cassettes adjacents à une intégrase défective et est dépourvu de la région 3' conservée, remplacée par les gènes nécessaires à la mobilité du transposon (figure 2C) (Hansson *et al*., 2002). Les trois gènes cassettes présents sont *dfr*A1, *sat* et *aad*A1 codant une résistance à la triméthoprime, à la streptothricine et à la streptomycine / spectinomycine, respectivement.

La classe 3 ne recense pour l'instant qu'un seul intégron identifié sur un plasmide de la souche *Serratia marcescens* AK9373, agent pathogène opportuniste à l'origine de maladies nosocomiales (figure 2D) (Arakawa *et al*., 1995). Cet intégron ne possède que la région 5' conservée ainsi qu'un seul gène cassette codant une metallo-β-lactamase.

La classe 4 a été crée avec la découverte d'un intégron de très grande taille (environ 130 kb), possédant 179 gènes cassettes, lors du séquençage des deux chromosomes de la souche *Vibrio cholerae* El Tor N16961 (Heidelberg *et al*., 2000). Ces gènes cassettes codent essentiellement des fonctions de virulence.

Les gènes cassettes sont des unités génétiques simples de 500 à 1000 pb constituées d'une ORF (rarement deux) de fonction variable, suivie d'un site de recombinaison, appelé *attC*, reconnu par l'intégrase des intégrons (Recchia et Hall, 1995). Tous les gènes cassettes ne possèdent pas le même site *attC*. La taille et la séquence de ce dernier varient en effet, avec, cependant, la conservation de deux régions inversées à ses extrémités de type RYYYAAC en 5' et GTTRRRY en 3' (figure 2A). Malgré des différences entre intégrases et sites *attI*, des gènes cassettes identiques peuvent être retrouvés dans des intégrons de différentes classes. Plusieurs gènes cassettes peuvent être insérés dans un même intégron, formant en quelque sorte un opéron sous la dépendance de l'unique promoteur situé dans la région 5' de l'intégron. Leur expression est dans ce cas tributaire de leur position par rapport au promoteur, les gènes localisés dans des cassettes éloignées étant faiblement exprimés. Les gènes cassettes insérés dans un intégron peuvent, après excision (catalysée également par l'intégrase), exister sous forme de molécules d'ADN libres, circulaires, non réplicatives, forme propice à une nouvelle insertion dans un nouvel intégron.

1.1.4. Les prophages

Les prophages sont des ADNs de bactériophages insérés dans un chromosome bactérien, se répliquant en même temps que ce dernier et transmis aux cellules filles lors de la division bactérienne. Les bactériophages sont des virus infectant les bactéries et comme tous les virus, ils se caractérisent par la possession d'un seul acide nucléique empaqueté dans une structure protéique et par un parasitisme intracellulaire obligatoire. La nature de leur acide nucléique (ADN simple ou double brin, ou ARN simple ou double brin) ainsi que leur structure sont les critères de leur classification. Seuls des bactériophages de l'ordre des Caudovirales peuvent intégrer leur ADN dans un chromosome bactérien. Ils sont qualifiés pour cela de bactériophages tempérés et établissent ainsi avec les bactéries des rapports de longue durée, éventuellement réversibles, nommés lysogénies. Les bactériophages tempérés insèrent leur ADN dans un chromosome bactérien grâce à l'intervention d'une intégrase codée par leur propre génome. Cette insertion peut être site-spécifique comme pour l'ADN du phage lambda qui ne s'intègre qu'entre les opérons *gal* et *bio* du chromosome des bactéries *E. coli* (Gottesman et Weisberg, 2004), ou être spécifique à plusieurs sites d'intégration comme pour l'ADN du phage P2 des bactéries *E. coli* (Bertani et Six, 1988). Elle peut être également totalement aléatoire (fruit d'un évènement de recombinaison transpositionnellle) comme pour l'ADN du phage Mu-1 des entérobactéries (Taylor, 1963). Naturellement (à une fréquence de 10^{-6} à 10^{-2}) ou sous l'induction de certains facteurs environnementaux (physiques ou chimiques), les prophages peuvent s'exciser de leur chromosome bactérien hôte toujours par le biais de l'intégrase accompagnée d'une excisionase. Il existe des cas particuliers de prophages non insérés dans un chromosome bactérien, comme celui de l'ADN du phage P1 des bactéries *E. coli*, restant à l'état libre dans le cytoplasme de la cellule, sous forme circulaire comme un plasmide (Lobocka *et al*., 2004).

Les prophages peuvent présenter une organisation et une séquence génétique très différentes de celles des ADN des bactériophages dont ils sont originaires. En effet, plusieurs évènements de mutations géniques ou chromosomiques, survenant dans le chromosome hôte, peuvent altérer leur structure. Des exemples de gènes isolés de prophages sont ainsi retrouvés dans le chromosome de la souche *S. agalactiae* NEM316, agent pathogène à l'origine de maladies néonatales (Glaser *et al*., 2002). La souche *Lactococcus lactis* IL1403, très utilisée en industrie laitière, présente quant à elle dans son génome trois prophages qualifiés de vestiges, c'est-à-dire ayant perdu une partie des gènes originels (notamment ceux contrôlant la morphologie des bactériophages et la lyse des cellules hôtes lors d'un cycle lytique) (Chopin *et al*., 2001). Enfin d'autres prophages arborent des structures chimériques, combinant des gènes issus d'ADN de bactériophages de familles différentes : c'est l'exemple du prophage SfV identifié dans le génome des souches *Shigella flexneri*, agents entéropathogènes, constitué de gènes typiques de l'ADN du phage lambda et de gènes typiques de l'ADN du phage Mu (Allison *et al.*, 2002).

Les bactéries contiennent souvent dans leur chromosome plusieurs prophages qui constituent alors une part non négligeable de leur génome total. Un exemple frappant est celui de la souche entéropathogène *E. coli* O157:H7 Sakaï : cette souche possède 18 prophages qui représentent 16 % de son gènome total (Hayashi *et al*., 2001). Des cas aussi impressionnants sont retrouvés avec des bactéries de l'espèce *S. pyogenes* (agents à l'origine de diverses pathologies chez l'homme), possédant 4 à 6 prophages totalisant jusqu'à 12 % de leur génome (Banks *et al*., 2002). Cependant, des bactéries dont les génomes sont dépourvus de prophages sont parfois identifiées. Il s'agit de parasites intracellulaires appartenant à la classe des γ-Protéobactéries telles que des bactéries du genre *Buchnera* (van Ham *et al*., 2003), ou au groupe des bactéries à Gram positif à haut pourcentage en paires de bases GC telle que *Mycobacterium leprae* (Cole *et al*., 2001), ou encore au groupe des bactéries à Gram positif à faible pourcentage en paires de bases GC telle que *Mycoplasma* (Sasaki *et al*., 2002).

1.1.5. Les îlots génomiques

Les îlots génomiques sont définis comme des zones de chromosome bactérien, ou parfois de plasmide, présentant des caractéristiques différentes de celles du reste du génome. Pouvant contenir des structures de différentes natures, dont des éléments génétiques mobiles ou mobilisables décrits précédemment ou des vestiges de ces derniers, les îlots génomiques présentent une organisation souvent chimérique et il est ainsi difficile de retracer les différents avènements ayant conduit à leur création. Ils semblent être ubiquitaires dans le monde bactérien. Un îlot génomique regroupe en général des gènes impliqués dans une même grande fonction. Des fonctions très diverses et dépendantes de la niche écologique des bactéries hôtes sont répertoriées. Sont désignés ainsi des îlots de pathogénicité (porteurs de fonctions de virulence), des îlots de symbiose (porteurs de fonctions permettant à une bactérie d'établir une relation symbiotique avec un autre organisme), des îlots de résistance (porteurs de fonctions de résistance à des antibiotiques), ou encore des îlots métaboliques (porteurs de fonctions apportant un métabolisme nouveau à une bactérie). L'expression de ces différentes fonctions peut être régulée par des composants codés, soit par l'îlot lui-même, soit par le génome hôte. Six caractéristiques communes permettent d'identifier ces éléments génétiques (Hacker et Kaper, 2000).

Leur première particularité est d'être "souche spécifique". Une bactérie peut porter un îlot génomique que des bactéries de la même espèce ne possèdent pas et que des bactéries très éloignées phylogénétiquement possèdent. Au sein de l'espèce *E. coli* par exemple, les souches K12 MG1655 (commensale non pathogène), CFT073 (uropathogène) et O157:H7 EDL933 (entéropathogène) possèdent des îlots génomiques dont 47,3 %, 61,7 % et 61,2 %, respectivement, des séquences codantes leurs sont spécifiques (Ou *et al*., 2006).

Les îlots occupent de grandes régions génomiques : des tailles de 10 à 200 kb sont généralement recensées. Un îlot d'une taille exceptionnelle de 500 kb a été identifié dans le chromosome de la souche symbiotique *Mesorhizobium loti* ICMP3153, représentant 10 % de son génome total (Sullivan et Ronson, 1998).

Leur contenu en paires de bases guanine et cytosine et leur usage des codons différent de ceux du reste du génome. La souche entéropathogène *E. coli* E2348/69 possède ainsi un génome et un îlot de pathogénicité dont les contenus en paires de bases guanine et cytosine sont de 51 % et 39 % respectivement (McDaniel *et al*., 1995).

Les îlots génomiques sont bordés par des séquences répétées directes d'une taille de 16 à 20 pb en moyenne ; une longueur de 135 pb a été remarquée pour un îlot d'une souche *E. coli* uropathogène (Redford et Welch, 2002). Ces séquences répétées directes sont probablement créées lors de l'insertion des îlots dans leur génome hôte par un phénomène de recombinaison site-spécifique (détaillé dans le paragraphe 1.3.2 de ce chapitre).

Dans plus de 75 % des cas, les îlots génomiques sont adjacents à des gènes codant des ARNts. Le site cible de leur intégration, qui se retrouve dupliqué après l'évènement de recombinaison, est ainsi constitué des 15 à 20 derniers nucléotides de ces gènes, comprenant la boucle T et les nucléotides CCA terminaux conservés des ARNts. Les gènes codant des ARNts ne sont pas tous utilisés équitablement en tant que points d'insertion chromosomique d'ADN étranger. Une étude analysant l'intégrité des loci d'ARNt d'une collection de souches de l'espèce *E. coli* a en effet montré que la majorité de ces loci sont et restent intacts et que seul un petit lot d'entre eux constitue une cible fréquente d'intégration (Germon *et al*., 2007). Ces derniers sont qualifiés de "points chauds" d'intégration. Les loci restant généralement intacts sont ceux situés à l'intérieur de structures opéroniques et polycistroniques, ceux transcrits à haute fréquence, ou encore ceux codant des ARNts reconnaissant des codons fréquemment utilisés. L'explication de cette spécificité d'intégration n'est pas encore parfaitement établie ; trois hypothèses sont avancées (Hou, 1999). La première suggère qu'un gène d'ARNt soit associé à un îlot afin que l'ARNt codé puisse lire les codons portés par l'îlot et rarement utilisés pour traduire les autres gènes du chromosome. Ceci a été observé avec le gène *leu*X codant le rare $ARNt^{leuX}$ (Ritter *et al*., 1997) : l'expression de *leu*X est nécessaire à la synthèse de facteurs de virulence codés par l'îlot de pathogénicité PAI_{536} qui lui est adjacent au sein du chromosome de la souche uropathogène *E. coli* 536. Cependant, étant donné que d'autres gènes de la cellule hôte, indépendants de l'îlot, requièrent également l'$ARNt^{leuX}$, cette première hypothèse ne semble pas favorable. La deuxième hypothèse émise prétend que les loci d'ARNt constituent de multiples points d'insertion possibles pour les îlots car ils sont présents en multiple copie sur le chromosome. Cette idée est de nouveau à rejeter puisque le locus *selC*, par exemple, très fréquemment utilisé pour l'insertion d'îlots chez des souches des espèces *E. coli*, *Shigella* spp. et *Salmonella enterica*, est présent en une seule copie sur le chromosome. La troisième et la plus probable hypothèse

suggère que la structure secondaire conservée des gènes d'ARNt représente un motif structural qui facilite l'intégration (et éventuellement l'excision) des îlots. Cette structure secondaire implique notamment la complémentarité des extrémités 5' et 3' des gènes d'ARNt, formant en quelque sorte des répétitions inversées reconnues par l'enzyme de mobilité qui catalyserait l'intégration des îlots par un phénomène de recombinaison.

Les îlots génomiques possèdent souvent des gènes codant des fonctions de mobilité telles que des intégrases, des transposases, des résolvases, ainsi que des éléments génétiques mobiles décrits précédemment tels que des séquences d'insertion, des transposons, des prophages. Ces différents gènes et éléments contrôlent la propre mobilité (intégration et excision) des îlots, expliquent leur organisation souvent chimérique (contenu très souvent de structures de différentes natures) et dictent leur plus ou moins bonne stabilité. Lorsqu'elles sont fonctionnelles, ces fonctions de mobilité constituent en effet une source de réarrangements génétiques des îlots et font de ces derniers des structures instables. Ceci est illustré, par exemple, avec les îlots de pathogénicité de différentes souches des espèces *Helicobacter pylori*, *Yersinia* spp. et *E. coli*, qui peuvent être délétés à des fréquences de 10^{-4}, fréquence plus élevées que celles des mutations géniques qui sont de l'ordre de 10^{-6}. Le plus souvent, cependant, les fonctions de mobilité codées par les îlots génomiques sont défectueuses et ne sont plus exprimées. Cette inactivation rend les îlots stables et leur intégration chromosomique permanente. C'est le cas par exemple des intégrases retrouvées dans le génome de la souche *E. coli* K12 MG1655, non fonctionnelles de par la présence de codons stop prématurés dans leurs séquences nucléotidiques.

1.2. Mécanismes responsables de leur mobilité intercellulaire

Il existe chez les bactéries trois mécanismes de transfert génétique intercellulaire. Deux d'entre eux correspondent à des transferts indirects : la transformation, où une bactérie capte et incorpore un ADN libre dans le milieu extérieur ; et la transduction, où une bactérie reçoit un nouvel ADN via son infection par un bactériophage. Le troisième est un transfert direct, la conjugaison, où une bactérie acquiert un nouvel ADN issu d'une autre bactérie, après avoir établi un contact physique avec elle. Chacun de ces moyens de transfert ne semble pas, cependant, prépondérant chez toutes les bactéries. Il apparaît en effet que, selon leur espèce, celles-ci utilisent de façon plus fréquente l'un ou l'autre de ces mécanismes. Ainsi, chez *Helicobacter pylori*, de nombreux évènements de transformation sont décrits, tandis que des évènements de transduction et de conjugaison sont beaucoup plus rarement rapportés (Israel, 2001). Chez *E. coli*, les évènements de tranduction semblent au contraire prédominer, étant donné le nombre important de prophages identifiés dans les génomes de plusieurs souches de cette espèce (Ohnishi *et al*., 2001).

1.2.1. La transformation

La transformation est un processus par lequel une cellule bactérienne se lie à un ADN libre dans son milieu environnant et l'incorpore de manière définitive dans son génome. Elle est le premier mécanisme d'échange génétique rapporté chez les bactéries (Griffith, 1928).

Les bactéries naturellement aptes à subir un évènement de transformation génétique sont dites compétentes. Elles développent pour cela un état physiologique particulier, la compétence, grâce à l'expression de gènes spécifiques. Plus de 40 genres bactériens, pour lesquels la compétence est un phénomène naturel, ont été identifiés (de Vries et Wackernagel, 2004). Pour la majorité d'entre eux, la compétence est un état physiologique transitoire. Seules les bactéries appartenant à l'espèce *Neisseria gonorrhoeae*, responsables chez l'homme de la gonococcie, expriment une compétence constitutive (Sparling, 1966). Les autres bactéries naturellement transformables développent un état de compétence de manière inductible, en réponse à des conditions environnementales spécifiques. Les phénomènes et les facteurs inducteurs sont divers et spécifiques à chaque souche, tels que : un stade de croissance particulier, une limitation ou un enrichissement en nutriments ou minéraux donnés, un changement de température ou de pH du milieu, l'apparition d'une nouvelle molécule (un antibiotique par exemple) dans le milieu…(Lorenz et Wackernagel, 1994). En plus de son impact évident et conséquent sur la diversification (Vestigian et Goldenfeld, 2005) et l'adaptation (Cohen *et al*., 2005) des génomes bactériens, la transformation naturelle serait un phénomène conservé par certaines bactéries, leurs procurant des ressources d'une part, en nucléotides, en atomes de carbone, d'azote et de phosphore (Redfield, 2001) et d'autre part, en séquences nucléotidiques pouvant être utilisées pour restaurer par recombinaison des gènes délétés ou dégradés (Szöllosi *et al*., 2006).

Le mécanisme de la transformation naturelle peut être divisé en quatre étapes (Lorenz et Wackernagel, 1994), chacune de ces étapes présentant un mode de déroulement et des acteurs souches spécifiques. La première étape, absente chez les bactéries *Neisseria gonorrheae* puisque celles-ci expriment une compétence constitutive, est la mise en place de l'état de compétence. Cette mise en place, inductible, fait appel à un groupe de gènes régulateurs agissant en cascade. Chez les streptocoques (bactéries où la transformation a été initialement découverte et où elle est la plus étudiée), ces gènes sont au nombre de trois (*com*C, *com*D et *com*E) et se situent sur un opéron (Cheng *et al*., 1997). La deuxième étape est la liaison de l'ADN libre environnant à la surface de la bactérie compétente. De nombreuses différences sont observées à ce stade entre les bactéries, notamment dans la spécificité de structure et de séquence de l'ADN. *S. pneumoniae*, agent pathogène majeur de la pneumonie et *Bacillus subtilis*, bactérie du sol, lient des ADNs uniquement doubles brins sans spécificité de séquence (Cvitkovitch, 2001). *Haemophilus influenzae*, bactérie responsable de diverses pathologies (tels que les otites, les méningites, les sinusites, les pharyngites et les conjonctivites) et *Neisseria gonorrheae* ne lient au

contraire que des ADNs présentant une séquence spécifique d'une dizaine de paires de bases environ et retrouvée fréquemment sur leur chromosome (Solomon et Grossman, 1996). *Neisseria gonorrheae* lie de plus aussi bien des ADNs doubles brins que simples brins. La troisième étape correspond au traitement et à l'entrée de l'ADN dans la cellule bactérienne. L'ADN subit des coupures sur l'un de ses brins et le brin intact est incorporé de la direction 3' vers 5' dans la cellule, où il est aussitôt pris en charge par des protéines de liaison le protégeant contre les attaques de nucléases intracellulaires (figure 3). Chez *S. pneumoniae*, l'enzyme catalysant les coupures simples brins de l'ADN lié est l'endonucléase membranaire EndA (Puyet *et al*., 1990).

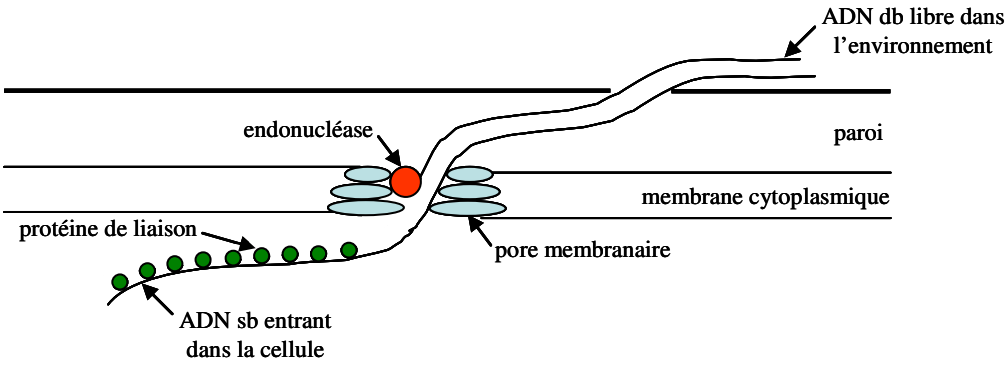


Figure 3 : Entrée d'un ADN libre de l'environnement dans une cellule bactérienne compétente par le processus de transformation, d'après Dreiseikelmann, 1994.

Un brin de l'ADN db présent dans l'environnement subit des coupures par une endonucléase membranaire. Le brin intact entre dans le cytoplasme de la cellule bactérienne par le biais d'un pore membranaire et est aussitôt pris en charge par des protéines de liaison.

Chez les bactéries à Gram négatif, cette troisième étape est accompagnée de la formation de petites structures membranaires visibles à l'extérieur de la cellule, appelées transformasomes, où l'ADN est incorporé avant d'être internalisé dans la cellule (figure 4) (Kahn *et al*., 1983).

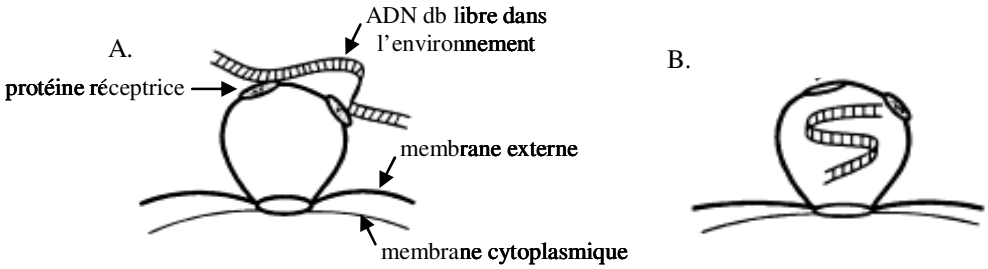


Figure 4 : Liaison d'un ADN libre de l'environnement sur la membrane externe des cellules bactériennes à Gram négatif, avant son entrée dans le cytoplasme par le processus de transformation, d'après Dreiseikelmann, 1994.

Chez les bactéries à Gram négatif, lors de la transformation, l'ADN db libre dans l'environnement se lie au préalable sur des protéines réceptrices de la membrane externe (A), est incorporé ensuite dans une structure appelée transformasome (B) et son entrée dans le cytoplasme est alors possible.

La dernière étape de la transformation naturelle est l'intégration de l'ADN incorporé dans le chromosome de la cellule hôte. Cette intégration se fait par un mécanisme de recombinaison homologue et/ou homéologue RecA dépendant et un système de réparation peut parfois intervenir en cas de mésappariements entre le nouveau brin d'ADN et le brin d'ADN hôte. Par exemple, chez *S. pneumoniae*, la recombinaison est dépendante de la protéine RecA (Mortier-Barrière *et al*., 1998) et le système de réparation éventuellement utilisé est le système Hex (Claverys et Lacks, 1986). L'utilisation d'un processus de recombinaison homologue implique que les bactéries naturellement compétentes réalisent des transformations génétiques qu'avec des ADNs de souches phylogénétiquement proches.

Quelques gènes ressemblant à ceux codant les différentes fonctions cellulaires requises pour la transformation naturelle ont été identifiés dans les génomes d'espèces bactériennes pour qui la compétence naturelle n'a jamais été observée. Il s'agit notamment des espèces *E. coli*, *Salmonella typhi*, *Klebsiella pneumoniae* et *P. aeruginosa* (Finkel et Kolter, 2001). Quelques espèces sont également connues pour devenir compétentes dans certaines situations (Demaneche *et al*., 2001a) et peuvent ainsi être capable d'incorporer de l'ADN libre environnant. Par exemple, *E. coli* peut devenir compétente lorsqu'elle est placée dans de l'eau minérale, ou dans un milieu carencé en nutriments, ou encore dans un sol fortement exposé à la lumière (Bauer *et al*., 1999 ; Finkel et Kolter, 2001 ; Demaneche *et al*., 2001b).

De grandes quantités d'ADN libre sont présentes dans l'environnement, aussi bien dans des milieux aquatiques que terrestres (Lorenz et Wackernagel, 1994). Leurs origines sont diverses, comme par exemple la lyse des cellules bactériennes. Ces ADNs libérés peuvent persister longtemps dans l'environnement en se complexant généralement à des sédiments tels que le quartz, le feldspath, ou divers métaux lourds (Austin *et al*., 1997). La vitesse de formation de ces complexes ADN / sédiment et leur stabilité dépendent de la nature du sédiment, du pH du milieu environnant, ou encore de la présence de certains cations dans le milieu environnant. Il a été démontré qu'une fois associés à des sédiments, les ADNs présentent une résistance aux attaques hydrolytiques des DNases, tout en conservant leur pouvoir transformant, c'est-à-dire tout en restant accessibles aux bactéries compétentes. Des évènements de transformation ont ainsi été mis en évidence entre des bactéries *Bacillus subtilis*, *P. stutzeri* ou *Acinetobacter calcoaceticus* compétentes et de l'ADN chromosomique adsorbé sur des particules de sable (Lorenz *et al*., 1988 ; Lorenz et Wackernagel, 1990 ; Chamier *et al*., 1993). Pour les bactéries *Bacillus subtilis*, le taux de transformation obtenu avec l'ADN chromosomique adsorbé était supérieur à celui obtenu avec le même ADN chromosomique libre non adsorbé ; pour les bactéries *P. stutzeri* et *Acinetobacter calcoaceticus*, les taux de tranformation étaient identiques dans les deux cas.

1.2.2. La conjugaison

La conjugaison est un processus de transfert unidirectionnel d'ADN d'une bactérie donatrice vers une bactérie réceptrice, par un mécanisme requérant un contact spécifique. Les gènes codant les fonctions nécessaires à ce phénomène sont uniquement portés par la bactérie donatrice et peuvent être retrouvés sur deux types d'éléments génétiques mobiles particuliers décrits dans le paragraphe 1.1 de ce chapitre : les plasmides conjugatifs et les transposons conjugatifs. Découverte en 1946 sur des cultures de bactéries de l'espèce *E. coli* (Tatum et Ledenberg, 1947), la conjugaison est un processus de transfert génétique très large, s'opérant aussi bien entre deux bactéries à Gram négatif ou deux bactéries à Gram positif d'espèces et de genres différents, qu'entre une bactérie à Gram négatif et une bactérie à Gram positif (Mazodier et Davies, 1991), qu'entre une bactérie et un organisme eucaryote inférieur (Heinemann et Sprague, 1989), ou qu'entre une bactérie et un organisme eucaryote supérieur (Waters, 2001). Le mécanisme du transfert génétique par conjugaison peut être décomposé en trois grandes étapes, chacune de ces étapes présentant un mode de déroulement et des acteurs spécifiques des types de bactéries impliquées (différences entre bactéries à Gram différents, mais également entre bactéries à Gram identiques) et de plasmides conjugatifs impliqués (différences entre plasmides de groupes d'incompatibilité différents). A l'heure actuelle, le mécanisme est le mieux compris, car étudié depuis plus longtemps, chez les bactéries à Gram négatif. Les exemples très souvent cités comme modèles sont le plasmide F d'*E. coli* appartenant au groupe d'incompatibilité IncF (Frost *et al*., 1994) et le plasmide RP4 à large spectre d'hôte appartenant au groupe d'incompatibilité IncP (Pansegrau *et al.,* 1994). Chez les bactéries à Gram positif, la compréhension du mécanisme est moins avancée, mais des études comparatives de séquences nucléotidiques partielles ou complètes de plasmides conjugatifs isolés de bactéries à Gram positif montrent des homologies avec certains acteurs conjugatifs des bactéries à Gram négatif (Morton *et al*., 1993 ; Berg *et al*., 1998 ; Dougherty *et al*., 1998 ; Schwarz *et al*., 2001).

La première étape du processus de conjugaison repose sur l'établissement d'un contact physique entre une bactérie détentrice d'un élément génétique conjugatif (bactérie donneuse ou mâle) et une bactérie qui en est dépourvue (bactérie réceptrice ou femelle) et à la stabilisation de ce contact (figure 5).

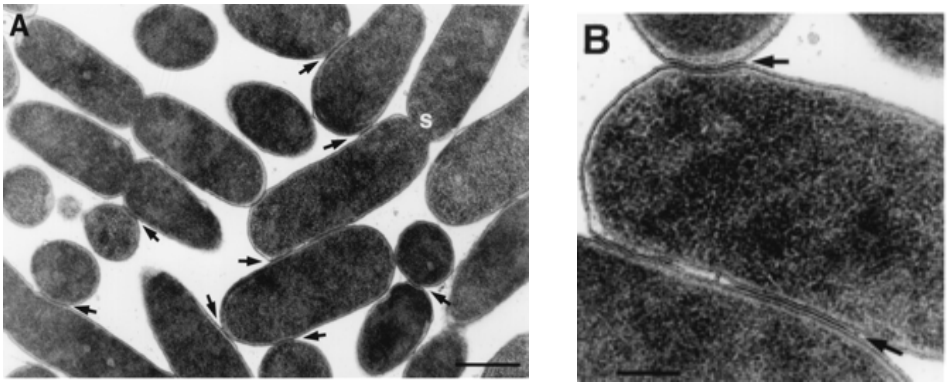


Figure 5 : Accolement de cellules bactériennes donneuses et réceptrices lors d'un processus de conjugaison, d'après Samuels *et al*., 2000.

Chez les bactéries à Gram négatif, le contact s'établit grâce à des appendices extracellulaires ancrés dans la membrane externe de la bactérie donneuse, appelés pili sexuels. Les fonctions nécessaires à la synthèse et à l'assemblage structural des pili et à la stabilisation du contact sont codées sur les plasmides conjugatifs ; elles sont par exemple au nombre de 13 sur le plasmide F (Frost *et al*., 1994) et de 11 sur le plasmide RP4 (Haase *et al*., 1995 ; Samuels *et al*., 2000). Les pili reconnaissent par leurs extrémités distales des zones de contact à la surface des bactéries réceptrices sur lesquelles ils se fixent. Ces zones de contact sont chez les bactéries réceptrices *E. coli* les protéines OmpA (Manoil et Rosenbusch, 1982). Un évènement de conjugaison ne peut avoir lieu entre deux bactéries donneuses à Gram négatif du fait de l'existence d'un système d'exclusion codé par les plasmides conjugatifs. Chez le plasmide F, ce système est porté par le gène *tra*T codant une lipoprotéine de la membrane externe, TraT. Cette lipoprotéine présente une homologie de séquence avec les protéines OmpA et surtout, une meilleure affinité pour les pili. Une fois le pilus d'une bactérie donneuse fixé sur une protéine TraT d'une autre bactérie donneuse, la fixation sur les protéines OmpA est inhibée et le processus de conjugaison est arrêté (Frost *et al*., 1994).

Chez les bactéries à Gram positif, différents mécanismes établissant la première étape de contact sont identifiés (Grohmann *et al*., 2003). Chez les entérocoques (bactéries commensales des tractus digestif et génital) par exemple, un mécanisme dépendant de la sécrétion de phéromones est utilisé. La bactérie réceptrice produit dans le milieu extérieur des phéromones induisant chez la bactérie donneuse la synthèse d'une protéine de surface formant de grands agrégats entre les deux bactéries (Clewell *et al*., 2002). Chez d'autres bactéries à Gram positif, le contact se fait toujours par agrégation, mais indépendamment de la production de phéromones. C'est le cas par exemple des bactéries du sol des espèces *Bacillus thuringiensis* subsp. *israelensis* (Andrup *et al*., 1993) et *Lactococcus lactis* (van der Lelie *et al*., 1991). Chez les bactéries du genre *Streptomyces* l'induction d'un contact et donc l'existence de gènes spécifiques impliqués dans ce contact ne sont pas rapportées

(Grohmann *et al*., 2003). Leur mode de vie sous forme de mycélium lie naturellement les cellules entre elles.

La deuxième étape du processus correspond au transfert du matériel génétique de la bactérie mâle vers la bactérie femelle, impliquant un traitement spécifique de l'ADN à transférer et l'établissement d'une voie de transit entre les deux bactéries (figure 6).

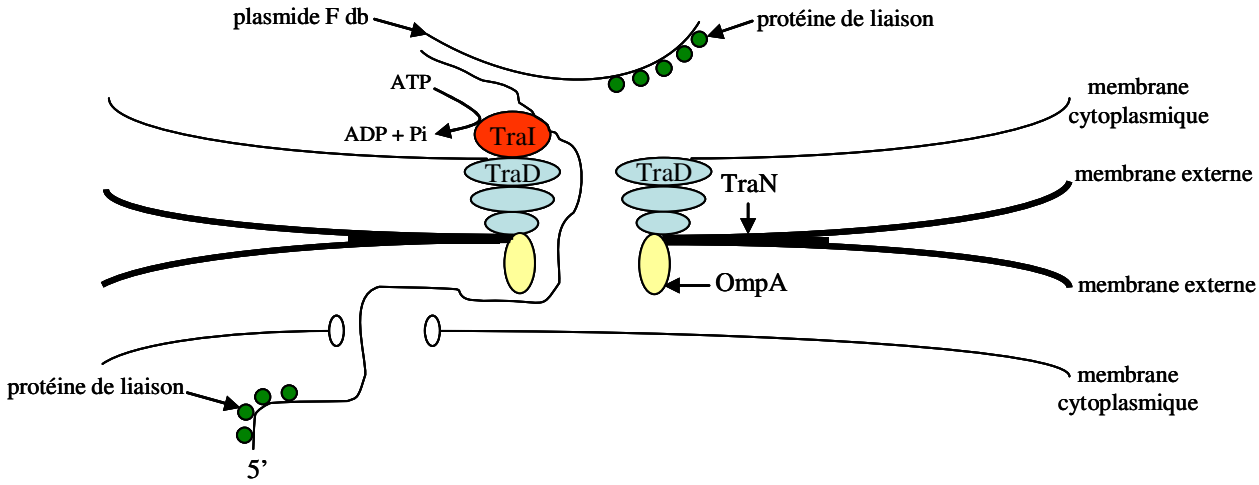


Figure 6 : Transfert par conjugaison du plasmide F, d'après Dreiseikelmann, 1994.

Après une coupure sb par une nucléase à l'origine de transfert oriT, le brin néoclivé est séparé de la molécule bicaténaire et amené à un pore membranaire par l'hélicase TraI.

Pour le traitement de l'ADN, les études avancées effectuées chez les bactéries à Gram négatif et celles moins avancées effectuées chez les bactéries à Gram positif montrent un mécanisme globalement similaire dans les deux cas. L'ADN est pris en charge par un complexe multiprotéique, le relaxosome, comportant notamment une fonction nucléase et une fonction hélicase. Un brin de l'ADN est ainsi coupé spécifiquement (à l'origine de transfert *ori*T) et transféré dans la cellule réceptrice. Une exception existe, chez les bactéries du genre *Streptomyces*, où l'ADN transféré est double brin (Possoz *et al*., 2001). Le transfert de l'ADN simple brin se fait par un pore reliant les deux bactéries. La structure de ce pore s'apparente aux systèmes de sécrétion de type IV (Cao et Saier, 2001) et les composants nécessaires à sa synthèse et à son assemblage sont en partie les mêmes que ceux impliqués dans la synthèse et la structure des pili. Sur les plasmides conjugatifs de bactéries à Gram positif, plusieurs protéines homologues à des protéines appartenant à des systèmes de sécrétion de type IV on été identifiées (Grohmann *et al*., 2003).

La troisième et dernière étape du processus correspond à la synthèse d'un brin complémentaire et à la ligation sous forme circulaire des ADNs mobilisés à la fois dans la cellule donneuse et dans la cellule réceptrice. Dans la cellule donneuse, cette troisième étape est en fait concomitante à la deuxième. Dès le début du transfert du brin clivé, un brin complémentaire à celui restant dans la

cellule donneuse est synthétisé selon le mécanisme du cercle tournant, grâce aux ADN polymérases hôtes ; un second clivage du brin transféré met fin au transfert et à cette nouvelle synthèse. Dans la cellule réceptrice, l'ADN simple brin introduit est stabilisé par des protéines de type SSB (Single Strand DNA Binding protein). Certains plasmides conjugatifs codent leurs propres protéines SSB qui accompagnent le simple brin d'ADN tout au long de son transfert, comme le plasmide F (Frost *et al*., 1994) ; d'autres sont dépourvus de cette fonction et le simple brin d'ADN est alors pris en charge par des protéines SSB de la cellule réceptrice. La synthèse du brin complémentaire au brin nouvellement introduit dans la cellule réceptrice se fait de manière discontinue (figure 7).

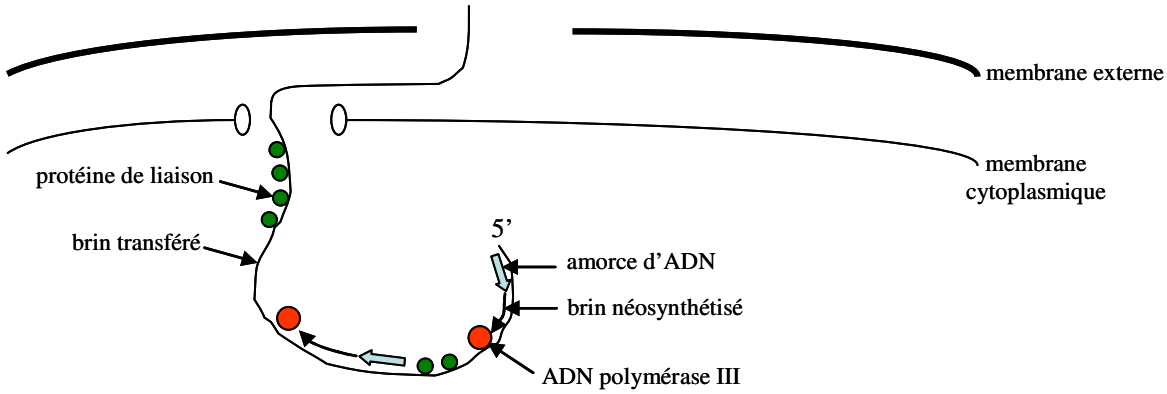


Figure 7 : Synthèse d'un brin complémentaire au brin d'ADN nouvellement reçu par une cellule bactérienne réceptrice après un processus de conjugaison, d'après Willetts et Wilkins, 1984.

Une fois arrivé dans le cytoplasme de la cellule réceptrice, le brin transféré d'un plasmide conjugatif est pris en charge par des protéines de liaison. Puis des primases synthétisent des amorces à partir desquelles l'ADN polymérase III débute la synthèse du brin complémentaire au brin transféré.

Des protéines primases synthétisent des amorces à partir desquelles les ADN polymérases hôtes peuvent démarrer leur synthèse. Ces protéines primases peuvent, comme dans le cas des protéines SSB, être codées soit par le plasmide conjugatif de la cellule donneuse et donc accompagner le simple brin durant son transfert, soit par le génome de la cellule réceptrice. Le premier cas cité est illustré avec le plasmide RP4, codant une protéine TraC, à activité primase, transféré en même temps que le simple brin (Lanka et Barth, 1981) ; et le second cas avec le plasmide R391, où aucune primase n'est mobilisée durant le transfert (Dalrymphe, 1982). Une fois les brins complémentaires synthétisés et ligués dans chaque cellule, des topoisomérases hôtes redonnent une structure superenroulée aux plasmides.

Certains plasmides conjugatifs peuvent s'intégrer dans le chromosome de la bactérie qui les héberge, de par la possession de transposons ou de séquences d'insertion capables de recombinaison. Que leur intégration chromosomique soit réversible ou au contraire stable, ces épisomes conjugatifs peuvent alors être à l'origine d'une dissémination des gènes du chromosome hôte en même temps que

leurs propres gènes. Lors d'une intégration réversible, l'excision peut parfois être incorrecte et inclure une petite fraction adjacente du chromosome hôte, fraction qui sera automatiquement transférée lors d'un évènement de conjugaison ultérieur. Lors d'une intégration stable, un évènement de conjugaison conduit forcément au transfert du chromosome hôte vers la bactérie réceptrice (transfert restant cependant incomplet étant donné les temps de contact courts entre les bactéries durant un évènement de conjugaison).

1.2.3. La transduction

La transduction est un transfert d'ADN entre bactéries par l'intermédiaire de bactériophages. Après avoir infecté une première bactérie, un bactériophage peut se libérer en empaquetant par erreur dans sa capside un fragment du génome de ce premier hôte, fragment qui peut être introduit dans un nouveau génome hôte lors de l'infection d'une deuxième bactérie par le bactériophage. Compte tenu de la spécificité existant entre les bactériophages et les bactéries, ces transferts se font essentiellement entre bactéries appartenant à la même espèce. Cependant, c'est un processus extrêmement efficace du fait de la protection de l'ADN dans les particules virales lors du passage de l'ADN d'une bactérie à une autre et d'un système de délivrance très performant par les bactériophages. Ceux-ci sont de plus très abondants dans l'environnement. Par exemple, dans les eaux marines, une concentration de 10^7 particules phagiques par millilitre est estimée (Brüssow et Hendrix, 2002) ; sachant que la fréquence de transduction dans ces milieux est évaluée à 10^{-8} (Jiang et Paul, 1998), un taux de 20.10^{15} évènements de transferts génétiques gérés par les bactériophages par seconde se produirait dans les océans. Sur les trois mécanismes à l'origine de transfert génétique entre bactéries, la transduction est celui le plus récemment découvert, en 1952 chez des bactéries de l'espèce *Salmonella typhimurium* (Zinder et Lederberg, 1952). Depuis, elle a été observée chez de nombreuses bactéries à Gram négatif et à Gram positif. Il existe deux types de transduction, la transduction généralisée et la transduction spécialisée, reposant sur les deux modes d'infection possibles des bactériophages : le mode virulent, encore appelé lytique et le mode tempéré, encore appelé lysogène.

1.3. Mécanismes responsables de leur insertion chromosomique

Trois mécanismes peuvent permettre l'insertion chromosomique d'éléments génétiques : la recombinaison homologue, la recombinaison site-spécifique et la recombinaison transpositionnelle. L'utilisation de l'un ou de l'autre de ces mécanismes dépend de la nature de l'élément génétique et parfois de l'état physiologique de la bactérie réceptrice.

1.3.1. La recombinaison homologue

La recombinaison homologue permet à un fragment d'ADN étranger de s'insérer dans un chromosome sous réserve qu'il présente une séquence homologue à une portion de celui-ci. Catalysée exclusivement par des acteurs apportés par l'hôte, elle est fortement dépendante du degré d'homologie des séquences nucléotidiques des deux partenaires génétiques impliqués. Quelques études ont été entreprises afin d'estimer le degré d'homologie nécessaire à une recombinaison efficace chez différents types de bactéries. Ainsi, chez les entérobactéries (Vulic *et al*., 1997, 1999) et les bactéries des genres *Bacillus* (Majewski et Cohan, 1998) et *Streptococcus* (Majewski *et al.*, 2000), la fréquence de recombinaison augmente de manière exponentielle lorsque la similarité des séquences nucléotidiques des deux partenaires d'ADN augmente. Dans la majorité des cas, la recombinaison échoue si les séquences divergent de plus de 1-2 %. Cette barrière peut toutefois être levée lors de conditions de croissance particulières ou d'avènements de mutations de certains gènes du chromosome de la bactérie, qui se trouve alors dans un état dit mutateur, où ses systèmes de "surveillance génétique" sont affaiblis. Un exemple bien connu d'état mutateur amenant à une levée du contrôle de la recombinaison homologue, par exemple, est l'altération du système MMR (pour Methyl directed Mismatch Repair) (Evans et Alani, 2000). Ce système, initialement identifié chez *E. coli* et retrouvé de façon ubiquitaire dans le monde bactérien, joue un rôle important dans le maintien de l'intégrité et de la stabilité du chromosome, notamment en inhibant les évènements de recombinaison entre séquences non homologues (Rayssiguier *et al*., 1989). Ainsi, des mutations dans le système MMR peuvent occasionner des évènements de recombinaison entre séquences dont le pourcentage de divergence est élevé, comme cela a été observé entre des bactéries de genres différents telles que *E. coli* et *Salmonella typhimurium* (Rayssiguier *et al*., 1991). Outre ces évènements de mutations altérant directement l'activité du système MMR, il a été montré que ce dernier n'est pas exprimé de manière constitutive. Une diminution de la concentration cellulaire de deux de ses constituants (MutS et MutH) a notamment été observée lors de la phase stationnaire de croissance des bactéries (Feng *et al*., 1996), s'accordant en parallèle avec une diminution du contrôle de la recombinaison homologue.

1.3.2. La recombinaison site-spécifique

La recombinaison site-spécifique permet l'insertion d'un fragment d'ADN étranger dans un chromosome au niveau de sites spécifiques reconnus sur chacun de ces deux partenaires. Ces sites sont des séquences identiques, mais de taille trop petite pour permettre à l'évènement d'intégration d'être pris en charge par la protéine RecA dirigeant les phénomènes de recombinaison homologue. Les ADN étrangers concernés par ce type de mécanisme sont les plasmides, les intégrons, les

transposons conjugatifs, les ADN de bactériophages tempérés à intégration spécifique et les îlots génomiques. Ces ADN portent eux-mêmes la machinerie impliquée dans leur intégration chromosomique.

Le mécanisme d'intégration par recombinaison site-spécifique repose sur la reconnaissance et le rapprochement des sites identiques partagés par l'ADN étranger et le chromosome hôte, puis sur l'échange des brins des deux partenaires suite à des coupures et des ligations contrôlées par des réactions de transéstérification (Grindley *et al*., 2006). Ce mécanisme est conservatif, il n'engendre aucune perte ou gain de nucléotides sur les deux partenaires et ne requiert aucun cofacteur énergétique tel que l'ATP. Bien souvent, une seule enzyme, une intégrase, catalyse l'ensemble de ces réactions. Les intégrases sont plus généralement des recombinases site-spécifique. Sur la base de leur structure primaire et de leur mécanisme catalytique, les recombinases site-spécifique sont partagées entre deux familles : la famille des recombinases à tyrosine (Nunes-Duby *et al*., 1998) et la famille des recombinases à sérine (Smith et Thorpe, 2002). La propriété commune à ces deux familles est la formation d'un intermédiaire covalent avec les partenaires d'ADN grâce à des attaques nucléophiles de leur acide aminé catalytique conservé sur les liaisons phosphodiesters des ADNs. La différence est que les recombinases à tyrosine forment des liaisons 3'phosphoryltyrosine, alors que les recombinases à sérine forment des liaisons 5'phosphorylsérine. Une autre différence est que les recombinases à tyrosine procèdent à deux évènements de clivage où un des deux brins de chaque partenaire d'ADN est clivé à la fois et entre lesquels une croix de Holliday se forme ; alors que les recombinases à sérine clivent les deux brins de chaque partenaire en même temps et dictent leur échange non pas par la formation d'une croix de Holliday, mais par un évènement de rotation (figures 8 et 9). Les recombinases à tyrosine sont très répandues chez les bactéries et présentent un domaine N-terminal variable de liaison d'ADN et un domaine C-terminal catalytique conservé contenant le résidu tyrosine nucléophile essentiel. Proche de ce dernier, cinq autres résidus conservés ont pu être identifiés dans la structure tertiaire de ces enzymes : deux résidus histidine, deux résidus arginine et un résidu lysine. Les recombinases à sérine constituent quant à elles une famille plus hétérogène de protéines, avec des organisations variables de domaines.

La recombinaison site-spécifique ne concerne pas uniquement l'intégration d'un ADN étranger dans un chromosome hôte, mais également ses possibles excision et inversion. Dans ces cas, les recombinases impliquées ne sont plus des intégrases, mais des excisionases et des invertases, respectivement. Malgré tout, intégrases, excisionases et invertases appartiennent toutes à l'une des deux familles de recombinase site-spécifique, donc présentent des mécanismes catalytiques similaires. Leur distinction se fait principalement dans la position des sites spécifiques qu'elles reconnaissent : sites identiques de même orientation portés par une seule molécule d'ADN pour les excisionases, sites identique d'orientation opposée portés par une seule molécule d'ADN pour les invertases et sites

identique portés par deux molécules d'ADN distinctes pour les intégrases. Certains motifs présents dans la structure primaire de ces enzymes dictent leur spécificité de reconnaissance.

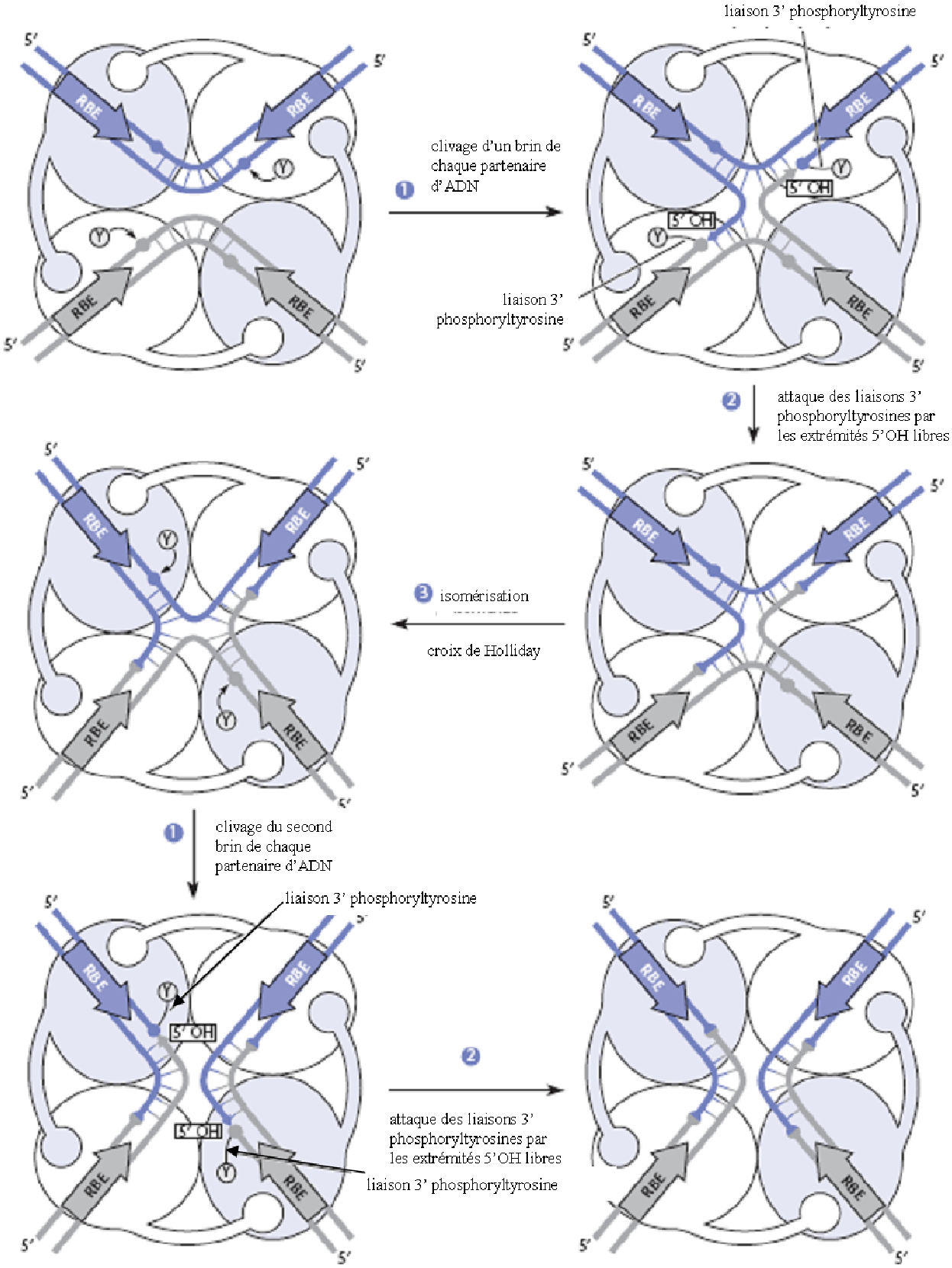


Figure 8 : Mécanisme de la recombinaison site-spécifique catalysée par les recombinases à tyrosine, d'après Snyder, 2007.

RBE : recombinase binding element ; Y : résidu tyrosine catalytique.

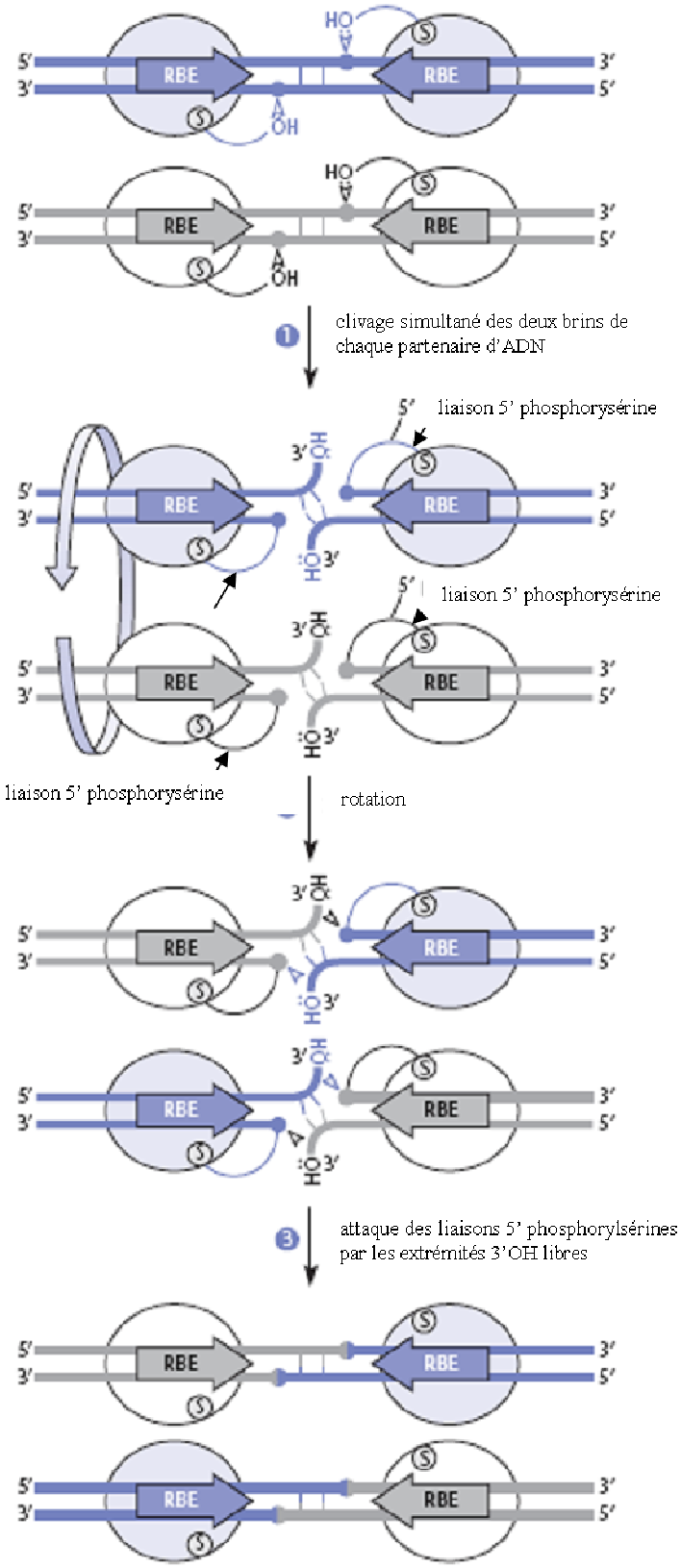


Figure 9 : Mécanisme de la recombinaison site-spécifique catalysée par les recombinases à sérine, d'après Snyder, 2007.

RBE : recombinase binding element ; S : résidu sérine catalytique.

Selon la nature du fragment d'ADN étranger, plusieurs sites spécifiques d'intégration ont été identifiés. Le plus souvent, le site d'intégration correspond à l'extrémité 3' de gènes codant des ARNt (Campbell, 1992 ; Hacker *et al.*, 1997). Dans ce cas, la séquence nucléotidique cible d'insertion est remplacée par une séquence identique apportée par l'élément qui s'intègre et le gène codant l'ARNt

garde ainsi son intégrité. Plusieurs hypothèses, présentées dans le paragraphe 1.1.5 de ce chapitre, ont été avancées pour expliquer cette spécificité d'intégration auprès des gènes codant des ARNt. Dans d'autres cas, la cible d'intégration correspond à l'extrémité 3' d'un gène qui code une protéine. La séquence nucléotidique cible d'insertion est alors souvent remplacée par une séquence, apportée par l'élément qui s'intègre, qui ne lui est que similaire (Campbell, 1992). Ainsi, l'ADN du bactériophage ΦLC3 de *Lactococcus lactis* s'intègre à l'extrémité 3' d'une ORF de fonction inconnue en remplaçant quinze nucléotides en amont du codon stop. Le gène recombinant formé code une protéine de longueur identique à celle de la protéine d'origine, mais dans laquelle les cinq derniers acides aminés ont été substitués (Lillehaug et Birkeland, 1993). L'ADN du bactériophage ΦL54a de *Staphylococcus aureus* s'intégre quant à lui dans la partie 3' du gène *geh* qui code une lipase. Cette intégration conduit à la formation d'une protéine inactive tronquée de 46 acides aminés dans la partie C-terminale (Lee et Iandolo, 1986). Enfin, quelques bactériophages, comme λ, intégrent leur ADN au niveau de régions intergéniques.

1.3.3. La recombinaison transpositionnelle

La recombinaison transpositionnelle, ou transposition, permet l'insertion d'un fragment d'ADN étranger dans un chromosome sans qu'aucun de ces deux partenaires ne présentent de similitude de séquences. Elle est donc le mécanisme capable de générer le plus de diversité et concerne les séquences d'insertion, les transposons composites, les transposons non composites et les ADN de bactériophages tempérés à intégration non spécifique. Comme pour la recombinaison site-spécifique, la machinerie impliquée dans le processus de la transposition est portée par les ADNs étrangers à intégrer.

Trois types de mécanismes de transposition peuvent être distingués (Snyder, 2007) : la transposition non réplicative ou par "couper-coller", la transposition réplicative et la transposition par cercle tournant. Ces trois types de mécanismes dépendent de l'enzyme catalysant la réaction, la transposase, présente sur l'ADN étranger.

Les transpositions non réplicative et réplicative présentent plusieurs caractéristiques communes. Leur signature quasi universelle est la duplication de quelques paires de bases du site cible de l'insertion. Les transposases qui gouvernent leur réaction présentent toutes de manière conservée trois acides aminés essentiels pour leur activité : deux résidus acide aspartique (D) et un résidu acide glutamique (E). Ces résidus, éloignés dans la structure primaire des transposases, se retrouvent rapprochés dans la structure tertiaire et ont pour rôle de chélater des ions magnésium participant aux réactions de clivage lors de la recombinaison. Aucune liaison covalente ne s'établit entre les transposases à motifs DDE et les ADNs partenaires. Enfin, leur dernière propriété commune est la reconnaissance et le

rapprochement dans l'espace, par la transposase, des extrémités du fragment d'ADN étranger à intégrer et du site cible non spécifique d'intégration pour former une structure appelée transposome. La distinction entre ces deux types de transposition se fait dans la nature des clivages des brins d'ADN de chaque partenaire (donneur et receveur) et dans la conservation ou non du fragment d'ADN mobilisé dans la molécule donneuse (Figure 10).

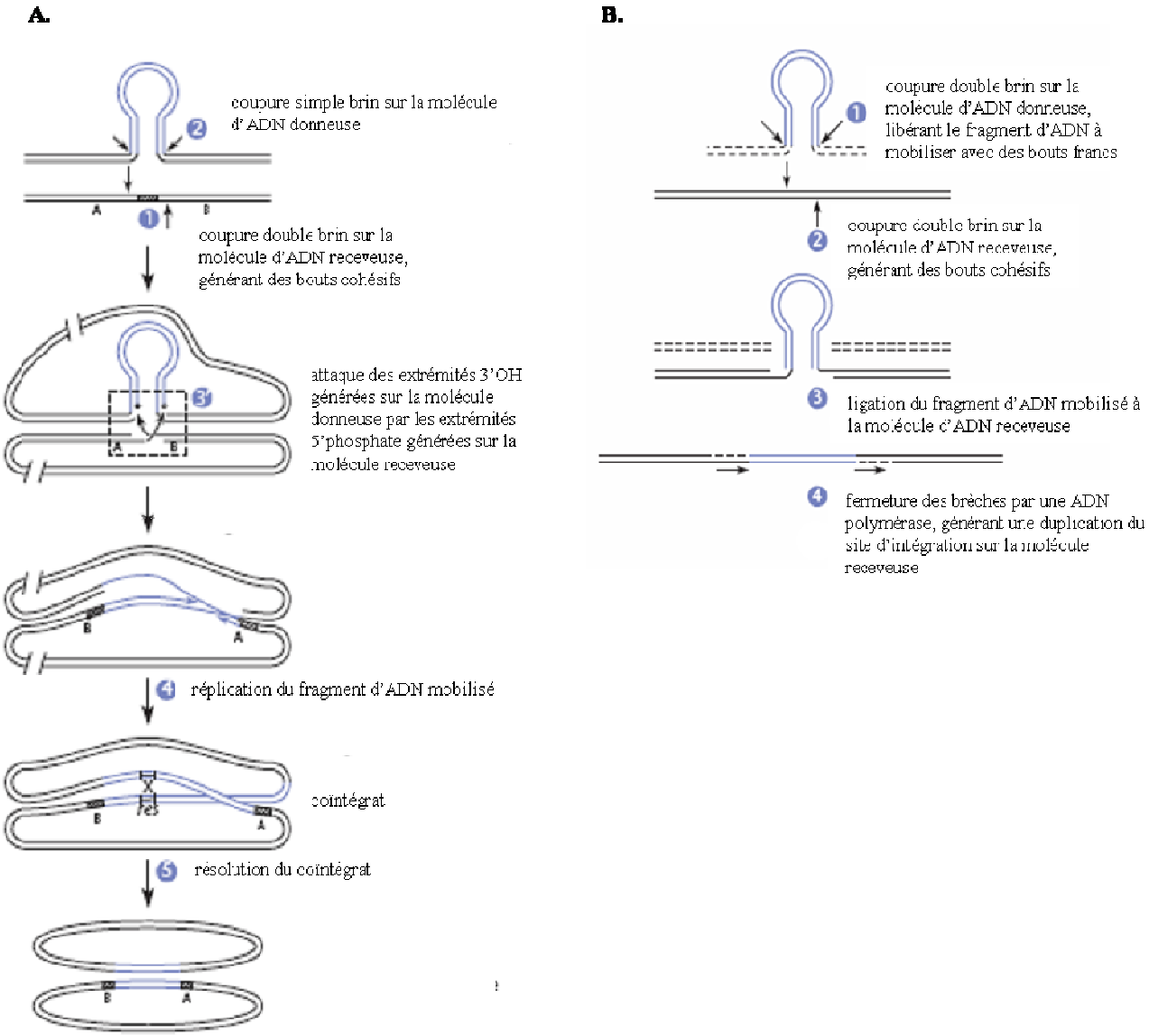


Figure 10 : Mécanismes des recombinaisons transpositionnelles réplicative (A) et non réplicative (B), d'après Snyder, 2007.

Lors d'une transposition non réplicative, le fragment d'ADN mobilisé est entièrement libéré de la molécule donneuse par une coupure double brin générant des bouts francs à chacune de ses extrémités, puis réinséré dans la molécule receveuse ayant subi au préalable une coupure double brin générant des bouts cohésifs. La brèche laissée au niveau des bouts cohésifs est ensuite refermée par réplication. Ceci conduit à la perte du fragment d'ADN mobilisé au niveau de la molécule donneuse.

Lors d'une transposition réplicative, le fragment d'ADN mobilisé n'est clivé que sur un brin de chacune de ses extrémités. Les extrémités 3'OH créées se lient aux extrémités 5' phosphate formées au niveau du site cible de la molécule receveuse par une coupure double brin générant des bouts cohésifs. Une réplication du fragment d'ADN mobilisé a ensuite lieu avec les extrémités 3'OH du site cible servant comme amorces. Les extrémités 3'OH des brins nouvellement synthétisés sont alors liguées aux extrémités 5' libres de la molécule donneuse pour former un cointégrat. Ce cointégrat, dans lequel les molécules donneuse et receveuse sont liées avec chacune une copie du fragment d'ADN mobilisé, est résolu grâce à l'intervention soit de protéines codées par la molécule receveuse, soit d'une résolvase codée par le fragment lui-même. Les résolvases sont des recombinases site-spécifique reconnaissant des séquences identiques de même orientation. Ceci conduit donc à l'acquisition du fragment mobilisé par la molécule receveuse sans que la molécule donneuse n'en perde une copie. La direction vers une transposition non réplicative ou réplicative semble être dictée par très peu de résidus d'acides aminés au niveau de la structure primaire des transposases à motif DDE. En effet, il a été montré qu'une mutation ponctuelle au niveau du gène codant la transposase du transposon Tn*7* engendrait la conversion de sa transposition d'un mode non réplicatif à un mode réplicatif (May et Graig, 1996).

La transposition par cercle-tournant est encore peu connue et comprise (Snyder, 2007). Elle conduit également à la conservation du fragment mobilisé dans la molécule donneuse grâce à la réplication de celui-ci selon un mode de cercle-tournant. Les transposases impliquées présentent deux résidus tyrosine essentiels à la catalyse, se liant de manière covalente au fragment mobilisé. L'intégration de ce dernier n'engendre pas une duplication du site cible de la molécule receveuse.

2. Les apports fonctionnels des éléments génétiques mobiles ou mobilisables bactériens

Selon Charles Darwin, l'évolution des organismes est gouvernée par un constant besoin d'augmenter leur fitness, c'est-à-dire de s'adapter au mieux et le plus rapidement possible à leur environnement et à son changement. Les éléments génétiques mobiles ou mobilisables décrits précédemment s'inscrivent parfaitement dans cet optique, dans la mesure où ils apportent aux bactéries qui les acquièrent divers gènes codant généralement de nouvelles fonctions favorisant leur adaptation aux conditions extérieures. Les fonctions mobilisées répertoriées à ce jour sont en effet très souvent des fonctions symbiotiques, métaboliques, de résistance aux antibiotiques, ou encore de virulence. Les gènes indispensables à la vie des bactéries, souvent appelés gènes constants et porteurs

de fonctions dites essentielles à la vie cellulaire (comme les gènes codant les ARNt ou les ARNr), semblent quant à eux beaucoup plus rarement mobilisés.

2.1. La symbiose

Les éléments génétiques mobiles porteurs de fonctions symbiotiques sont très souvent trouvés dans les génomes de bactéries vivant dans le sol, établissant alors des relations de symbiose avec d'autres organismes de cette niche écologique, aussi bien d'origine végétale que d'origine animale.

Chez les végétaux, les bactéries concernées peuvent notamment être les rhizobiums, capables d'induire la formation de nodosités au niveau des racines ou des tiges des légumineuses. L'analyse complète ou partielle des génomes de différentes rhizobiums a montré que leurs fonctions symbiotiques sont très souvent codées par deux types d'éléments génétiques mobiles possibles : des plasmides ou des îlots (MacLean *et al*., 2007). Généralement, les plasmides sont retrouvés chez les souches des genres *Rhizobium* et *Sinorhizobium* (tableau 1) et les îlots chez les souches des genres *Mesorhizobium* et *Bradyrhizobium* (tableau 2).

Tableau 1 : Caractéristiques de quatre plasmides symbiotiques identifiés dans le génome de rhizobiums.

Caractéristiques des plasmides	**Souches bactériennes détentrices**			
	***Rhizobium* sp NGR234**	***Sinorhizobium meliloti* 1021**	***Rhizobium leguminosarum* biovar *viciae* 3841**	***Rhizobium elti* CNF42**
Nom	pNGR234	pSymA	pRL10	p42d
Taille (kb)	536	1354	488	371
GC% (GC% du génome total de la souche détentrice)	58,5 (62,2)	60 (62)	59,6 (60,86)	58,35 (60,54)
Nombre d'ORFs	416	1296	471	354
Référence	Freidberg *et al*., 1997	Barnett *et al*., 2001	Young *et al*., 2006	Gonzalez *et al*., 2006

Tableau 2 : Caractéristiques de trois îlots symbiotiques identifiés dans le génome de rhizobiums.

Caractéristiques des îlots	Souches bactériennes détentrices		
	***Mesorhizobium loti* MAFF303099**	***Mesorhizobium loti* R7A**	***Bradyrhizobium japonicum* USDA110**
Taille (kb)	611	502	681
GC% (GC% du génome total de la souche détentrice)	59,7 (62,7)	59,3 (Non disponible)	59,4 (64,1)
Nombre d'ORFs	580	414	655
ARNt utilisé comme site d'intégration	$ARNt^{Phe}$	$ARNt^{Phe}$	$ARNt^{Val}$
Impact de la fonction de mobilité	19% des ORFs impliquées	8% de la séquence nucléotidique totale de l'îlot impliquée	15% des ORFs impliquées
Référence	Kaneko *et al.*, 2000	Sullivan *et al.*, 2002	Kaneko *et al.*, 2002

Ces deux types d'éléments génétiques symbiotiques spécifiques des rhizobiums possèdent plusieurs caractéristiques communes et l'hypothèse d'un ancêtre commun peut ainsi être avancée. Ils semblent tout d'abord partager les séquences codant les fonctions de nodulation et de fixation de l'azote atmosphérique (Kaneko *et al.*, 2000). Tous deux présentent également une structure très chimérique, expliquée par un contenu important en éléments génétiques mobiles tels que des séquences d'insertion, des transposons et des gènes codant différentes enzymes de mobilité (transposase, recombinase), suggérant l'avènement de plusieurs transferts génétiques à l'origine de leur formation. Comme le montre la représentation du plasmide pNGR234 en figure 11, ces éléments génétiques mobiles se disposent entre plusieurs blocs de gènes, chaque bloc contenant des gènes impliqués dans une même fonction.

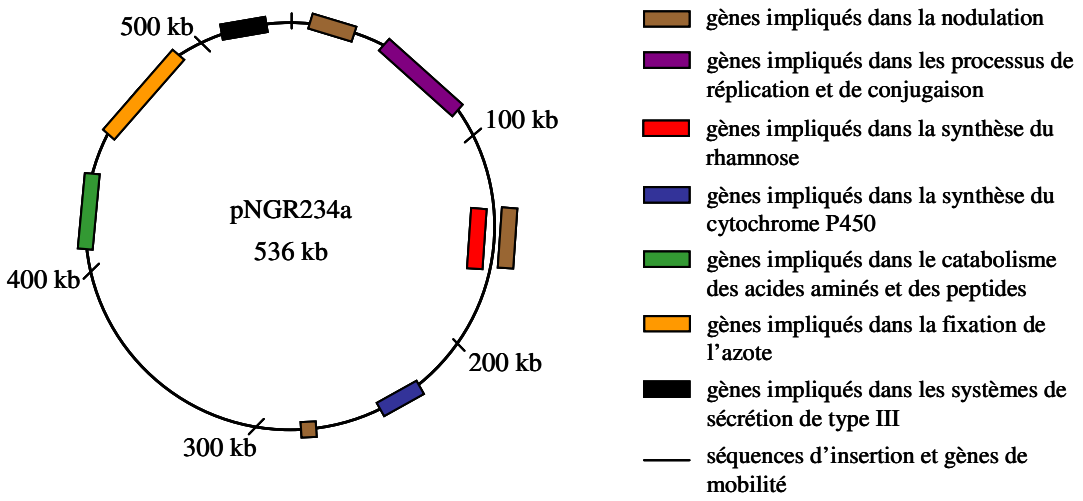


Figure 11 : Carte génétique du plasmide circulaire pNGR234a, d'après Freidberg *et al.*, 1997.

Chez les animaux, le cas des bactéries *Photorhabdus luminescens*, vivant dans l'intestin de nématodes pathogènes pour les insectes, peut être pris comme exemple. Dans cette relation symbiotique, lors de l'infection d'un insecte par un nématode, les bactéries sont libérées dans les vaisseaux sanguins de l'insecte et les tuent rapidement grâce à la production de toxines. Leur sécrétion concomitante de lipases et de protéases attaquant les tissus hôtes permet de fournir des ressources nutritives nécessaires à leur multiplication, mais également utilisables par les nématodes pour leur reproduction. Une fois ces ressources épuisées, bactéries et nématodes se réassocient, quittent la carcasse d'insecte et partent à la recherche d'un nouvel hôte (Forst et Nealson, 1996). Pour assurer ces différentes fonctions, les génomes de ces bactéries comprennent de nombreux éléments génétiques acquis par transfert horizontal (Ffrench-Constant *et al*., 2003). Chez la souche *Photorhabdus luminescens* W14, le gène *ngr*A, codant une protéine impliquée dans la synthèse de nutriments et de facteurs signaux essentiels à la croissance et à la reproduction des nématodes, peut être pris comme exemple. Ce gène est situé sur le chromosome entre deux régions de composition en paires de bases guanine et cytosine atypique : un îlot de pathogénicité codant plusieurs toxines et inséré en aval d'un gène codant un $ARNt^{Asp}$; et un opéron codant des fimbriae. Ffrench-Constant *et al*. suggèrent ainsi que le gène *ngr*A ait été acquis lors d'un co-transfert de ces deux régions.

2.2. Le métabolisme

De nombreux éléments génétiques mobiles porteurs de fonctions métaboliques sont identifiés à ce jour, permettant de couvrir l'utilisation d'une très large gamme de substrats. Cette dernière est en relation avec la niche écologique des bactéries qui les possèdent : intestins des animaux à sang chaud, sédiments, profondeurs terrestres ou océaniques…

2.2.1. Métabolisme des composés organiques "naturels"

Les composés organiques "naturels" ne peuvent pas tous être métabolisés par tous les organismes vivants. Par exemple au sein des entérobactéries, la plupart des souches de l'espèce *E. coli* et du genre *Salmonella* sont incapables d'utiliser le saccharose comme source de carbone (Brenner, 1984). L'analyse des rares souches capables de métaboliser ce sucre révèle la présence dans leur génome de deux types d'éléments génétiques mobiles possibles codant les fonctions nécessaires à ce métabolisme : un plasmide conjugatif ou un transposon conjugatif. Le plasmide conjugatif est nommé pUR400 et présente une taille de 70 kb (Schmid *et* al., 1982). Le transposon conjugatif est quant à lui nommé Tn*scr*-94 et présente une taille de 100 kb (Hochhut *et* al., 1997). Chez les souches de l'espèce *E. coli*, il a été montré que ce transposon présente deux sites préférentiels d'insertion sur le

chromosome : immédiatement en aval du gène *phe*V et immédiatement en aval du gène *phe*U, codant tous deux des $ARNt^{Phe}$. Le plasmide pUR400 et la transposon Tn*scr*94 codent les mêmes fonctions responsables de l'utilisation du saccharose : un système phosphotransférase phosphoénol-pyruvate-dépendant (PTS) nécessaire à l'incorporation du sucre dans la cellule et les différentes enzymes assurant sa métabolisation. En plus du partage de ces fonctions, ces deux éléments génétiques présentent un même ordre et une grande similarité de séquences, suggérant l'existence d'un ancêtre commun.

2.2.2. Métabolisme des composés organiques " non naturels"

Depuis le dernier siècle, les activités industrielles humaines se sont largement développées et multipliées, en s'accompagnant d'un rejet dans la biosphère de nombreux composés chimiques toxiques naturellement absents sur la Terre. Ces composés sont souvent qualifiés de xénobiotiques. Certaines bactéries sont apparues, cependant, aptes à résister à leur effet toxique, voire même à les utiliser comme sources de carbone, d'azote et d'énergie. Il est suggéré que le transfert génétique horizontal ait joué un rôle important dans ce développement, ainsi que la création de nouvelles voies métaboliques par la modification ou la combinaison de celles déjà existantes (Copley, 2000 ; Top et Springael, 2003). Ainsi, de nombreux gènes cataboliques permettant la dégradation de composés xénobiotiques sont désormais découverts et identifiés sur des plasmides, des transposons et des îlots génomiques.

Les plasmides cataboliques identifiés jusqu'à présent présentent des tailles variées allant de 10 à plus de 1000 kb, des structures circulaires ou linéaires et sont conjugatifs ou non (Nojiri *et al*., 2004). Leurs contenus génétiques codent les acteurs de voies cataboliques de nombreux substrats différents comme : le camphre, le naphtalène, l'octane, le phénol, la nicotine, l'atrazine, le chlorobenzoate, le toluène, l'aniline, le nitrobenzène, l'isopropylbenzène, l'acide cinnamique, le xylène, le benzène et bien d'autres encore. De nombreux éléments génétiques mobiles, tels que des séquences d'insertion et des transposons composites ou non composites, sont retrouvés au sein de ces plasmides, suggérant leur création par plusieurs avènements de transferts génétiques, en adéquation avec l'évolution des niches écologiques des bactéries. Le plasmide pWWO, isolé de la souche *P. putida* mt-2, est un bon exemple pour illustrer ceci. Il a été montré que les 116 kb qui composent ce plasmide pouvaient être divisés en deux régions d'origine différente : une région ancestrale de 46 kb, comportant les gènes de réplication, de maintenance et de mobilité par conjugaison et une région plus récente de 70 kb, acquise par la première et portant des fonctions impliquées dans le catabolisme du toluène (Figure 12) (Greated *et al*., 2002). Cette région plus récente correspond à un transposon non composite appelé Tn*4653*, formé à la suite d'au moins deux avènements d'insertion au sein d'un

transposon ancestral de 11 kb appelé Tn*4653*A : les insertions du transposon non composite Tn*4651*, d'une taille de 56 kb et portant les opérons *xyl* codant les fonctions cataboliques et de la séquence d'insertion IS*tol* d'une taille de 3 kb.

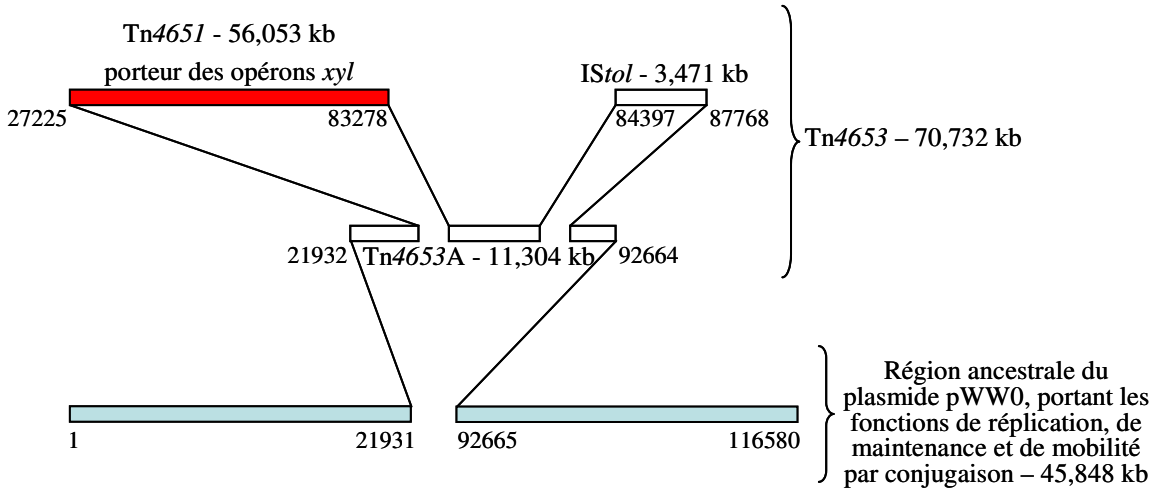


Figure 12 : Présentation des différents avènements d'intégration ayant créé le plasmide pWW0, d'après Greated *et al.*, 2002.

Ainsi, les fonctions cataboliques des plasmides impliqués dans la dégradation des composés xénobiotiques sont très souvent codées par des transposons composites ou non composites (tableau 3).

Tableau 3 : Transposons composites et non composites codant différents gènes impliqués dans la dégradation de composés xénobiotiques et présents sur des plasmides.

	Nom	Taille (kb)	Gènes cataboliques	Substrats	Origine (plasmide)	Référence
Transposons composites	Tn*5280*	8,5	*tcbAaAbAcAdB*	Chlorobenzène	*Pseudomonas* sp. P51 (p51)	van der Meer *et al.*, 1991
	Tn*5542*	12	*bed*	Benzène	*P. putida* ML2 (pHMT112)	Fong *et al.*, 2000
	Tn*5707*	15	*cbnRABCD*	3-Chlorobenzoate	*Ralstonia eutrophus* NH9 (pENH91)	Ogawa et Miyashita, 1999
	Tn*5271*	17	*cbaABC*	3-Chlorobenzoate	*Comamonas testosteroni* BR60 (pBRC60)	Wyndham *et al.*, 1988
	Tn*5271*-like	16	*cbaABC*	Chlorobenzoates	*Alcaligenes* sp. CPE3 (pCPE3)	Gioia *et al.*, 1998
	Tn*Had1*	8,9	*dehH1*	Haloacetate	*Delftia acidovorans* B (pUO1)	Sota *et al.*, 2002
	Tn*5530*	41	*tfd*, *mdc*	Malonate, acide 2,4-dichlorophenoxyacétique	*Burkholderia cepacia* 2a (pIJB1)	Xia *et al.*, 2001 ; Poh *et al.*, 2002
Transposons non composites	Tn*4651*	56	*xyl*	Xylène, toluène	*P. putida* mt-2 (pWW0)	Tsuda et Iino, 1987 ; Tsuda *et al.*, 1989
	Tn*4653*	70	*xyl*	Xylène, toluène	*P. putida* mt-2 (pWW0)	Tsuda et Iino, 1988 ; Tsuda *et al.*, 1989
	Tn*4656*	39	*xyl*	Xylène, toluène	*P. putida* MT53 (pWW53)	Keil *et al.*, 1985 ; Tsuda et Genka, 2001
	Tn*4655*	38	*nah*	Naphtalène	*P. putida* G7 (NAH7)	Tsuda et Iino, 1990 ; Tsuda, 1996
	Tn*Had2*	16	*dehH1*, *dehH2*	Haloacétate	*Delftia acidovorans* B (pUO1)	Sota *et al.*, 2002
	Tn*4676*	73	*carABCDEF*, *antABC*	Carbazole	*P. resinovorans* CA10 (pCAR1)	Maeda *et al.*, 2003

Un des exemples bien connus d'îlots génomiques impliqués dans le catabolisme de composés xénobiotiques est l'élément *clc*, identifié pour la première fois au sein du génome de la souche *Pseudomonas* sp.B13 vivant dans les eaux usées ou d'égouts. D'une taille de 103 kb et présent en deux copies sur le chromosome en aval de gènes codant des ARNt^{Gly}, l'élément *clc* est composé de 107 ORFs dont le pourcentage global en paires de bases guanine et cytosine est de 62,5 %, avec cependant des variations individuelles de 52,1 % à 72 % (Gaillard *et al.*, 2006). L'annotation de cet îlot révèle deux régions distinctes (figure 13).

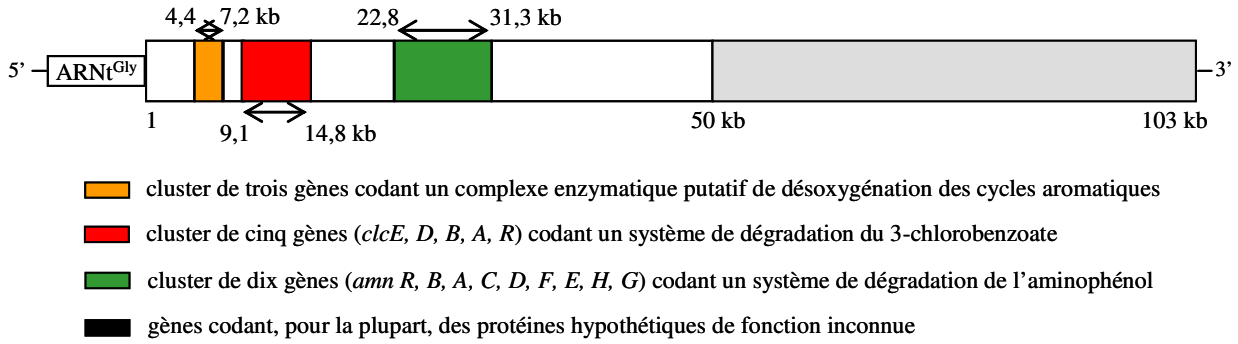


Figure 13 : Carte génétique de l'élément *clc,* d'après Gaillard *et al.*, 2006.

La première région, comprenant les 50 premiers kb immédiatement en aval de l'ARNt^{Gly}, est essentiellement impliquée dans des fonctions cataboliques. La deuxième région comprend les 50 derniers kb et code des protéines aux fonctions encore inconnues pour la plupart. Une des hypothèses formulées est que cette deuxième région serait impliquée dans la mobilité de l'îlot. Plusieurs observations appuient en effet le fait que l'îlot *clc* ait conservé ses fonctions de mobilité. Il a été montré tout d'abord qu'il pouvait s'auto-transférer vers des souches placées en eaux contaminées au 3-chlorobenzoate et qui ne le possèdent pas (Springael *et al.*, 2002). Ensuite, de nombreux éléments qui lui sont similaires ont été identifiés chez d'autres bactéries dégradant les composés aromatiques chlorés et d'origines géographiques diverses, indiquant sa large distribution (van der Meer et Sentchilo, 2003).

2.2.3. Métabolisme des métaux lourds

Les métaux lourds tels que le mercure, l'arsenic, le cadmium, le plomb ou le nickel, naturellement présents sous forme de minerais dans la croûte terrestre, peuvent, une fois libérés, devenir très toxiques pour les organismes vivants. Les activités humaines ne cessent d'induire leur libération en grande quantité dans l'environnement, sans connaître en contrepartie un système de détoxification rapide de ces composés persistants et non dégradables. Une des solutions possibles à ce

problème, encore à l'état de réflexion, serait une bioremédiation via l'utilisation de bactéries. En effet, de nombreuses bactéries résistantes à divers métaux lourds ont été identifiées (Silver et Ji, 1994). Cette résistance est expliquée par une capacité à convertir ces composés toxiques en dérivés non toxiques. Les différents acteurs impliqués dans cette conversion sont codés par des gènes situés la plupart du temps sur des éléments génétiques mobiles.

Un des métaux lourds pour lequel la résistance bactérienne est la mieux connue est le mercure, avec notamment l'identification de l'opéron *mer* contenant plusieurs gènes codant la transformation de l'ion mercurique Hg^{2+} en mercure volatil Hg^{0} (Nascimento et Chartone-Souza, 2003). Cet opéron *mer* a été mis en évidence dans le génome de diverses bactéries et ceci avec des structures et des localisations variables (figure 14). Ces dernières peuvent en effet être aussi bien plasmidiques et faire partie intégrante ou non d'un transposon, que chromosomiques au sein d'un transposon.

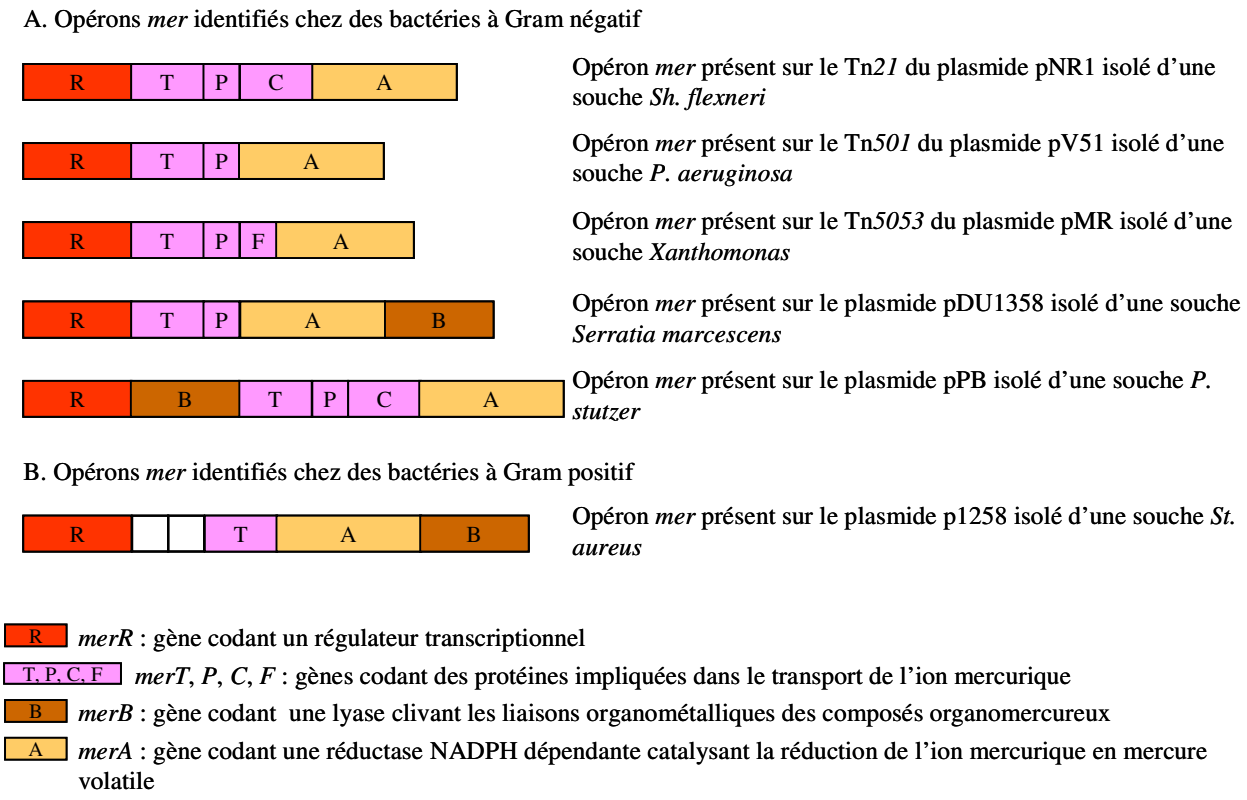


Figure 14 : Représentation schématique de différentes structures de l'opéron *mer* identifiées chez des bactéries à Gram négatif (A) et à Gram positif (B), d'après Nascimento et Chartone-Souza, 2003.

2.3. La résistance aux antibiotiques

Actuellement, un des problèmes majeurs de la médecine est la résistance des bactéries aux antibiotiques. En effet, bien que ces derniers soient très nombreux, d'origines variées et de spectres d'action très divers, il semble que les bactéries soient capables de développer une résistance contre

chacun d'entre eux. Lors des premières introductions d'antibiotiques dans le traitement de maladies infectieuses, la résistance bactérienne était peu répandue et qualifiée de naturelle. Puis, avec l'usage répété et systématique des antibiotiques, cette résistance s'est diversifiée, accrue et répandue très rapidement. Cette deuxième forme de résistance bactérienne est qualifiée d'acquise et peut reposer sur différents modes opérationnels. Créée par plusieurs évènements de mutations géniques et chromosomiques, la résistance acquise des bactéries aux antibiotiques s'est surtout très répandue grâce au processus du transfert génétique horizontal. Les gènes de résistance sont en effet très fréquemment retrouvés sur des éléments génétiques mobiles. Parmi ces derniers, les intégrons et les transposons apparaissent comme les premiers porteurs de fonctions de résistance, eux-mêmes étant situés dans la plupart des cas au sein d'îlots génomiques ou d'autres transposons à localisation plasmidique ou chromosomique, comme l'illustrent les quatre exemples de résistance bactérienne aux antibiotiques qui vont être présentés ici.

2.3.1. L'élément SXT

L'élément SXT est un transposon conjugatif porteur de fonctions de résistance à quatre antibiotiques : le sulfaméthoxazole, la triméthoprime, le chloramphénicol et la streptomycine. Initialement découvert chez une souche *Vibrio cholerae* de sérogroupe O139 (Waldor *et al.*, 1996), cet élément est désormais retrouvé chez pratiquement tous les isolats cliniques de cette espèce bactérienne (Hochhut *et al.*, 2001). De localisation chromosomique, ce transposon conjugatif présente une taille d'environ 100 kb, 87 ORFs potentielles et un pourcentage en paires de bases guanine et cytosine de 47,1 %, proche de celui du génome total qui est de 47,6 % (Beaber *et al.*, 2002). Cette proximité de pourcentage cache en fait une grande variation au sein même du transposon conjugatif, où des taux en paires de bases guanine et cytosine de 42 à 51,9 % sont rencontrés, soulignant sa structure chimérique, c'est-à-dire sa création par plusieurs avènements de transfert génétique horizontal. Ainsi sont identifiés dans l'élément SXT des gènes de plasmides, de bactériophages et d'autres origines encore inconnues dans la plupart des cas. Ces différents gènes semblent être regroupés dans des clusters de même fonction, donnant une organisation modulaire à l'élément (figure 15).

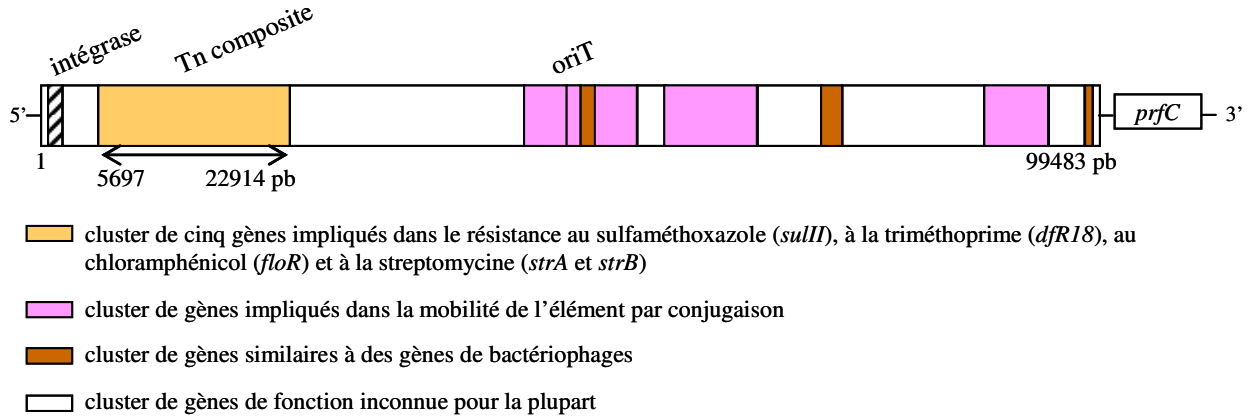


Figure 15 : Carte génétique de l'élément SXT, d'après Beaber *et al.*, 2002.

Vers l'extrémité 5' de ce dernier, une région d'environ 17 kb rassemble, entre autre, les cinq gènes impliqués dans la résistance aux quatre antibiotiques cités précédemment. Cernée par deux gènes codant des transposases, cette région de résistance peut être considérée comme un transposon composite à elle seule, possédant le plus haut pourcentage en paires de bases guanine et cytosine (51,9 %). Dans la moitié 3' de l'élément SXT, trois opérons renfermant des gènes homologues aux gènes impliqués dans le transfert par conjugaison des plasmides R27 et F isolés chez *Salmonella enterica* serovar Typhi et *E. coli*, respectivement, sont retrouvés. Il a été montré que cette région de transfert est toujours fonctionnelle : l'élément SXT peut se propager par conjugaison, après s'être excisé du chromosome pour passer à un état circulaire cytoplasmique incapable d'auto-réplication. Cette excision ainsi que la nouvelle intégration chromosomique sont catalysées par une seule enzyme, une intégrase, codée par le premier gène situé en 5' de l'élément. L'intégration semble d'ailleurs être site-spécifique, jusque là toujours retrouvée en 5' du gène *prfC* (Hochhut et Waldor, 1999).

2.3.2. L'élément SGI1

L'élément SGI1 est un îlot génomique porteur de fonctions de résistance à sept antibiotiques : l'ampicilline, le chloramphénicol, le florfénicol, la streptomycine, la spectinomycine, le sulfonamide et la tétracycline. Cet îlot a été initialement identifié au sein du génome de la souche *Salmonella enterica* serovar Typhimurium DT104 (Boyd *et al.*, 2000). De localisation chromosomique, il présente une taille d'environ 43 kb, 44 ORFs potentielles et un pourcentage en paires de bases guanine et cytosine de 49,17 %, tandis que celui du génome total est de 51 à 53 % (Boyd *et al.*, 2001). Comme dans le cas de l'élément SXT, ce pourcentage global cache des disparités à l'intérieur de l'îlot, illustrant sa structure chimérique et suggérant une nouvelle fois sa création par plusieurs avènements de transfert génétique horizontal. Les fonctions de résistance aux multiples antibiotiques sont rassemblées dans une région couvrant les 13 derniers kb de l'îlot, appelée MDR (pour Multi Drug Resistance) (figure 16). Cette région MDR est composée de deux intégrons de classe 1. Son

pourcentage en paires de bases guanine et cytosine de 58,7 % et la présence à ses extrémités 5' et 3' d'une résolvase et d'une transposase, respectivement, suggère que la région MDR ait été acquise par l'îlot SGI1 lors d'un évènement de transposition.

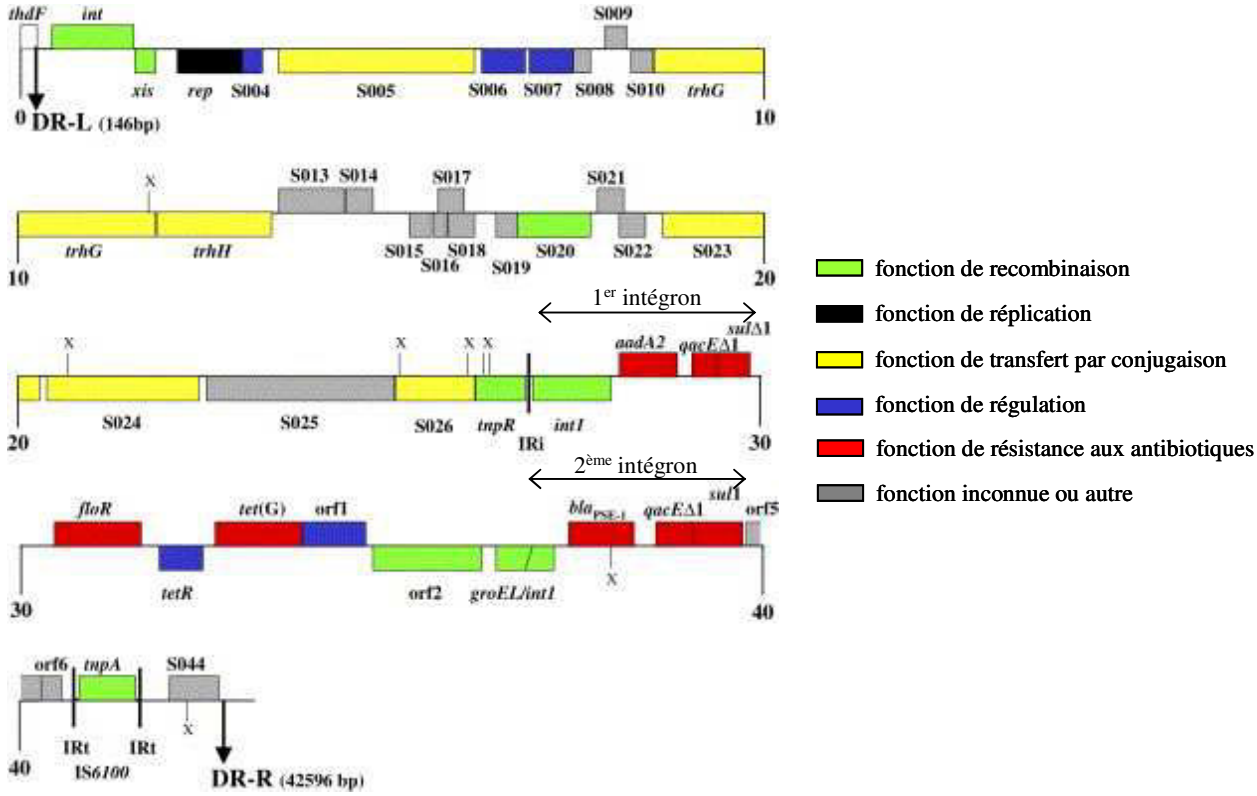


Figure 16 : Carte génétique de l'élément SGI-1, d'après Mulvey *et al.*, 2006. Les gènes *tnpR* et *tnpA* délimitent la région MDR composée notamment de deux intégrons.

int : intégrase ; *xis* : excisionase ; rep : protéine de réplication ; *tnpR* : résolvase ; *tnpA* : transposase ; *aadA2* : gène de résistance à la streptomycine / spectinomycine ; *qacEΔ1* : gène délété de résistance aux antiseptiques de type ammonium quaternaire; *sulΔ1* : gène délété de résistance aux sulfonamides; *floR* : gène de résistance au chloramphénicol / florfénicol ; *tetR, tetG* : gènes de résistance à la tétracycline ; bla_{PSE-1} *:* gène de résistance à l'ampicilline ; *sul1* : gène de résistance aux sulfonamides.

L'îlot SGI1 a été mis en évidence au sein de nombreux autres génomes de souches *Salmonella enterica* du sérotype Typhimurium, mais également de sérotypes différents (Mulvey *et al.*, 2006). Cette dissémination montre donc une capacité de transfert. Pour cela, il a été montré que l'îlot SGI1 pouvait se transférer d'une bactérie à une autre par le biais de la conjugaison, mais d'un façon mobilisable, c'est-à-dire seulement si un plasmide conjugatif résidait dans la même cellule pour lui apporter les fonctions de mobilité nécessaires qu'il ne possède pas au sein de sa séquence (Doublet *et al.*, 2005). Chez tous les isolats où l'îlot SGI1 a pu être mis en évidence, l'insertion chromosomique est conservée en 3' du gène *thdF*. Cependant, chez ces mêmes isolats, de nombreux variants de la région MDR ont pu être identifiés, amenant à classer les îlots SGI1 de SGI1-A à SGI1-L (Mulvey *et al.*, 2006 ; Cloeckaert *et al.*, 2006). Ces variants différent dans leur contenu en gènes de résistance : certains ont perdus quelques gènes, comme l'îlot SGI1-C ne contenant plus que les gènes de résistance

à la streptomycine, à la spectinomycine et aux sulfonamides ; certains ont au contraire gagné un gène de résistance, comme l'îlot SGI1-A conférant une résistance supplémentaire à la triméthoprime ; enfin d'autres semblent avoir échangé de gènes cassettes. Il est suggéré que ces variations structurales à l'intérieur de la région MDR se soient produites par des évènements de recombinaison homologue. Les intégrons de classe 1 sont en effet très répandus dans le monde bactérien, distribués à la fois sur les chromosomes et les plasmides et leurs parties 5' et 3' conservées permettent ce genre de recombinaison.

2.3.3. L'élément SCCmec

L'élément SCC*mec* est défini comme une cassette chromosomique conférant une résistance à la méthicilline aux souches de l'espèce *Staphylococcus aureus*. L'introduction de la méthicilline en vue d'une thérapie anti-staphylocoques s'est faite en 1959 ; cependant, seulement deux ans après, en 1961, les premières souches résistantes sont apparues (Hiramatsu *et al*., 2001). Cette résistance réside dans l'acquisition de l'élément SCC*mec* qui, depuis sa première apparition, s'est diversifié en cinq types différents (figure 17 et tableau 4).

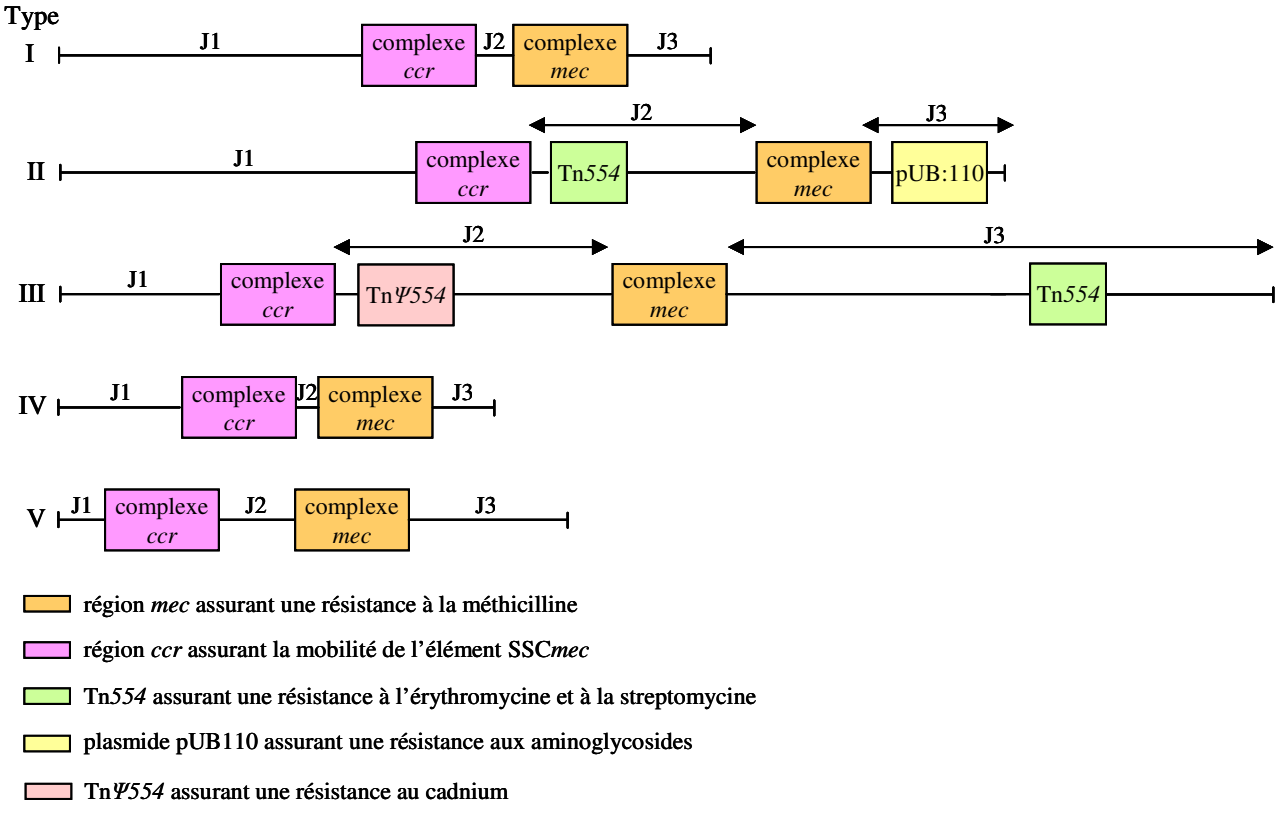


Figure 17 : Représentation schématique de différents types d'éléments SSC*mec* identifiés, d'après Martins et Cunha, 2007 et Deurenberg *et al*., 2006.

Tableau 4 : Caractéristiques des cinq types d'éléments SCC*mec* recensés, d'après Martins et Cunha, 2007 et Deurenberg *et al*., 2006.

Type	Taille (kb)	Fonctions de résistance portées
I	34,3	Méthicilline
II	53	Méthicilline + érythromycine + streptomycine + aminoglycosides
III	66,9	Méthicilline + érythromycine + streptomycine + cadnium
IV	20,9-24,3	Méthicilline
V	28	Méthicilline

Ces cinq types partagent trois caractéristiques communes : un complexe de gènes *mec*, un complexe de gènes *ccr* et une intégration chromosomique spécifique en 3' d'une ORF de fonction inconnue appelée *orfX* (Martins et Cunha, 2007). Le complexe de gènes *mec* est responsable de la résistance à la méthicilline ; il est composé d'une séquence d'insertion, du gène *mecA* codant une PBP (penicillin binding protein) et des gènes *mecR1* et *mecI* intacts ou tronqués codant des régulateurs transcriptionnels du gène *mecA*. Le complexe de gènes *ccr* assure la mobilité de l'élément SCC*mec* ; il renferme pour cela un ou deux gènes *ccr* codant une transposase catalysant à la fois l'intégration et l'excision chromosomiques de l'élément. Les cinq types différents d'éléments SCC*mec* reposent sur l'existence de cinq et quatre variants des complexes de gènes *ccr* et *mec*, respectivement. A cela s'ajoutent des différences dans les régions autres que *mec* et *ccr* de l'élément, appelées régions J : J1 pour la région séparant le complexe *ccr* de l'extrémité 3' de l'élément, J2 pour la région séparant les complexes *ccr* et *mec* et J3 pour la région séparant le complexe *mec* de l'extrémité 5' de l'élément. En effet, alors que pour les premier, quatrième et cinquième types d'éléments SCC*mec* identifiés (SCC*mec* I, IV et V), aucun gène de résistance additionnel n'est présent dans les trois régions J, les éléments SCC*mec* II et III, grâce à l'insertion de plasmides et de transposons dans ces régions, ont gagné plusieurs gènes de résistance (figure 17).

2.3.4. L'élément pNR1

L'élément pNR1 est un plasmide conjugatif porteur de fonctions de résistance à cinq antibiotiques : la tétracycline, le chloramphénicol, la streptomycine, la spectinomycine et le sulfonamide. Initialement découvert à la fin des années 1950 chez une souche de l'espèce *Sh. flexneri* au Japon (Nakaya *et al*., 1960), ce plasmide est l'archétype d'une large collection de plasmides similaires découverts par la suite à travers le monde et appelés plasmides R, de par leur capacité d'auto-transfert par conjugaison et leur propriété de résistance à de multiples antibiotiques. Sur les 94,5 kb composant le plasmide pNR1, deux tiers (62,1 kb) sont dédiés aux fonctions de réplication et

de transfert et un tiers (32,4 kb) aux fonctions de résistance à divers antibiotiques (figure 18) (Liebert *et al*., 1999).

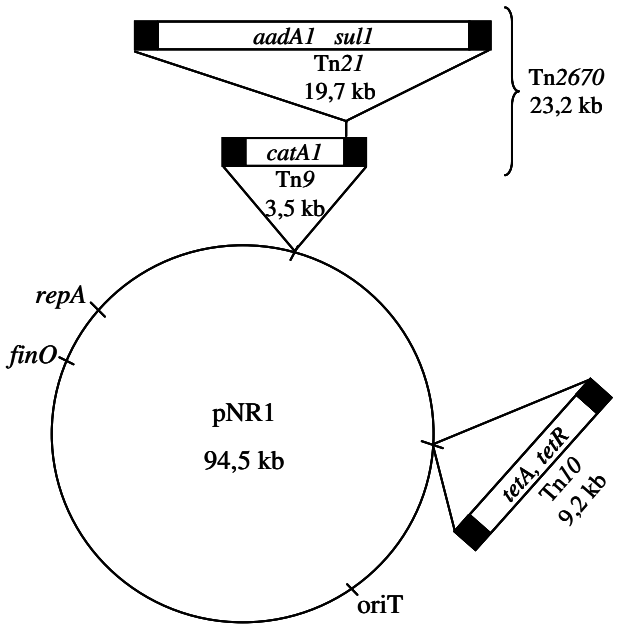


Figure 18 : Représentation schématique du plasmide pNR1, d'après Normark et Normark, 2002.

tetA : gène de résistance à la tétracycline ; *tetR* : gène régulateur ; *catA1* : gène de résistance au chloramphénicol ; *aadA1* : gène de résistance à la streptomycine / spectinomycine ; *sul1* : gène de résistance aux sulfonamides ; oriT : origine de transfert ; *finO* : gène de régulation ; *repA* : gène de réplication.

Ces fonctions de résistance sont partagées de façon inéquitable sur deux transposons composites : les transposons Tn*10* et Tn*2670*. Le transposon Tn*2670* est composé lui-même deux transposons : le transposon composite Tn*9* et le transposon non composite Tn*21* renfermant un intégron de classe 1 appelé In2.

2.4. La virulence

La notion de virulence doit être perçue ici comme le caractère pathogène de certaines bactéries. Ce caractère pathogène correspond à l'expression de facteurs capables d'induire une maladie chez des organismes hôtes. Ces facteurs peuvent être répartis dans six groupes différents selon la fonction qu'ils portent. Ainsi les facteurs d'adhésion permettent aux bactéries de s'attacher aux surfaces hôtes. Les sidérophores assurent une solubilisation et une entrée des ions Fe^{3+} dans les cellules bactériennes. Les capsules constituent une barrière protectrice contre les acteurs et les actions du système immunitaire hôte. Les exotoxines altèrent ou détruisent même les cellules hôtes en modifiant leurs structures cytoplasmiques ou leurs outils de communication. Les invasines permettent l'entrée des bactéries dans les cellules hôtes. Enfin, les systèmes de sécrétion de type III et IV peuvent être requis pour délivrer certains composants bactériens dans les cellules hôtes. La particularité de la pathogénicité bactérienne est qu'une souche d'une espèce donnée peut être pathogène sans que les autres souches de la même espèce le soient. Ceci s'explique par le fait que la majorité des gènes codant les différents facteurs de virulence décrits précédemment sont retrouvés sur des éléments

génétiques mobiles. Des exemples concernant trois types d'éléments génétiques mobiles, les plasmides, les prophages et les îlots génomiques, vont être présentés ici.

2.4.1. Les plasmides porteurs de fonctions de virulence

Les fonctions de transfert impliquées dans les évènements de conjugaison et codées sur les plasmides conjugatifs peuvent contribuées à elles seules à la virulence des bactéries. L'une d'elles, la lipoprotéine membranaire TraT, est codée par le gène *traT* présent sur les plasmides conjugatifs des bactéries à Gram négatif (Sukupolvi et O'Connor, 1990). Il a été démontré que l'expression de ce gène pouvait permettre aux bactéries de résister au sérum de l'organisme hôte (Moll *et al*., 1980) et à l'action phagocytaire de certaines de ses cellules immunitaires (Aguero *et al*., 1984). Pour cela, la lipoprotéine TraT serait notamment capable d'entraver la cascade d'activation du système du complément (Sukupolvi et O'Connor, 1990). Hormis ces gènes de mobilité, les plasmides bactériens peuvent contenir d'autres gènes impliqués dans la virulence.

Il semble exister en particulier une catégorie de bactéries pour lesquelles les plasmides jouent un grand rôle dans leur pathogénicité : les bactéries phytopathogènes. Une revue éditée en 2001 par Vivian *et al*. présente ainsi l'importance des plasmides pour six genres bactériens à l'origine de maladies chez les plantes (Vivian *et al*., 2001). Ces six genres bactériens sont *Burkholderia*, *Erwinia*, *Pantoea*, *Pseudomonas*, *Ralstonia* et *Xanthomonas*. Diverses pathologies peuvent être attribuées à ces bactéries, telles que la rouille, la galle, la moisissure, la pourriture, et ceci chez de nombreuses plantes comme la tomate, l'aubergine, le poivre, le coton, la pomme, la poire, l'oignon, la banane, la pomme de terre, le maïs et le blé. Au sein du genre *Pseudomonas*, un facteur de virulence, appartenant au groupe des exotoxines et dont le gène codant est très souvent de localisation plasmidique, est particulièrement répandu : la coronatine. Grâce à cette dernière, les bactéries semblent pouvoir induire de larges lésions sur les tissus végétaux et se développer plus rapidement. La production de cette toxine, concomitante à l'infection des plantes, est dictée par un cluster de gènes *cor* d'une taille d'environ 30 kb (Bender *et al*., 1996).

2.4.2. Les prophages porteurs de fonctions de virulence

Des bactériophages porteurs de facteurs de virulence ont été découverts dès le milieu du XX^{ème} siècle. Deux exemples sont les bactériophages gamma et C1, codant des facteurs de virulence clés pour les bactéries *Corynebacterium diphteriae* et *Clostridium botulinum* (Freeman, 1951 ; Barksdale et Arden, 1974). Ensuite, avec la découverte du bactériophage CTXΦ porteur de la toxine du choléra (Waldor et Mekalanos, 1996), il est très vite apparu que les bactériophages jouent un rôle très

important dans la virulence de nombreuses bactéries pathogènes. Le tableau 5 présente ainsi quelques exemples de facteurs de virulence codés par des gènes présents dans des génomes de bactériophages et les bactéries hôtes où ces derniers peuvent s'installer sous forme de prophages. Ces gènes de virulence ne sont pas indispensables au cycle de vie des bactériophages. Ils appartiennent plus généralement à une catégorie de gènes appelés morons, ou encore gènes de conversion lysogénique, (Brüssow *et al*., 2004), responsables d'un changement phénotypique des bactéries qui les accueillent. La virulence de nombreuses bactéries pathogènes pour l'homme provient ainsi de la présence de morons. Les exemples les mieux connus sont les bactéries *Salmonella enterica*, *Vibrio cholerae*, *E. coli*, *Clostridium botulinum*, *S. pyogenes*, *St. aureus* et *Corynebacterium diphteriae* responsables entre autre d'intoxications alimentaires, du choléra, d'infections intestinales et urinaires, du botulisme, de la scarlatine, de nombreuses infections nosocomiales et de la diphtérie, respectivement.

Tableau 5 : Exemples de facteurs de virulence codés par des gènes d'origine bactériophagique, d'après Brüssow *et al*., 2004. O-Ag : O-Antigène ; OMP : outer membrane protein ; *C.* : *Corynebacterium* ; *Cl.* : *Clostridium* ; *V.* : *Vibrio* ; *N.* : *Neisseria* ; *Sa.* : *Salmonella* .

Facteurs de virulence		**Gènes**	**Bactériophages porteurs**	**Bactéries hôtes**
Toxines extracellulaires	Toxine de la diphtérie	*tox*	β-phage	*C. diphteriae*
	Neurotoxine	C1	Phage C1	*Cl. botulinum*
	Shiga toxines	*stx1, stx2*	H-19B	*E. coli*
	Entérohémolysine	*hly2*	ΦFC3208	*E. coli*
	Cytotoxine	*ctx*	ΦCTX	*P. aeruginosa*
	Entérotoxine	*see*, *sel*	NA	*St. aureus*
	Entérotoxine P	*sep*	ΦN315	*St. aureus*
	Entérotoxine A	*entA*	Φ13	*St. aureus*
	Entérotoxine A	*sea*	ΦMu50A	*St. aureus*
	Toxine exfoliative A	*eta*	ΦETA	*St. aureus*
	Toxine de type A	*speA*	T12	*S. pyogenes*
	Toxine de type C	*speC*	CS112	*S. pyogenes*
	Toxine du choléra	*ctxAB*	CTXΦ	*V. cholerae*
	Leucocidine	*pvl*	fPVL	*St. aureus*
	Superantigènes	*speA1, speA3, speC, speI, speH, speM, speL, speK, ssa*	8232.1	*S. pyogenes*
Protéines altérant l'antigénicité	Protéines membranaires	Mu-like	Pnm1	*N. meningitidis*
	Protéines glycosylées	*rfb*	ε^{34}	*Sa. enterica*
	Protéines glycosylées	*gtr*	P22	*Sa. enterica*

Tableau 5 : Suite.

Facteurs de virulence		Gènes	Bactériophages porteurs	Bactéries hôtes
Protéines altérant l'antigénicité	O-Ag acétylase	*oac*	Sf6	*Sh. flexneri*
	Glycosyl transférase	*gtrII*	SfII, SfV, SfX	*Sh. flexneri*
Invasines	Protéine effectrice de type III	*sopE*	SopEΦ	*Sa. enterica*
	Protéine effectrice de type III	*sseI*	GIFSY-2	*Sa. enterica*
	Protéine effectrice de type III	*sspH1*	GIFSY-3	*Sa. enterica*
Enzymes	Superoxide dismutase	*sodC*	Sp4, Sp10	*E. coli* O157
	Superoxide dismutase	*sodC-I*	GIFSY-2	*Sa. enterica*
	Superoxide dismutase	*sodC-III*	Fels-1	*Sa. enterica*
	Neuraminidase	*nanH*	Fels-1	*Sa. enterica*
	Hyaluronidase	*hylP*	H4489A	*S. pyogenes*
	Leucocidine	*pvl*	ΦPVL	*St. aureus*
	Staphylokinase	*sak*	Φ13	*St. aureus*
	Phospholipase	*sla*	315.4	*S. pyogenes*
	DNase/streptodornase	*sdn, sda*	315.6, 8232.5	*S. pyogenes*
Facteurs de résistance au sérum	OMP	*bor*	λ	*E. coli*
	OMP	*eib*	λ-like	*E. coli*
Adhésines	Vir	*vir*	MAV1	*Mycoplasma arthritidis*
	Protéines de capside bactériophagique	*pblA, pblB*	SM1	*S. mitis*

2.4.3. Les îlots génomiques porteurs de fonctions de virulence

Les îlots génomiques porteurs de facteurs de virulence sont plus souvent appelés îlots de pathogénicité ou PAIs pour "pathogenicity islands" (Hacker *et al.*, 1990). Ce terme est désormais couramment utilisé pour décrire les régions génomiques de certaines bactéries pathogènes où sont rassemblés plusieurs gènes de virulence de manière continue. Des îlots de pathogénicité ont pu être identifiés dans le génome de diverses souches bactériennes éloignées d'un point de vue phylogénétique, à la fois à Gram négatif et à Gram positif et pathogènes pour l'homme, les animaux ou les plantes. La liste des PAIs caractérisés à ce jour est déjà longue ; le tableau 6 présente les propriétés de certains d'entre eux. Il est intéressant de noter que pour plus des deux-tiers des exemples présentés dans ce tableau, le gène cible d'insertion est un gène codant un ARNt, site prédominant

d'insertion lors d'évènements de recombinaison site-spécifique comme il l'a été dit dans le paragraphe 1.3.2 de ce chapitre. Pour le tiers restant, le gène cible d'insertion est un gène codant une protéine. En général, un PAI donné est spécifique à une souche bactérienne pathogène donnée. Par exemple, les PAIs I_{536} et III_{536} sont jusqu'à présent exclusivement trouvés dans le génome de la souche *E. coli* uropathogène 536 et les PAIs I_{CFT073} et II_{CFT073} dans celui de la souche *E. coli* uropathogène CFT073 (Guyer *et al*., 1998). Toutefois, quelques PAIs peuvent être partagés par plusieurs souches bactériennes. L'exemple bien connu est le HPI (pour "High Pathogenicity Island"), initialement découvert dans le génome de souches du genre *Yersinia* (Carniel *et al*., 1996), puis dans celui de souches des espèces *E. coli* (Schubert *et al*., 1998), *Citrobacter diversus* et de différentes espèces du genre *Klebsiella* (Bach *et al*., 2000). Enfin, une souche bactérienne pathogène peut posséder dans son génome plus d'un PAI : jusqu'à 12 et 4 PAIs ont pu être détectés chez des souches *Salmonella enterica* et la souche *E. coli* 536, respectivement.

Tableau 6 : Caractéristiques de certains îlots de pathogénicité bactériens, d'après Schmidt et Hensel, 2004. * : gène codant un ARNt. PAI : pathogenicity island ; LEE : locus of enterocyte effacement ; CNF : cytotoxic necrotising factor ; PTS : phosphotransferase system ; ABC : ATP binding cassette ; Ag : antigène ; LPS : lipopolysaccharide ; *Y.* : *Yersinia* ; *Sa.* : *Salmonella* ; *He.* : *Helicobacter* ; *V.* : *Vibrio* ; *N.* : *Neisseria* ; *Li.* : *Listeria* .

Souche/espèce détentrice	Nom	Taille (kb)	Gène cible d'insertion	GC%	Facteurs de virulence
***E. coli* 536**	PAI I_{536}	76,8	*selC**	46	α-hémolysine, fimbriae, adhésine
	PAI II_{536}	102,2	*leuX**	46	α-hémolysine, P-fimbriae, adhésines
	PAI III_{536}	68,1	*thrW**	47	S-fimbriae, adhésine, système de capture du fer, protéase
	PAI IV_{536}	30,2	*asnT**	57	Système de capture du fer
***E. coli* J96**	PAI I_{J96}	>170	*pheV**	Non publié	α-hémolysine, P-pilus
	PAI II_{J96}	>110	*pheU**	Non publié	α-hémolysine, CNF1, hémagglutinine
***E. coli* CFT073**	PAI I_{CFT073}	123	*pheV**	47	α-hémolysine, P-fimbriae, aérobactine
	PAI II_{CFT073}	52	*pheU**	48	P-fimbriae
***E. coli* C5**	PAI I_{C5}	Non publié	*leuX**	Non publié	α-hémolysine, CNF1, hémagglutinine
***E. coli* RS218**	GimA	20,3	*yjiD/yiiE*	46,2	Enzymes pour le métabolisme du glycérol, système PTS
***E. coli* AL862**	PAI I_{AL862}	61	*pheR**	46,4	AfaE-VIII adhésine, enzymes pour le métabolisme du désoxyribose, système PTS
	PAI II_{AL862}	Non publié	*pheV**	Non publié	AfaE-III adhésine
***E. coli* EDL933**	LEE	43,6	*selC**	40,9	Système de sécrétion de type III

Tableau 6 : Suite.

Souche/espèce détentrice	Nom	Taille (kb)	Gène cible d'insertion	GC%	Facteurs de virulence
***E. coli* 68-24**	TAI	>8	Non pubié	Non publié	Adhésine, protéines pour la résistance au tellure
***E. coli* EDL933**	OI#7	35,1	*aspV**	Non publié	Toxine
	OI#28	25,1	*ybaT*	Non publié	Toxine, système de transport
	OI#43	87,6	*clp*	Non publié	Uréase
	OI#47	31,7	*ycdT*	Non publié	Adhésine
	OI#48	87,5	*ycdU*	Non publié	Uréase
	OI#115	16,9	Z4179	Non publié	Invasines
	OI#122	23,4	*pheV**	Non publié	
	OI#148	43,4	*selC**	Non publié	LEE
***E. coli* E2348/69**	LEE	35,6	*selC**	38,3	Intimine, système de sécrétion de type III
	EspC	15,2	*ssrA**	40,5	Protéase
***E. coli* RIMD-0509952**	SPLE1	86,2	*serX**	Non publié	=TAI
	SPLE2	13,4	*yeeX*	Non publié	Non publié
	SPLE3	23,4	*pheI**	Non publié	
	SPLE4	43,4	*selC**	Non publié	=LEE
	SPLE5	10,2	*leuX**	Non publié	
	SPLE6	34,1	*leuX**	Non publié	
***E. coli* RDEC-1**	LEE	34	*lifA*	Non publié	Intimine, système de sécrétion de type III
***E. coli* RW1372**	LEE	80	*pheV**	Non publié	Intimine, système de sécrétion de type III
***E. coli* HT10407**	Tia-PAI	46	*selC**	43,7	Invasine
Diverses *E. coli*	HPI	Non publié	*asnT**	Non publié	Système de capture du fer
***Citrobacter rodentium* DBS100**	LEE	36,1	IS679/ABC	38	Intimine, système de sécrétion de type III
***Y. pestis* 6/69M**	HPI	119	*asnT**	59	Système de capture du fer
***Y. enterocolitica* YE8081**	HPI	43	*asnT**	60	Système de capture du fer
***Y. pseudotuberculosis* IP32637**	HPI	36	*asnT**, *asnU**, *asnW**	60	Système de capture du fer
***Sa. enterica* serovar Typhi**	SPI-7	134-146	*pheU**	Non publié	Ag, pili de type IV
***Sa. enterica* DT104**	SGI	152,4	*thdF*	41-58	Protéines pour une résistance à de multiples antibiotiques

Tableau 6 : Suite.

Souche/espèce détentrice	Nom	Taille (kb)	Gène cible d'insertion	GC%	Facteurs de virulence
***Sa. enterica* serovar Typhimurium**	SPI-1	40	*fhlA*	42	Invasines, système de sécrétion de type III
	SPI-2	40	*valV**	44,3-55,5	Protéines pour une prolifération intracellulaire, système de sécrétion de type III
	SPI-3	17	*selC**	47,5	Protéines pour une prolifération intracellulaire, système de capture d'ions Mg^{2+}
	SPI-4	25	*ssb/soxSR*	37-54	Toxine, système de sécrétion de type I
	SPI-5	7	*serT**	43,6	Protéines effectrices
***Sa. enterica* ssp. III, IV**	HPI	Non publié	*asnT**	Non publié	Système de capture du fer
***Sa. enterica* serovar Typhi CT18**	SPI-6	59	*asp**	Non publié	Fimbriae
	SPI-8	6,8	*phe**	Non publié	Pseudogènes
	SPI-9	16	*ssrA**	Non publié	Toxine, système de sécrétion de type I
	SPI-10	33	*leu**	Non publié	Fimbriae
***P. aeruginosa* PAO1**	PAGI-1	48,9	PAO2217	54,9-63,7	Déhydrogénase, protéines de détoxification d'entités oxigénées
***P. aeruginosa* C**	PAGI-2(C)	105	*gly**	64,8	Protéines de transport de métaux lourds et de résistance au mercure
***P. aeruginosa* SG17M**	PAGI-3(SG)	103,3	*glu**	59,8	Protéines de transport d'acides aminés, coenzymes, porphyrines
***P. aeruginosa* PAK**		16	*flgL/fliC*	63,3	Protéines pour la glycosylation de flagelline
P. syringae	HRP	21	*leu**	46-59	Protéines effectrices
Dichelobacter nodosus	*vap*	11,9	*serV**	Non publié	Antigènes
	vrl	27	Non publié	Non publié	Antigènes
He. pylori	*cag*	37-40	*glr*	35	Système de sécrétion de type IV, Protéine CagA
***Sh. flexneri* Y53**	SHI-O	10,6	*thrW**	40	Protéines pour la synthèse du LPS
***Sh. flexneri* YSH6000**	SRL	66,2	*serX**	49,8	Protéines pour une résistance à de multiples antibiotiques
***Sh. flexneri* YSH6000T**	SHI-1	46,6	*pheV**	49	Toxines, protéases
***Sh. flexneri* SA100/M90T**	SHI-2	23,8	*selC**	48,6	Système de capture du fer
***Sh. boydii* O1392**	SHI-3	21	*pheU**	51	Système de capture du fer

Tableau 6 : Suite.

Souche/espèce détentrice	Nom	Taille (kb)	Gène cible d'insertion	GC%	Facteurs de virulence
***V. cholerae* EL Tor O1**	VPI-1	40	*ssrA**	35	Adhésine
	VPI-2	57,3	*ser**	42	Neuraminidase
	VSP-1	16	VC0174/VC0186	40	Phospholipase A
	VSP-2	7,5	Non publié	41	Non publié
N. gonorrhoeae	*atlA*	8,9	Non publié	43	Hydrolase de peptidoglycane
N. meningitidis	*exl1*	1,9-6	Non publié	Non publié	Lipoprotéine
Porphyromonas gingivalis	*rag*	9,5	IS*1126*	41	Récepteur d'antigène
Bacteroides fragilis	BFPAI	6	*bfmB/bfmC*	35	Toxine
Li. monocytogenes	LIPI-1	9	Non publié	Non publié	Phospholipase C, lysteriolysine O
Li. ivanovii	LIPI-2	22	Non publié	Non publié	Sphingomyelinase
St. aureus	SCC*mec*	36-53	Non publié	Non publié	Protéines pour une résistance à de multiples antibiotiques
	SaPI1	15,2	*tyrB*	35	Superantigène
	SaPI2	Non publié	Non publié	Non publié	Superantigène
	SaPI3	15,9	Non publié	41,4	Toxines
	SaPIbov	15,9	*gmps*	Non publié	Toxines, superantigène
	*v*Saα	25-40	Non publié	Non publié	Toxines
	*v*Saβ	30-40	Non publié	Non publié	Toxines, superantigène
	etd	9	Non publié	Non publié	Toxines exfoliatives
	*v*Sa3	Non publié	Non publié	Non publié	Toxines
	*v*Sa4	Non publié	Non publié	Non publié	Toxines, superantigène
***S. pneumoniae* 0100993**	PPI-1	27,2	OrfB	32,6	Système de capture du fer

2.5. Conclusion

Les différents exemples qui viennent d'être exposés montrent que les éléments génétiques bactériens mobiles ou mobilisables sont porteurs de fonctions très diverses et qu'ils sont largement répandus dans le monde bactérien. Chaque type d'élément présenté dans le paragraphe 1.1 de ce chapitre (plasmide, transposon, intégron, prophage et îlot génomique) participe à la mobilité de ces diverses fonctions. Par ailleurs, il semble que ces derniers s'associent très fréquemment et forment ainsi des structures complexes et chimériques, capables de véhiculer de plus grandes informations.

3. Les éléments génétiques mobiles ou mobilisables et leur impact sur la diversification de l'espèce *Escherichia coli*

La bactérie désormais connue sous le nom *Escherichia coli* a été décrite pour la première fois par un pédiatre allemand, Theodor Escherich, à la fin du XIX^{ème} siècle (Escherich, 1885). Lors d'une série d'études pionnières sur la flore intestinale des nouveaux-nés, ce médecin décrivit un microbe présent chez des individus sains, qu'il nomma *Bacterium coli* commune et qui, par la suite, fut renommé en son honneur *Escherichia coli* (Sojka, 1965). *E. coli* est la première espèce bactérienne commensale intestinale anaérobie facultative décrite. Certaines souches de cette espèce ont été pour cela très utilisées dans la création de traitements probiotiques contre une variété de maladies intestinales, allant des diarrhées infectieuses aux ulcères. Elle a également joué un rôle important dans le développement de la biologie moléculaire, servant comme premier organisme modèle pour la compréhension des aspects fondamentaux de la régulation génétique. De nos jours encore, elle est un outil essentiel de la biologie moderne, utilisée par exemple comme hôte de clonage. Cependant, ce sont les effets nuisibles des bactéries *E. coli* qui ont suscité le plus d'intérêt auprès des microbiologistes et du grand public. En effet, certaines d'entres-elles peuvent être responsables de différentes maladies, parfois fatales. En 1945, John Bray, également pédiatre, isola les premières souches pathogènes d'*E. coli*, à l'origine de diarrhées chez des nouveaux-nés (Bray, 1945). Depuis, sont recensés au sein de cette espèce, de nombreux pathotypes différents pour l'homme et les animaux. Le tableau 7 présente les caractéristiques des six principaux pathotypes intestinaux et des deux principaux pathotypes extra-intestinaux (ExPEC) recensés chez l'homme (Kaper *et al*., 2004), sachant que des pathotypes similaires ou presque sont retrouvés chez les animaux, mais ceux-ci sont encore mal définis (problème de redondance et de confusion de nomenclature). Ainsi, *E. coli* est une espèce caractérisée par une grande diversité, renfermant aussi bien des souches commensales intestinales pour l'homme et les animaux à sang chaud, que des souches hautement pathogènes pour ces mêmes organismes vivants. L'analyse comparative de génomes de différentes souches *E. coli* peut facilement expliquer cette diversité. Tous semblent en effet d'une part partager un pool de gènes aux fonctions cellulaires majoritairement essentielles et d'autre part posséder des gènes spécifiques à leur niche écologique (Dougan *et al*., 2001).

Tableau 7 : Caractéristiques cliniques, pathologiques et épidémiologiques des maladies causées par les principaux pathotypes intestinaux et extra-intestinaux humains de l'espèce *E. coli*, d'après Robins-Browne et Hartland., 2002. ETEC : *E. coli* entérotoxinogène ; EIEC : *E. coli* entéroinvasive ; EPEC : *E. coli* entéropathogène ; EHEC : *E. coli* entérohémorragique ; EAEC : *E. coli* entéroagrégative ; DAEC : *E. coli* à adhésion diffuse ; UPEC : *E. coli* uropathogène ; MNEC : *E. coli* associées à des méningites chez les nouveaux nés.

	Pathotypes	Signes cliniques	Signes pathologiques	Épidémiologie
Intestinaux	**ETEC**	Diarrhée d'eau	Néant	Enfants et touristes des pays en voie de développement
	EIEC	Dysenterie	Inflammation de la muqueuse du côlon	Toutes les personnes sont susceptibles
	EPEC	Gastro-entérite	Lésion de la muqueuse intestinale	Enfants de moins de deux ans dans les pays en voie de développement
	EHEC	Diarrhée avec traces de sang	Lésion de la muqueuse intestinale et nécrose des cellules épithéliales	Enfants et personnes âgées des pays développés
	EAEC	Diarrhée persistante	Inflammation de la muqueuse intestinale	Enfants et touristes des pays en voie de développement
	DAEC	Diarrhée d'eau	-	Enfants entre deux et cinq ans dans les pays en voie de développement
Extra-intestinaux	**UPEC**	Infections urinaires	Inflammation de la vessie, du rein ou du bassinet	Toutes les personnes sont susceptibles
	MNEC	Méningites	Inflammation des méninges	Nouveaux nés et patients de neurochirurgie

3.1. Situation de l'espèce *E. coli* au sein de la famille *Enterobacteriaceae*

Au sein du règne des bactéries, *E. coli* est l'espèce type de la famille *Enterobacteriaceae* créée en 1937 par Rahn. Dans la hiérarchie de la classification bactérienne, cette famille appartient à l'embranchement *Proteobacteria*, à la classe *Gamma Proteobacteria* et à l'ordre *Enterobacteriales*. Elle est l'une des plus importantes familles de bactéries, autant d'un point de vue quantitatif que qualitatif. Elle regroupe ainsi 49 genres très ubiquitaires, allant des commensales (comme *Proteus*, *Klebsiella*, ou *Escherichia* dont l'espèce *E. coli* constituerait 80 % de la flore aérobie intestinale humaine) aux saprophytes (comme *Proteus*, *Enterobacter*, ou *Serratia*), en passant par des pathogènes pour les plantes (comme *Erwinia*, ou *Pantoea*), les hommes et les animaux à sang chaud (comme *Salmonella*, *Yersinia*, *Serratia*, *Klebsiella*, ou *Escherichia*). Sept critères définissent classiquement les bactéries de cette famille, également appelées entérobactéries : un Gram négatif, une culture facile, l'absence d'oxydase, la capacité à réduire les nitrates, une aéro-anaérobie facultative, la capacité à fermenter le glucose et leur immobilité ou leur mobilité grâce à la possession d'une ciliature

péritriche. Les différences entre les nombreux genres et espèces viennent de critères plus précis, comme la capacité à fermenter différents sucres, la production ou non de sulfures, la présence ou l'absence d'enzymes métaboliques spécifiques, ou la réalisation d'un type de fermentation bien précis. En janvier 2008, étaient dénombrés cent quatre génomes d'entérobactéries entièrement séquencés ou en cours de séquençage sur le site NCBI Microbial Genome (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html). Trente et un, huit et vingt quatre d'entre eux appartenaient à des souches des genres *Escherichia*, *Shigella* et *Salmonella*, respectivement.

3.1.1. Proximité des genres *Escherichia*, *Shigella* et *Salmonella*

Au sein de la famille *Enterobacteriaceae*, *Salmonella* est le genre le plus proche d'un point de vue phylogénétique du genre *Escherichia* ; tous deux se seraient séparés d'un ancêtre commun il y a 100 ou 140 millions d'années (Ochman et Wilson, 1987 ; Doolittle *et al*., 1996). La proximité de ces deux genres est illustrée par le fait que les données concernant leur ARNr 16S ne permettent pas de différencier certaines de leurs souches. Les séquences nucléotidiques du gène codant l'ARNr 16S de trois souches *Salmonella enterica*, *Salmonella enterica* serovar Enteriditis ES22, *Salmonella enterica* serovar Paratyphi A ATCC 54388 et *Salmonella enterica* serovar Paratyphi B P1, présentent en effet 99,6 %, 97,4 % et 97,2 % d'identité, respectivement, avec la séquence nucléotidique du gène codant l'ARNr 16S de la souche *E. coli* ATCC 25922 (Fukushima *et al*., 2002). Or, d'après Stackebrandt et Goebel, il est admis que deux souches qui présentent moins de 97 % d'identité entre les séquences des gènes codant leurs ARNr 16S, n'appartiennent pas à la même espèce bactérienne (Stackebrandt et Goebel, 1994). Ainsi d'autres critères moléculaires doivent être utilisés pour distinguer les souches des genres *Escherichia* et *Salmonella*. L'un d'entre eux est le taux d'hybridation de leurs ADNs. Deux souches appartiennent à la même espèce si le taux d'hybridation ADN-ADN de leurs génomes dépasse 70 % (Wayne *et al*., 1987). Konstantinidis et Tiedje ont montré que ce taux de 70 % d'hybridation ADN-ADN correspond approximativement à un taux moyen de 95 % d'identité nucléotidique des séquences génomiques partagées par deux souches (Konstantinidis et Tiedje, 2005). Or, les séquences génomiques partagées entre différentes souches *E. coli* et différentes souches *Salmonella* ne montrent pas plus de 80 % d'identité nucléotidique en moyenne (Konstantinidis *et al.*, 2006), ce qui approuve donc la distinction de ces deux genres. La séparation des genres *Salmonella* et *Escherichia* semble s'être opérée par plusieurs avènements d'inversions et de délétions chromosomiques et surtout par des acquisitions de nouvelles séquences d'ADN par transfert génétique horizontal (Wain *et al*., 2001).

Le genre *Shigella* a été introduit pour la première fois en 1898 par Kiyoshi Shiga à la suite de la découverte d'un microbe responsable de la dysenterie bacillaire (Shiga, 1898). Bien que les

ressemblances entre ces bactéries et celles décrites par Theodor Escherich étaient très fortes, deux noms différents leurs avaient été attribués étant donné le caractère pathogène des premières et le caractère commensal des secondes. Cependant, l'avancée des études phénotypiques et génotypiques de ces deux bactéries, plus d'un siècle après leur découverte, montre désormais qu'elles pourraient être placées au sein d'une seule et même espèce. Par exemple, de nombreux antigènes O portés par des souches *Shigella* sont identiques à ceux portés par des souches *E. coli* (Wang *et al*., 2001). Or, au sein des entérobactéries, l'antigène O, partie externe glucidique des lipopolysaccharides, est souvent considéré comme spécifique à l'espèce, voire même aux souches et seule l'endotoxine, partie membranaire lipidique, semble conservée et identique chez toutes. Un autre fait appuyant la relation étroite entre les bactéries *E. coli* et *Shigella* est la connaissance désormais de souches *E. coli* pathogènes, dont certains pathotypes présentent des modes d'action similaires à ceux des bactéries *Shigella* (Johnson, 2000). Les données permettant d'affirmer de la manière la plus évidente et la plus sûre l'appartenance de ces deux bactéries à une seule et même espèce concernent les analyses de leurs génomes. Les séquences nucléotidiques des gènes codant les ARNr 16S de différentes souches *E. coli* et *Shigella* présentent en effet entre elles toujours plus de 97 % d'identité (Fukushima *et al*., 2002 ; Amano *et al*., 2005 ; Lee et Côté, 2006). Les taux d'hybridation ADN-ADN observés entre leurs génomes dépassent également 70 % : 95 %, 98 %, 95 % et 92,6 % d'hybridation sont ainsi obtenus entre les génomes de la souche *E. coli* GTC 503 et la souche *Sh. dysenteriae* GTC 786, la souche *Sh. flexneri* GTC 780, la souche *Sh. boydii* GTC 779 et la souche *Sh. sonnei* GTC 781, respectivement (Amano *et al*., 2005). Ces taux d'hybridation supérieurs à 70 % sont en relation avec les pourcentages moyens d'identité nucléotidique des séquences génomiques partagées des souches *E. coli* et *Shigella*, qui atteignent des valeurs supérieures à 95 % (Konstantinidis *et al.*, 2006).

3.1.2. *Escherichia coli* au sein du genre *Escherichia*

Bien que les études du genre bactérien *Escherichia* restent majoritairement concentrées sur l'espèce *E. coli*, cinq autres espèces ont pu être décrites : *E. albertii*, *E. blattae*, *E. fergusonii*, *E. hermanii* et *E. vulneris* (tableau 8) (Bettelheim, 2002 ; Abbott *et al*., 2003). Ces six espèces se distinguent par différents caractères métaboliques et surtout par leurs pourcentages d'hybridation ADN-ADN comme le montre le tableau 9.

Tableau 8 : Présentation des cinq espèces, autres qu'*E. coli*, du genre *Escherichia.*

Espèce	Souche type	Isolement	Référence
E. albertii	*E. albertii* Albert 19982	Enfant de neuf mois atteint de diarrhée	Huys *et al*., 2003
E. blattae	*E. blattae* ATCC 29907	Intestin de blattes en bonne santé	Burgess *et al*., 1973
E. fergusonii	*E. fergusonii* ATCC 35469	Sang, urines, plaies abdominales, fèces, vésicule biliaire de l'homme et d'animaux	Farmer *et al*., 1985
E. hermannii	*E. hermannii* ATCC 35650	Sang, selles, plaies, appareil respiratoire, liquide péritonéal de l'homme ; environnement ; aliments (viande de porc et coquilles d'œufs)	Brenner *et al*., 1982a
E. vulneris	*E. vulneris* ATCC 33821	Homme et animaux	Brenner *et al*., 1982b

Tableau 9 : Pourcentages d'hybridation ADN-ADN entre souches de différentes espèces du genre *Escherichia*, d'après Huys *et al*., 2003. - : non disponible.

	E. coli	***E. albertii***	***E. blattae***	***E. fergusonii***	***E. hermanii***	***E. vulneris***
E. coli	100	61-64	18	63	-	-
E. albertii	55-59	100	14-15	35-42	20	22-23
E. blattae	20	17-18	100	18	-	-
E. fergusonii	59	44-45	17	100	-	-
E. hermanii	40-46	18-20	-	-	100	28
E. vulneris	39	16-17	-	-	27	100

3.2. Etude comparative des génomes de différentes souches de l'espèce *E. coli*

Le premier génome d'une souche *E. coli* entièrement séquencé fût celui de la souche commensale non pathogène très utilisée en laboratoire K12 MG1655 (Blattner *et al*., 1997). Actuellement, dix génomes de souches *E. coli* sont entièrement séquencés et disponibles (tableau 10) et vingt et un sont en cours de séquençage (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html). Bien que ces trente et un génomes soient très probablement peu représentatifs du nombre total de génomes différents d'*E. coli* sur la Terre, une grande variabilité les caractérise d'ors et déjà. Des tailles de 4,6 Mb à 5,6 Mb, par exemple, sont enregistrées, signifiant un gain ou une perte nucléotidique possible de plus de 20 % d'un génome à l'autre.

Tableau 10 : Caractéristiques des dix génomes entièrement séquencés de l'espèce *E. coli*. [a] : sur le chromosome.

Souche	Taille du génome (pb)	GC% du génome	Nombre de chromosome	Nombre de plasmides	Nombre de gènes [a]	Nombre d'ORFs [a]	Nombre d'ARNt [a]	Nombre d'opérons ribosomiaux [a]	Pourcentage de séquences codantes [a]	Référence
K12 MG1655	4639675	50,8	1	0	4466	4289	86	7	85	Blattner *et al*., 1997
W3110	4646332	50,8	1	0	4444	4227	86	7	86	Hayashi *et al*., 2006
O157:H7 Sakaï	5498450	50,5	1	2	5395	5253	102	7	85	Hayashi *et al*., 2001
EDL933	5528445	50,3	1	1	5453	5324	98	7	87	Perna *et al*., 2001
CFT073	5231428	50,5	1	0	5589	5379	89	7	87	Welch *et al*., 2002
UTI89	5065741	50,6	1	1	5154	5044	88	7	88	-
536	4938920	50,5	1	0	4789	4629	81	7	87	-
APEC01	5082025	50,5	1	4	4573	4457	94	7	86	Johnson *et al*., 2007
HS	4643538	50,8	1	0	4636	4384	88	7	86	-
E24377	4979619	50,6	1	6	4987	4755	91	7	85	-

En 2002, la comparaison du quatrième génome entièrement séquencé d'*E. coli* avec deux autres génomes d'*E. coli* précédemment séquencés a montré que sur les 7461 protéines différentes recensées, seules 2996 étaient communes aux trois génomes (soit 40 % d'entre elles), 911 étaient communes à deux des trois génomes (soit 12 % d'entre elles) et 3554 étaient exclusivement présentes dans l'un des trois génomes (soit 48 % d'entre elles) (Welch *et al*., 2002). En 2005, une étude concernant neuf génomes entièrement séquencés à l'époque dont six de souches *E. coli* et trois de souches *Shigella*, a révélé quant à elle que sur environ 8050 gènes différents annotés au total, seuls 3000 environ étaient communs aux neuf génomes (soit 37 % d'entre eux) et 5050 environ étaient exclusivement présents dans l'un des neuf génomes (soit 63 % d'entre eux) (Konstantinidis et Tiedje, 2005). Il semble donc que plus le nombre de génomes *E. coli* séquencés augmente, plus le nombre de gènes souche spécifique augmente, tandis que le nombre de gènes constants partagés par toutes les souches tend à stagner (autour de 3000 environ). Ainsi l'espèce *E. coli* renferme des souches dont les génomes présentent des propriétés très variables. Ces différences de propriétés, en relation avec les différentes niches écologiques des souches, s'expliquent par des contenus en éléments génétiques mobiles ou mobilisables très variables et très différents dans leurs génomes. Une catégorie d'éléments génétiques mobiles ou mobilisables particulièrement impliqués dans la diversité de l'espèce *E. coli* est celle des îlots génomiques et particulièrement celle des îlots de pathogénicicté, dont un grand nombre a été présenté précédemment dans le tableau 6.

3.2.1. Génome de la souche *E. coli* K12 MG1655

Le génome de la souche commensale non pathogène K12 MG1655 a été publié en 1997 (Blattner *et al*., 1997). Il correspond à un unique chromosome circulaire de 4639221 pb, sur lequel sont identifiés 4289 gènes codant des protéines, 86 gènes codant des ARNt et 22 gènes codant des ARNr rassemblés dans sept opérons ribosomaux (tableau 10). Son pourcentage global en paires de bases guanine et cytosine est de 50,8 %. Les séquences nucléotidiques des 4289 gènes codant des protéines totalisent 87,8 % de la séquence complète du chromosome et les 108 gènes codant des ARNt et des ARNr 0,8 %. Le reste des séquences est attribué à des régions de régulation, des régions répétées et des régions de fonction inconnue. La distance moyenne séparant deux gènes est de 118 pb, toutefois 70 régions intergéniques de plus de 600 pb sont observées. La totalité des gènes annotés se répartissent sur 2584 opérons.

Plusieurs séquences de composition nucléotidique différente peuvent être retrouvées en de nombreux sites sur le chromosome. La séquence Chi nécessaire à la recombinaison homologue et faisant partie d'un grande famille d'octamères ayant conservé le trimère CTG, est ainsi comptabilisée 761 fois. L'élément *Rhs*, dont la taille varie de 5,7 à 9,6 kb, est quant à lui répété cinq fois, représentant 0,8 % de la taille du chromosome. Les séquences les plus fréquentes sont les séquences palindromiques de 40 pb appelées REP, totalisant 0,54 % de la taille du chromosome. Les séquences ERIC, BoxC, RSA, *Ter*, LDR et *iap* sont quant à elles répétées 19, 26, 6, 8, 2 et 2 fois, respectivement, sur le chromosome.

Les 4289 protéines codées sur le chromosome sont réparties dans 23 familles de différentes fonctions, sachant que presque 40 % d'entre-elles se situent dans une famille dite de fonction inconnue et 25 %, 12,5 % et 20 % d'entre elles sont impliquées dans le métabolisme de molécules de petite taille, dans le métabolisme de molécules de grande taille et dans différents processus cellulaires (réplication, transcription, traduction…), respectivement. La taille moyenne de ces protéines est de 317 acides aminés. Cependant, quatre protéines possèdent entre 1500 et 1700 acides aminés, 51 protéines entre 1000 et 1500 acides aminés et 381 protéines moins de 100 acides aminés. La plus grande présente une taille de 2383 acides aminés. En admettant que deux protéines sont paralogues si elles présentent au moins 30 % d'identité sur au moins 60 % de leur séquence, 1345 de ces 4289 protéines auraient au moins une séquence paralogue sur le chromosome.

Plusieurs éléments génétiques mobiles sont identifiés sur le chromosome d'*E. coli* K12 MG1655. L'identification de ces éléments, sur la base de leurs caractéristiques différentes de celles du reste du génome (notamment leur pourcentage en paires de bases guanine et cytosine), a permis

d'estimer que 755 des 4289 protéines codées recensées (17,6 %) seraient d'origine étrangère et acquises par transfert génétique horizontal (Lawrence et Ochman, 1997). Cependant, ce chiffre ne concerne que les gènes récemment acquis et est donc probablement sous-estimé. En effet, au fil du temps, les séquences transférées, soumises au même processus de mutations qui touche les gènes du génome hôte, tendent à modifier leur composition pour la rapprocher de celle du génome hôte et leur origine étrangère passe ainsi inaperçue. Chez *E.* coli K12 MG1655, 42 séquences d'insertion de 10 types différents et 11 prophages ou éléments similaires à des prophages (tableau 11) sont ainsi, entre autres, recensés.

Tableau 11 : Prophages ou éléments similaires recensés sur le chromosome de la souche *E. coli* K12 MG1655, d'après Blattner *et al*., 1997 et Hayashi *et al*., 2001. * : éléments intégrés en 3' d'un gène codant un ARNt.

Nom	Taille (kb)	Nature
Lambda		Prophage lambda
DLP12*	21,3	Prophage cryptique de type lambda
Rac	23	
Qin	20,4	
CP4-57*	22	Prophage cryptique de type P4
CP4-6*	39,9	
CP4-44	12,9	
Eut/CPZ-55	6,8	
E14	15,2	Élément similaire à un prophage
KpLE1*	12,6	
KpLE2*	40,1	

Aucun de ces éléments ne porte de fonction de virulence. Ils renferment majoritairement des gènes codant des protéines hypothétiques de fonction inconnue, des protéines de bactériophages, des protéines impliquées dans la réplication et la réparation de l'ADN, ou encore des régulateurs transcriptionnels. Des gènes codant des protéines de résistance aux antibiotiques, des protéines de choc froid et des protéines de transport du fer sont également identifiés dans les éléments DLP12, Qin, et KpLE2, respectivement.

3.2.2. Génome de la souche *E. coli* O157:H7 Sakaï

E. coli O157:H7 Sakaï est une souche pathogène intestinale. Elle a été isolée en 1996 au Japon dans la ville de Sakaï, où elle a été responsable d'une épidémie de colites hémorragiques chez plus de 6000 enfants, dont trois en sont morts. Son génome, publié en 2001, comprend un chromosome

circulaire de 5498450 pb, un plasmide pO157 de virulence de 92721 pb et un plasmide pOSAK1 cryptique de 3306 pb (tableau 10) (Hayashi *et al*., 2001). Le chromosome circulaire présente un pourcentage en paires de bases guanine et cytosine de 50,5 %, 5361 gènes codant des protéines (totalisant 88,1 % de sa séquence totale et d'une taille moyenne de 904 pb), 102 gènes codant des ARNt et 22 gènes codant des ARNr rassemblés dans sept opérons ribosomaux. Sur les 102 gènes codant des ARNt, 82 sont conservés sur le chromosome de la souche *E. coli* K12 MG1655 et 20 sont spécifiques à la souche *E. coli* O157:H7 Sakaï (deux suite à une duplication de deux gènes conservés et 18 présents sur des prophages de type lambda acquis par transfert génétique horizontal). De même, sur les 86 gènes codant des ARNt recensés sur le chromosome de la souche *E. coli* K12 MG1655, 4 sont absents sur le chromosome de la souche *E. coli* O157:H7 Sakaï, les gènes *leu*P, *leu*V, *lys*Q et *pro*M.

La comparaison des séquences nucléotidiques des chromosomes des souches *E. coli* K12 MG1655 et *E. coli* O157:H7 Sakaï permet d'identifier 4,1 Mb communes aux deux souches avec la conservation de plus de 98 % d'identité nucléotidique. Ces séquences partagées ne sont pas disposées de façon continue, mais, au contraire, interrompues plusieurs fois par l'insertion de séquences spécifiques aux deux souches. Ainsi, sont comptabilisées 296 et 325 insertions de plus de 19 pb spécifiques à la souche O157:H7 Sakaï et à la souche K12 MG1655, respectivement. Parmi ces insertions, 203 se situent à des sites analogues sur les deux chromosomes. La somme des tailles des séquences spécifiques insérées chez la souche O157:H7 Sakaï est de 1,4 Mpb, soit 25 % de la taille totale du chromosome (contre 0,5 Mpb chez la souche K12 MG1655). Ces séquences spécifiques à la souche O157:H7 Sakaï présentent un pourcentage en paires de bases guanine et cytosine de 48,5 % et sont de deux types différents : 48 % d'entre elles correspondent à des prophages ou à des éléments similaires à des prophages et 52 % sont d'autres natures (IS, îlots).

Le chromosome d'*E. coli* O157:H7 Sakaï possède effectivement 18 prophages ou vestiges de prophages appelés Sp1 à Sp18 (pour Sakaï prophage) et 6 régions appelées SpLE1 à SpLE6 (pour Sakaï prophage-like elements) ne contenant aucun gène présentant une homologie significative avec des gènes de bactériophages, excepté ceux codant des protéines de type intégrases. Quelques unes des propriétés de ces prophages, notamment les propriétés de virulence portées par certains, sont présentées dans le tableau 12. Sur les 18 éléments Sp, treize sont de type lambda et très similaires, partageant pour certains d'entre eux plus de 20 kb identiques ou presque identiques. Cette forte ressemblance n'est pas encore expliquée, mais des phénomènes de duplication et de recombinaison entre prophages ancestraux sont avancés comme hypothèses. Sur les 6 éléments SpLE, deux (SpLE1 et SpLE4) ressemblent aux prophages cryptiques de type P4 identifiés chez la souche *E. coli* K12 MG1655. Ces 24 régions au total, dont onze se situent en aval de gènes codant des ARNt, témoignent du rôle prédominant des bactériophages dans l'évolution de la souche *E. coli* O157:H7 Sakaï.

Tableau 12 : Prophages ou éléments similaires recensés dans le génome de la souche *E. coli* O157:H7 Sakaï, d'après Ohnishi *et al*., 2001. * : éléments intégrés en 3' d'un gène codant un ARNt. Lom : facteur d'adhésion cellulaire ; SOD : superoxide dismutase ; EspF : protéase ; Stx2 : shiga toxine de type 2 ; Bor : facteur de résistance au sérum ; SopA : protéase ; YopM : protéine effectrice ; EmrE : système d'efflux d'antibiotiques ; Anm : adhésine ; TrcA : protéine de régulation ; Stx1 : shiga toxine de type 1 ; TraT : facteur de résistance au sérum ; TetR : facteur de résistance au tellure ; Iha : facteur d'adhésion ; AidA-I : adhésine ; PagC : invasine ; ShET2 : shigella entérotoxine de type 2 ; Efa1 : adhésine ; LEE : locus of enterocyte effacement.

Nom	Taille (kb)	Nature	Facteurs de virulence
Sp1*	10,6	Prophage de type lambda	-
Sp2*	12,9	Prophage de type P4	-
Sp3	38,6	Prophages de type lambda	Lom
Sp4*	49,6		SOD, Lom, EspF
Sp5	62,7		Stx2, Lom, Bor
Sp6	48,4		SopA, YopM
Sp7	15,4	-	-
Sp8	46,9	Prophages de type lambda	EmrE, Bor, Lom, Catalase
Sp9	58,2		Anm, Lom, TrcA
Sp10	51,1		SOD, Lom
Sp11	45,8		Lom
Sp12	46,1		Lom
Sp13*	21,1	Prophage de type P2	-
Sp14*	44	Prophages de type lambda	EspF, Lom
Sp15	47,9		Stx1, Lom
Sp16*	8,5	Prophage de type P22	-
Sp17*	24,2	Prophage de type lambda	TrcA
Sp18	38,7	Prophage de type Mu	-
SpLE1*	86,2	Prophage de type P4	TraT, Ter^R, Iha, AidA-I, Uréase
SpLE2	13,4	-	-
SpLE3*	23,4	-	PagC, ShET2, Efa1
SpLE4*	43,4	Prophage de type P4	LEE
SpLE5*	45	-	-
SpLE6*	34,1	-	-

Les 296 insertions spécifiques à la souche *E. coli* O157:H7 Sakaï renferment 1632 des 5361 gènes codant des protéines recensés sur son chromosome. Un grand nombre de ces protéines spécifiques sont des protéines de bactériophages et des protéines de séquences d'insertion. Plusieurs d'entre elles portent également des fonctions de virulence, tels que des fimbriae, des adhésines, des pili, des systèmes de sécrétion de type III, des toxines, des facteurs de résistance au sérum et des systèmes de capture du fer. La plupart de ces facteurs de virulence sont codés sur les éléments Sp et SpLE. Au

niveau des systèmes de transport et des métabolismes, les protéines spécifiques à la souche *E. coli* O157:H7 Sakaï lui permettent, contrairement à la souche *E. coli* K12 MG1655, d'exporter hors de sa cellule plusieurs antibiotiques, de résister au tellure, de dégrader des acides aminés aromatiques, de fermenter le glutamate et de synthétiser différents lipides ou lipopolysaccharides ou lipoprotéines spécifiques. Inversement, la souche *E. coli* O157:H7 Sakaï ne peut utiliser, contrairement à la souche *E. coli* K12 MG1655, la 2-phenylethylamine, la xanthosine, le D-galactonate, le L-idonate, le glycolate, le D-sorbitol et les acides gras à courtes chaînes, suite à l'absence ou à l'altération des gènes nécessaires. Enfin, les deux bactéries disposent de systèmes de modification / restriction d'ADN complètement différents.

Le génome d'une autre souche *E. coli* pathogène intestinale a été publié la même année que celui de la souche *E. coli* O157:H7 Sakaï. Il s'agit de celui de la souche *E. coli* EDL933, isolée de viande bovine hachée aux Etats-Unis en 1982 (Perna *et al*., 2001). Les chromosomes de ces deux souches pathogènes intestinales sont très similaires. Les quelques différences observées concernent chez la souche *E. coli* EDL933 l'inversion d'environ 420 kb à proximité du site de terminaison de la réplication, la duplication de l'élément SpLE1, l'absence de l'élément Sp18 et quelques différences au sein de seize régions chromosomiques, dont deux correspondent à des IS et dix à des séquences internes de prophages de type lambda.

3.2.3. Génome de la souche *E. coli* CFT073

E. coli CFT073 est une souche pathogène extra-intestinale. A l'origine d'infections du tractus urinaire, elle est plus précisément qualifiée de souche uropathogène. Elle a été isolée en 1997 au Centre Hospitalier Universitaire de Maryland aux Etats-Unis, à partir du sang d'une femme atteinte de pyélonéphrite. Son génome, publié en 2002, comprend un unique chromosome circulaire de 5231428 pb (taille similaire à celle des chromosomes des deux souches pathogènes intestinales citées précédemment) et présente un pourcentage en paires de bases guanine et cytosine de 50,47 % (tableau 10) (Welch *et al*., 2002). Sur ce chromosome, ont été identifiés 5533 gènes codant des protéines, 89 gènes codant des ARNt et 22 gènes codant des ARNr rassemblés dans sept opérons ribosomaux. Sur les 89 gènes codant des ARNt, 77 sont conservés sur le chromosome de la souche *E. coli* K12 MG1655 et 12 sont spécifiques à la souche *E. coli* CFT073. De même, sur les 86 gènes codant des ARNt recensés sur le chromosome de la souche *E. coli* K12 MG1655, 9 sont absents sur le chromosome de la souche *E. coli* CFT073, les gènes *ile*V, *ala*V, *arg*U, *leu*W, *met*T, *ile*Y, *thr*V, *ala*U, et *ile*U.

La comparaison des séquences nucléotidiques des chromosomes de la souche *E. coli* CFT073 et de la souche *E. coli* K12 MG1655 a permis d'identifier 3,92 Mpb communes, ayant conservé plus de 98 % d'identité nucléotidique. Comme dans le cas du chromosome de la souche *E. coli* O157:H7 Sakaï, ces séquences partagées ne sont pas disposées de façon continue, mais interrompues par 247 insertions spécifiques de plus de 50 pb sur le chromosome de la souche *E. coli* CFT073. Ces 247 insertions totalisent 1,303 Mb, soit 25 % de la taille totale du chromosome et renferment 2004 gènes. Parmi ces 2004 gènes, 204 sont présents sur le chromosome de la souche *E. coli* EDL933, dont deux tiers codent des protéines de fonction inconnue ou des protéines phagiques ou des protéines de séquences d'insertion et un tiers code des adhésines ou des protéines impliquées dans des systèmes d'incorporation du fer, des voies de biosynthèse d'acides gras, des systèmes de transfert de groupements phosphate et des systèmes de transport de type ABC.

Les insertions spécifiques à la souche *E. coli* CFT073 comptent cinq prophages cryptiques. Ces derniers présentent des tailles allant de 11 à 54 kb et des pourcentages en paires de bases guanine et cytosine de 49 à 51 %, très proches de celui du chromosome global. Quatre d'entre eux ne portent aucun gène connu de virulence et un, localisé à proximité du gène *icdA*, possède quatre gènes (*sitA*, *sitB*, *sitC* et *sitD*) impliqués dans le transport du fer et du manganèse et a ainsi été récemment renommé PAI-CFT073-*icdA* (Lloyd *et al*., 2007). Une recherche de ces prophages sur les chromosomes entièrement séquencés des autres souches *E. coli* montre que seulement des fragments de certains d'entre eux peuvent y être retrouvés. Parmi les autres insertions spécifiques à la souche *E. coli* CFT073, neuf îlots génomiques de plus de 30 kb ont pu être identifiés, dont huit se trouvent à proximité de gènes codant des ARNt (Lloyd *et al*., 2007). Sur ces neuf îlots, dont les caractéristiques sont présentées dans le tableau 13, six sont porteurs de fonctions de virulence en relation notamment avec la nature pathogène et le tropisme de la souche. Des îlots de pathogénicité ont pu être identifiés au sein du chromosome d'autres souches *E. coli* uropathogènes, mais ceux-ci ne montrent que très peu de similitudes avec ceux présents chez la souche *E. coli* CFT073. Certes quelques gènes codant certains facteurs de virulence sont partagés, mais présents au sein d'îlots aux contenus différents et aux localisation chromosomiques également différentes. Le seul îlot qui semble bien conservé au sein des chromosomes de toutes les souches uropathogènes jusqu'à présent étudiées est l'îlot PAI-CFT073-*asnT* de la souche *E. coli* CFT073, encore appelé HPI (pour High Pathogenicity Island) et codant un système d'incorporation du fer.

Tableau 13 : Caractéristiques des neuf îlots génomiques d'une taille de plus de 30 kb recensés dans le génome de la souche *E. coli* CFT073, d'après Lloyd *et al*., 2007. * : gène codant un ARNt.

Nom	Taille (kb)	GC%	Gène cible d'insertion	Contenu en facteurs de virulence
PAI-CFT073-*aspV*	100	47	*aspV**	Adhésine, protéase
PAI-CFT073-*serX*	113	49	*serX**	Fimbriae, système de capture du fer, Ag43
PAI-CFT073-*asnT*	32	57	*asnT**	Système de capture du fer
PAI-CFT073-*metV*	32	53	*metV**	Protéases
PAI-CFT073-*pheV*	123	47	*pheV**	Hémolysine, fimbriae, adhésine, toxine, systèmes de capture du fer, capsule, Ag43
PAI-CFT073-*pheU*	52	48	*pheU**	Fimbriae
GI-CFT073-*selC*	68	47	*selC**	Îlots non virulents
GI-CFT073-*asnW*	54	53	*asnW**	Îlots non virulents
GI-CFT073-*cobU*	44	50		Îlots non virulents

3.2.4. Génome de la souche *E. coli* APEC01

E. coli APEC01 est une souche pathogène également extra-intestinale, non plus de source humaine, mais animale. Elle a été isolée du poumon d'un poulet atteint de colisepticémie. Son génome, publié en 2007, comprend un chromosome circulaire et quatre plasmides (Johnson *et al*., 2007).

Le chromosome présente une taille de 5082025 pb (taille similaire à celle des chromosomes des souches pathogènes précédemment analysées) et un pourcentage en paires de bases guanine et cytosine de 50,55 % (tableau 10). Il renferme 4467 gènes codant des protéines, 93 gènes codant des ARNt et 22 gènes codant des ARNr rassemblés dans sept opérons ribosomaux. Sur les 4467 gènes codant des protéines, 3484 (78 %) et 4020 (90 %) sont conservés sur les chromosomes des souches *E. coli* K12 MG1655 et *E. coli* CFT073, respectivement. De plus amples analyses ont montré d'une manière générale que le chromosome de cette souche pathogène extra-intestinale animale ressemble plus à ceux des souches pathogènes extra-intestinales humaines qu'à ceux des souches pathogènes intestinales humaines ou encore qu'à ceux des souches commensales humaines non pathogènes. Les régions exclusivement partagées avec les souches pathogènes extra-intestinales humaines codent notamment des systèmes d'incorporation du fer, des fimbriae, des protéines capables d'auto-sécrétion et des systèmes de transport pour le manganèse. La proximité de la souche *E. coli* APEC01 des souches pathogènes extra-intestinales humaines est illustrée par le fait que sur les 43 insertions de plus de 4 kb identifiées sur son chromosome et absentes sur le chromosome de la souche *E. coli* K12 MG1655, seules quatre lui sont véritablement spécifiques, les 39 autres étant également présents sur le chromosome d'au moins une souche pathogène extra-intestinale humaine. L'analyse des quatre

régions spécifiques (totalisant 111046 pb, soit 2 % de la taille du chromosome) montre que trois d'entre elle correspondent à des prophages. Leur contenu reste encore de fonction inconnue.
Les quatre plasmides de la souche *E. coli* APEC01 diffèrent en taille et en fonction. Les plasmides pAPEC-01-Cryp1 et pAPEC-01-Cryp2, d'une taille de 101 kb et 49 kb et d'un pourcentage en paires de bases guanine et cytosine de 46,55 % et 44,62 %, respectivement, codent des protéines pour la plupart hypothétiques, ne conférant aucun phénotype apparent. Le plasmide pAPEC-01-ColBM, d'une taille de 174 kb et au pourcentage en paires de bases guanine et cytosine de 49,65 %, est un plasmide de virulence, portant un des cinq îlots de pathogénicité de la souche. Cet îlot code un facteur de résistance au sérum, un système de transport pour le manganèse et le fer, deux systèmes permettant l'incorporation de fer et une hémolysine. Nombre de ces facteurs de virulence sont retrouvés sur des îlots de pathogénicité à localisation chromosomique chez les souches pathogènes extra-intestinales humaines. Enfin, le plasmide pAPEC-01-R, d'une taille de 241 kb et au pourcentage en paires de bases guanine et cytosine de 46,43 %, est un plasmide apportant une résistance à huit antibiotiques différents.

3.2.5. Etude partielle des génomes des souches *E. coli* A0 34/86 et *E. coli* Nissle 1917

Les quatre génomes précédemment décrits semblent montrer que la présence d'éléments spécifiques, sur chacun, explique et est en relation avec la nature et le tropisme des souches qui les contiennent. Or l'existence de différents types de souches *E. coli* ne peut être expliquée si facilement comme le révèle les analyses partielles d'autres génomes. Une étude de la souche *E. coli* A0 34/86 a en effet révélé la présence dans son génome du HPI commun aux souches uropathogènes et de plusieurs opérons similaires à ceux trouvés dans un îlot de pathogénicité de la souche *E. coli* 536, opérons codant notamment des fimbriae, des adhésines et un système de transport pour le phosphoglycérate (Hejnova *et al*., 2006). Cette souche *E. coli* A0 34/86 est cependant commensale, utilisée depuis les années 1970 en Tchékoslovaquie pour la prévention des diarrhées nosocomiales chez les nouveaux nés (Lodinova *et al*., 1967, 1980). De même, plusieurs îlots génomiques présentant des propriétés communes avec quelques uns des îlots de pathogénicité de la souche *E. coli* CFT073 ont été mis en évidence dans le génome de la souche *E. coli* Nissle 1917 (Grozdanov *et al*., 2004). Cette dernière est également une souche commensale, utilisée dans la préparation d'aliments probiotiques pour le traitement de désordres intestinaux (Lodinova-Zadnikova *et al*., 1992 ; Lodinova-Zadnikova et Sonnenborn, 1997). Ainsi, il semble que certaines souches commensales non pathogènes puissent partager des éléments génétiques avec des souches pathogènes, sans que d'autres souches commensales ne possèdent ces éléments (il faut penser ici à la souche *E. coli* K12 MG1655).

3.2.6. Conclusion

Les éléments génétiques mobiles ou mobilisables semblent donc constituer une part non négligeable du pool génétique de l'espèce *E. coli*. Dans chacun des génomes analysés précédemment, un pourcentage élevé de ces éléments peut en effet être calculé. Selon les souches, certains types d'éléments semblent prédominants, comme, par exemple, les prophages chez la souche O157 :H7 Sakaï et les îlots génomiques chez la souche CFT073. L'analyse du contenu de ces éléments montre que ce sont eux qui apportent aux souches leurs propriétés spécifiques, donc que ce sont eux qui sont responsables de la diversité de l'espèce *E. coli*. Il apparaît également que les plus grandes tailles des génomes des souches pathogènes s'expliquent par un contenu plus important en éléments génétiques mobiles.

3.3. Description des différents pathotypes de l'espèce *E. coli*

3.3.1. Pathotypes intestinaux

De nombreuses souches *E. coli* pathogènes sont reconnues comme des agents responsables de syndromes diarrhéiques d'origine alimentaire ou hydrique. Six principaux pathotypes intestinaux sont ainsi recensés : les *E. coli* entérotoxinogènes ou ETEC, les *E. coli* entéroinvasives ou EIEC, les *E. coli* entéropathogènes ou EPEC, les *E. coli* entérohémorragiques ou EHEC, les *E. coli* entéroagrégatives ou EAEC et les *E. coli* à adhésion diffuse ou DAEC (Robins-Browne et Hartland, 2002). Les caractéristiques cliniques, pathologiques et épidémiologiques de ces six pathotypes ont été présentées dans le tableau 7. Comme l'illustrent quelques exemples qui vont être développés ci-dessous, leurs facteurs de virulence sont codés par des gènes portés par des éléments génétiques mobiles ou mobilisables. Ce fait montre donc de nouveau que ces éléments sont responsables de la spécificité de chaque souche de l'espèce *E. coli*.

La virulence des ETEC est liée à la production d'une ou de deux entérotoxines et de facteurs dits de colonisation (Qadri *et al*., 2005). La première entérotoxine, thermolabile, est appelée LT. D'une masse moléculaire de 84 kDa, elle possède une structure de type A_1B_5, où les cinq sous unités B ont pour rôle de se lier aux récepteurs de type ganglioside (GM1) des cellules épithéliales de la muqueuse de l'intestin grêle, afin que la sous unité A puisse être internalisée par ces dernières. Une fois dans le cytoplasme des cellules épithéliales, la sous unité A active de manière irréversible l'adénylate cyclase, conduisant à une accumulation d'AMPc dans le cytoplasme et d'électrolytes dans la lumière intestinale. La deuxième entérotoxine, thermostable, est appelée ST. Composée de 18 à 19

résidus d'acides aminés, elle se lie et active également de manière irréversible les guanylates cyclases membranaires des cellules épithéliales, conduisant à une accumulation de GMPc dans leur cytoplasme et d'électrolytes dans la lumière intestinale. Les facteurs dits de colonisation, appelés CFs, sont des structures de type pili ou fimbriae, permettant aux bactéries de se lier aux cellules épithéliales de la muqueuse de l'intestin grêle et de s'y multiplier. Ces trois types de facteurs de virulence des ETEC sont codés par des gènes à localisation plasmidique.

La virulence des EIEC est liée à la sécrétion de protéines via des systèmes de sécrétion de type III (Nataro et Kaper, 1998). Les différentes protéines sécrétées permettent aux bactéries de pénétrer par endocytose dans une cellule épithéliale de la muqueuse du côlon, de se libérer des vésicules d'endocytose, de se déplacer dans le cytoplasme et d'envahir d'autres cellules voisines. Elles sont capables également d'induire l'apoptose de macrophages. Comme précédemment, les systèmes de sécrétion de type III utilisés ainsi que les protéines effectrices sécrétées sont codées par des gènes à localisation plasmidique.

La virulence des EPEC est liée à la production de facteurs d'adhésion, de systèmes de sécrétion de type III et de protéines sécrétées effectrices (Mellies *et al*., 2007). Le premier contact avec les cellules épithéliales de la muqueuse de l'intestin grêle se fait grâce à une protéine codée par un gène plasmidique et appelée Bfp. Puis, les bactéries adhèrent très fortement aux cellules épithéliales grâce à une protéine nommée intimine, y produisent des lésions d'attachement et d'effacement caractérisées par la destruction localisée des micro villosités de la bordure en brosse et y induisent des altérations au niveau du cytosquelette grâce à des protéines nommées Esp sécrétées par un système de sécrétion de type III. Ces trois derniers facteurs, intimine, protéines Esp et systèmes de sécrétion de type III, sont codés par des gènes présents sur un îlot de pathogénicité appelé LEE (pour Locus of Enterocyte Effacement).

La virulence des EHEC ressemble à celle des EPEC, mais en étant toutefois beaucoup plus sévère de par la présence de facteurs de virulence supplémentaires (Mellies *et al*., 2007). Ainsi en plus des facteurs portés par l'îlot LEE, les EPEC sont capables, d'une part, d'exprimer des Shiga toxines appelées Stx1 et Stx2. Ces toxines, codées par des gènes localisés sur des prophages, présentent une structure de type A_1B_5, où les cinq sous unités B ont pour rôle de se lier aux récepteurs de type globotriaosyl ceramide (Gb3) des cellules épithéliales de la muqueuse du côlon, afin que la sous unité A puisse être internalisée par ces dernières. Une fois dans le cytoplasme des cellules épithéliales, la sous unité A gagne l'appareil de Golgi et le réticulum endoplasmique où elle y altère l'intégrité des ribosomes, ce qui conduit à l'arrêt du processus de traduction et à la mort cellulaire. Les EHEC possèdent, d'autre part, cinq facteurs de virulence supplémentaires codés par des gènes localisés sur

un plasmide appelé pO157. Ces facteurs sont une hémolysine (HlyA), une protéase à sérine (EspP) responsable du clivage du facteur V de coagulation humain, une protéine de surface (ToxB) permettant l'adhésion des bactéries aux cellules épithéliales, une catalase et une métalloprotéase à zinc (StcE) impliquée dans l'inactivation du système du complément et dans la colonisation et la destruction des tissus.

La virulence des EAEC repose sur plusieurs facteurs différents selon le stade de leur infection (Huang *et al*., 2006). Dans un premier temps, différents facteurs d'adhésion de type fimbriae (AAFs) sont exprimés, afin de permettre l'adhésion des bactéries aux cellules épithéliales intestinales. Ces facteurs AAFs sont codés par des gènes à localisation plasmidique. Ensuite la production de protéines codées par des gènes à localisation chromosomique entraîne la formation d'un biofilm bactérien à la surface de la muqueuse intestinale. Enfin, grâce à la sécrétion notamment de différentes toxines codées par des gènes plasmidiques, les bactéries peuvent être à l'origine de nécroses au pôle apical des villosités et éventuellement d'oedèmes inflammatoires et hémorragiques. Les trois toxines les plus étudiées sont la toxine Pet qui est une protéase à sérine, l'entérotoxine thermostable EAST1 similaire à la toxine ST des ETEC et l'entérotoxine ShET1 retrouvée chez les souches *Shigella*.

La virulence des DAEC est encore mal comprise et sujette à de nombreuses discussions et controverses. Ces souches *E. coli* pathogènes ont été identifiées grâce à leur capacité à adhérer de manière diffuse sur des cellules épithéliales cultivées *in vitro* (Scaletsky *et al*., 1984 ; Nataro *et al*., 1985 ; Nataro *et al*., 1987) et constituent un groupe aux propriétés très hétérogènes (Nataro *et al*., 1985 ; Czeczulin *et al*., 1999). Leurs facteurs d'adhésion sont des adhésines de la famille Afa/Dr (Nowicki *et al*., 2001), dont certaines d'entre elles permettent l'internalisation des bactéries par les cellules épithéliales, ou des protéines de la membrane externe appelées AIDA (pour adhesine involved in diffuse adherence) (Benz et Schmidt, 1993).

3.3.2. Pathotypes extra-intestinaux

Des souches *E. coli* pathogènes peuvent également être à l'origine d'infections extra-intestinales. Deux pathotypes extra-intestinaux (ExPEC) différents sont ainsi recensés : les *E. coli* uropathogènes ou UPEC à l'origine d'infection du tractus urinaire et les *E. coli* associées à des méningites chez les nouveaux nés ou MNEC (tableau 7). Comme dans le cas des pathotypes intestinaux, les facteurs de virulence de ces deux pathotypes extra-intestinaux sont codés par des gènes portés par des éléments génétiques mobiles ou mobilisables.

Les UPEC sont à l'origine de 90 % des cas d'infections survenant sur un arbre urinaire normal, telles que les cystites et les pyélonéphrites. Pour cela elles expriment différents facteurs de virulence, dont la plupart sont codés par des gènes localisés sur des îlots de pathogénicité (Yamamoto, 2007). Les premiers facteurs impliqués dans l'infection permettent l'adhésion des bactéries aux cellules uroépithéliales, grâce à la reconnaissance de récepteurs spécifiques à la surface de ces dernières. Ces facteurs d'adhésion sont des fimbriae de type 1, P, S et F1C et certains d'entre eux sont même capables d'induire l'invasion des cellules uroépithéliales par les bactéries. Les facteurs exprimés ensuite sont des toxines, tels que des hémolysines de type alpha, des facteurs cytotoxiques nécrotiques de type 1 (CNF1) et d'autres toxines capables de s'auto sécréter. Les hémolysines de type alpha permettent la lyse des cellules hôtes et des lymphocytes, mais également l'inhibition des processus de phagocytose et de chimiotactisme. La lyse des cellules hôtes a pour but notamment de fournir des nutriments et du fer aux bactéries, fer qu'elles captent grâce à la production de sidérophores. Les CNF1 constituent un moyen de défense des bactéries face au système immunitaire hôte, en inhibant les processus de phagocytose et l'action des granulocytes et sont également capables de tuer les cellules uroépithéliales humaines par un mécanisme apoptotique. Le dernier facteur de virulence important des UPEC est la production d'un antigène capsulaire diminuant la reconnaissance des bactéries par les anticorps, les macrophages ou les molécules du complément de l'organisme hôte.

Les MNEC sont à l'origine de 4 % des cas de méningites bactériennes touchant principalement les nouveaux nés et les patients de neurochirurgie. Un taux de 80 % d'entre elles exprime l'antigène capsulaire K1. Ce dernier est considéré comme le facteur de pathogénicité le plus important pour l'infection, présentant une activité antiphagocytaire élevée. Sa synthèse est dirigée par des protéines codées par des gènes localisés sur des îlots de pathogénicité.

3.3.3. Relation entre pathotypes et groupes phylogénétiques

Les souches de l'espèce *E. coli* peuvent être réparties dans quatre principaux groupes phylogénétiques : A, B1, B2 et D. La première distinction entre ces quatre groupes était basée sur des analyses de variants alléliques par migration électrophorétique d'enzymes cellulaires telles que l'adenylate kinase, l'aconitase, la phosphoglucose isomérase, l'alcaline phosphatase, l'isocitrate déhydrogénase, ou encore l'indophénol oxydase (Selander *et al*., 1986). Puis, une distinction basée sur l'analyse, par RFLP, du polymorphisme des opérons ribosomiques avait été introduite (Desjardins *et al*., 1995). Désormais, l'appartenance d'une souche à l'un de ces groupes peut être définie par la présence ou l'absence dans son génome de trois marqueurs moléculaires : le gène *chuA*, requis pour le transport de l'hème chez les EHEC, le gène *yjaA*, de fonction inconnue chez la souche *E. coli* K12 MG1655 et le fragment d'ADN TspE4C2, présent chez les MNEC (Clermont *et al*., 2000) (figure 19).

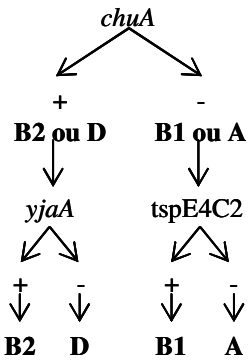


Figure 19 : Arbre dichotomique déterminant le groupe phylogénétique d'une souche *E. coli* selon sa possession ou non des gènes *chuA* et *yjaA* et du fragment d'ADN tspE4C2, d'après Clermont *et al*., 2000.

Des études ont montré qu'il existe un lien entre la nature, le tropisme et le groupe phylogénétique d'une souche *E. coli* pathogène (Picard *et al*., 1999 ; Escobar-Paramo *et al*., 2004). Ainsi les souches pathogènes extra-intestinales semblent exclusivement être associées aux groupes B2 et D, les souches ETEC, EHEC et EIEC exclusivement aux groupes A et B1 et les souches commensales non pathogènes, EPEC, EAEC et DAEC semblent quant à elles pouvoir appartenir aussi bien aux quatre groupes phylogénétiques, avec une prédominance, cependant, des groupes phylogénétiques A et B1 pour les souches commensales non pathogènes (tableau 14).

Tableau 14 : Distribution des différents pathotypes *E. coli* au sein des quatre principaux groupes phylogénétiques déterminés chez cette espèce (1ère colonne) et distribution des gènes codant les facteurs de virulence spécifiques de chaque pathotype au sein de ces mêmes quatre groupes phylogénétiques (2ème colonne), d'après Escobar-Paramo *et al*., 2004.

Pathotype	Groupe phylogénétique	Facteurs de virulence
ExPEC	B2 – D	A – B1 – B2 – D
EAEC DAEC EPEC	A – B1 – B2 – D	A – B1 – B2 – D
ETEC EHEC EIEC	A – B1	A – B1
Commensal	A – B1 – B2 – D	

L'analyse de la distribution des gènes codant les facteurs de virulence spécifiques à chacun des pathotypes de l'espèce *E. coli* a révélé quant à elle que : les gènes spécifiques de la virulence des souches extra-intestinales d'une part et des souches EPEC, EAEC et DAEC d'autre part, peuvent être retrouvés dans le génome de souches appartenant aussi bien aux quatre groupes phylogénétiques et que ceux spécifiques des souches ETEC, EHEC et EIEC étaient exclusivement retrouvés dans le génome de souches appartenant aux groupes phylogénétiques A et B1 (tableau 14). Ces différentes observations ont permis d'aboutir à la conclusion qu'il existe un lien entre la virulence et la

phylogénie d'une souche *E. coli*, c'est-à-dire un lien entre l'acquisition de facteurs de virulence et la nature du génome core d'une souche *E. coli*, mais également un lien entre l'expression des facteurs de virulence acquis et la nature de ce génome. Ainsi, l'acquisition des facteurs de virulence spécifiques des souches extra-intestinales ne semble pas dépendre du génome core, mais leur expression oui ; l'acquisition des facteurs de virulence spécifiques des souches intestinales EAEC, DAEC et EPEC ne semble également pas dépendre du génome core, ni leur expression ; enfin, l'acquisition des facteurs de virulence spécifiques des souches intestinales EHEC, ETEC et EIEC semble quant à elle dépendre du génome core.

4. Objectif du travail : caractérisation d'une souche *Escherichia coli* inductrice de mammites bovines

Il existe au sein de l'espèce *E. coli* un pathotype non cité dans le chapitre précédent qui renferme les souches *E.* coli inductrices de mammites bovines, encore appelées *E. coli* mammopathogènes ou MPEC. Contrairement à leurs consoeurs intestinales et extra intestinales, ces souches MPEC restent extrêmement peu caractérisées et même citées dans la littérature. Or, la pathologie dont elles sont responsables chez le bovin est celle la plus fréquemment rencontrée chez les vaches laitières et est à l'origine d'importantes pertes économiques pour les producteurs laitiers du monde entier. Ces pertes économiques sont expliquées par les coûts des soins vétérinaires et parfois de l'abattage des animaux infectés, par le fait que les animaux infectés (même qu'une seule fois dans leur vie) puis guéris ne regagnent jamais leur production de lait initiale et par le fait également que pendant l'infection, le lait produit est de moins bonne qualité et non autorisé à la consommation (Bradley, 2002).

4.1. Présentation des mammites bovines

4.1.1. Définition

Une mammite bovine se définit comme l'inflammation d'un ou de plusieurs quartiers de la mamelle, suite à une infection par un microorganisme. Cette pathologie se caractérise par des signes cliniques et une sévérité qui varient grandement selon le microorganisme inducteur. Cependant, elle entraîne systématiquement une augmentation du nombre de cellules somatiques présentes dans le lait (Harmon, 1994 ; Schepers *et al*., 1997). Ces cellules somatiques comprennent des cellules épithéliales et des globules blancs. Pour un lait normal, la concentration en cellules somatiques est généralement

inférieure à 10^5 cellules par millilitre (Burvenich *et al*., 2004) et contient majoritairement (80 %) des macrophages (Boutet *et al*., 2006). Pour un lait issu de mammite au contraire, cette concentration peut aller de 5.10^5 à plusieurs millions de cellules par millilitre (Paape *et al*., 1981) et renferme principalement (90 à 95 %) des neutrophiles (Boutet *et al*., 2006). Le décompte des cellules somatiques présentes dans son lait peut donc indiquer si une vache lutte contre une infection par un microorganisme.

4.1.2. Agents inducteurs

Un grand nombre de microorganismes peuvent être retrouvés sur et dans la mamelle d'une vache. En 1988, Watts a ainsi identifié 137 espèces et sous-espèces de microorganismes pouvant être associés à la glande mammaire bovine (Watts, 1988). Ces derniers sont très divers et peuvent être des bactéries, des levures, des champignons ou des algues. La majorité d'entre eux font partie de la flore microbienne normale du pis, n'y causant aucune maladie et pouvant même le protéger d'infections d'organismes pathogènes. Seul un petit nombre peut, par contre, provoquer l'infection de la glande mammaire. Les microbes les plus fréquemment inducteurs de mammites (causant environ 90 % des cas) sont ainsi d'origine bactérienne et appartiennent plus particulièrement aux cinq espèces bactériennes suivantes : *Staphylococcus aureus*, *Streptococcus agalactiae*, *Streptococcus dysgalactiae*, *Streptococcus uberis* et *Escherichia coli*.

En 1995, ces cinq principaux agents bactériens inducteurs de mammites bovines ont été répartis entre deux classes de pathogènes, l'une dite à "réservoir mammaire" et l'autre dite à "réservoir environnemental" (Blowey et Edmondson, 1995). Les microorganismes à "réservoir mammaire", qui prolifèrent sur la peau et à l'intérieur des mamelles, ont comme source principale des animaux infectés et sont transmis d'une vache à une autre lors de la traite. Leurs représentants principaux sont *St. aureus*, *S. agalactiae* et *S. dysgalactiae*. Les microorganismes à "réservoir environnemental" sont retrouvés quant à eux sur le sol, la litière, le fumier ou dans l'eau et atteignent donc le trayon uniquement lors de traites réalisées dans de mauvaises conditions d'hygiène. Leurs représentants principaux sont *S. uberis* et *E. coli*. La fréquence relative de ces cinq différents germes responsables de mammites présente de grandes variations géographiques. Ainsi vers la fin des années 1990, par exemple, les germes *St. aureus* et *E. coli* étaient responsables de mammites pour 5 % et 35 %, respectivement, des cas en Grande-Bretagne (Bradley et Green, 2001a), 46,5 % et 1,5 %, respectivement, des cas au Québec et 27 % et 31 %, respectivement, des cas en France (Seegers *et al*., 1997). A l'heure actuelle toutefois, l'ensemble des pays producteurs de lait s'accorde à dire que les mammites touchant les vaches laitières sont majoritairement causées par des germes de l'espèce *E. coli* (Wenz *et al*., 2001).

4.1.3. Conséquences

Les mammites peuvent ou non s'accompagner de signes physiologiques visibles chez le bovin et sont alors qualifiées de mammites cliniques ou de mammites sub-cliniques, respectivement. Autrefois, les mammites cliniques étaient souvent imputées aux germes à "réservoir environnemental" et celles sub-cliniques aux germes à "réservoir mammaire" (Blowey and Edmondson, 1995). Cependant, cette attribution n'est désormais plus valide car des cas de mammites sub-cliniques dues à des germes à "réservoir environnemental" sont de plus en plus rapportés (Bradley, 2002). Les signes physiologiques visibles des mammites cliniques sont une mamelle souvent enflée, rouge, chaude, douloureuse et un lait de constitution anormale. En fonction de la sévérité de la mammite, le lait peut être en partie coagulé, décoloré et contenir du sang, tandis que la vache peut présenter une déshydratation, une perte d'appétit, un rythme cardiaque élevé et être victime de fièvre et de diarrhées.

Qu'elles soient cliniques ou sub-cliniques, les mammites entraînent à la fois une baisse de production laitière et une modification importante de la composition du lait. La baisse de production laitière est liée à la diminution de l'activité sécrétrice des lactocytes, lieux de synthèse des principaux constituants du lait (Hamann and Krömker, 1997). Elle peut également être expliquée par l'altération voire la destruction des cellules sécrétrices par les globules blancs (Sordillo *et al.*, 1997) et par l'invasion de ces dernières par les agents pathogènes (Bayles *et al.*, 1998). La modification de la composition du lait est en grande partie due à l'altération physique de la barrière épithéliale mammaire séparant les compartiments "lait" et "sang" (Nickerson and Pankey, 1984), mais aussi, vraisemblablement, à la présence de facteurs issus des germes contaminants. Divers flux de molécules, ions, cellules, protéines et enzymes peuvent ainsi s'établir d'un compartiment à l'autre. De par la diminution de l'activité sécrétrice des lactocytes, le lait devient ainsi plus pauvre en lactose (Shuster *et al.*, 1991). De par l'altération physique de la barrière épithéliale mammaire, il s'appauvrit également en calcium, phosphore, potassium (Munro *et al.*, 1984) et s'enrichit en sodium et chlore (Munro *et al.*, 1984 ; Le Roux *et al.*, 1995). Une augmentation du passage de sérum albumine bovine, d'immunoglobulines et des composés du système plasminogène / plasmine est observée du sang vers le lait (Munro *et al.*, 1984 ; Kaartinen *et al.*, 1988). Enfin, une protéolyse importante est détectée : les teneurs en caséines α_s, β et κ diminuent fortement au profit des teneurs en caséines γ et en protéoses peptones, tous deux étant en partie des composés issus de l'hydrolyse des caséines (Michelutti *et al.*, 1999). Cette protéolyse est due à des dégradations enzymatiques endogènes (plasmine et protéases des cellules somatiques) (Michelutti *et al.*, 1999) et probablement exogènes (protéases bactériennes).

4.2. Connaissances actuelles des souches *E. coli* inductrices de mammites bovines

4.2.1. Variabilités génotypique et pathologique

A l'heure actuelle, aucun profil type de génotype et de facteurs de virulence ne semble se dégager des différentes études génotypiques et phénotypiques portées sur les souches *E. coli* inductrices de mammites bovines. Un grand polymorphisme génétique est en effet observé entre les souches isolées de cas indépendants de mammites. Une analyse menée en 1995 au Pays-Bas, utilisant la ERIC-PCR comme outil de typage moléculaire, a ainsi mis en évidence 13 génotypes différents pour seulement 30 isolats cliniques étudiés (Lipman *et al*., 1995). Plus récemment, aux Etats-Unis, la technique d'électrophorèse en champs pulsés, ou PFGE, a révélé 85 profils électrophorétiques différents pour 135 souches analysées (Srinivasan *et al*., 2007). Une aussi grande diversité génotypique a pu être observée chez des souches *E. coli* isolées de fourrage bovin grâce à la BOX-PCR, identifiant parmi 280 isolats 89 groupes taxonomiques différents (Yang *et al*., 2004). Ce dernier point confirme notamment le fait que les souches *E. coli* inductrices de mammites bovines proviennent de l'environnement.

Par ailleurs, la possession de gènes codant des facteurs de virulence semble extrêmement variée d'une souche à l'autre et est considérée même comme très rare. Le tableau 15 illustre bien ce fait en présentant les résultats de quelques recherches de gènes codant des facteurs de virulence chez des souches isolées de mammites bovines à travers le monde. Les fréquences des gènes recherchés, codant différentes toxines, adhésines, et intimines, restent très faibles. Cependant, un point très important à considérer ici, est que la recherche de ces gènes s'est faite par PCR grâce à l'utilisation d'amorces dessinées d'après les séquences nucléotidiques des gènes codant les facteurs de virulence identifiés et connus chez les souches *E. coli* pathogènes intestinales et extra-intestinales décrites dans le paragraphe 3.3. Ainsi les résultats négatifs obtenus signifient, certes, que les MPEC ne possèdent pas dans leur génome les gènes codant les facteurs de virulence connus chez les autres pathotypes d'*E. coli*, mais ne peuvent permettre d'affirmer, de manière définitive, qu'elles ne possèdent aucun ou presque aucun facteur de virulence. Des facteurs de virulence typiques des MPEC existent peut-être, mais restent pour l'instant impossibles à cribler étant donné qu'ils ne sont pas encore identifiés.

Tableau 15 : Prévalence de certains facteurs de virulence chez des souches *E. coli* inductrices de mammites bovines. Stx : shiga toxine ; CNF : cytotoxic necrotising factor ; eae : intimine ; cs31a : adhésine ; Afa : afimbrial adhésine ; F17 : fimbriae.

Facteurs de virulence		Pays / Année	Fréquence	Référence
Toxines	Stx1/2	USA / 2004	0/105 (0%)	Murinda *et al*., 2004
		Brésil / 2004	22/182 (12%)	Lira *et al*., 2004
		Nouvelle Zélande / 2004	25/80 (31%)	Bean *et al*., 2004
		France / 2005	0/123 (0%)	Raynaud *et al*., 2005
	CNF1/2	Irlande / 1996	1/62 (1,6%)	Burns *et al*., 1996
		Israël – Finlande / 2002	40/273 (14,65%)	Kaipainen *et al*., 2002
		Nouvelle Zélande / 2004	6/80 (7,5%)	Bean *et al*., 2004
		USA / 2006	14/123 (11,4%)	Wenz *et al*., 2006
Intimine	eae	Nouvelle Zélande / 2004	3/80 (3,75%)	Bean *et al*., 2004
		Brésil / 2004	11/182 (6%)	Lira *et al*., 2004
		USA / 2006	0/123 (0%)	Wenz *et al*., 2006
Adhésines	cs31a	Israël – Finlande / 2002	0/273 (0%)	Kaipainen *et al*., 2002
		USA / 2006	1/123 (0,8%)	Wenz *et al*., 2006
	Afa7D, 7E, F17a	Israël – Finlande / 2002	0/273 (0%)	Kaipainen *et al*., 2002
	F17b	Israël – Finlande / 2002	2/273 (0,7%)	Kaipainen *et al*., 2002
	F17c	Israël – Finlande / 2002	4/273 (1,4%)	Kaipainen *et al*., 2002
	F17d	Israël – Finlande / 2002	5/273 (1,8%)	Kaipainen *et al*., 2002
	F17II	Israël – Finlande / 2002	5/273 (1,8%)	Kaipainen *et al*., 2002
	S fimbriae	Israël – Finlande / 2002	13/273 (4,8%)	Kaipainen *et al*., 2002
	P fimbriae	Israël – Finlande / 2002	11/273 (4,1%)	Kaipainen *et al*., 2002
	Afa8D	Israël – Finlande / 2002	2/273 (0,7%)	Kaipainen *et al*., 2002
	Afa8E	Israël – Finlande / 2002	2/273 (0,7%)	Kaipainen *et al*., 2002

Ce fait n'est certainement pas improbable si l'on regarde plus attentivement les résultats concernant l'analyse de la résistance au sérum et aux antibiotiques des souches MPEC. Sur la base d'un test phénotypique consistant à observer ou non une croissance bactérienne dans un sérum bovin, la résistance au sérum a été identifiée plusieurs fois comme la seule caractéristique prédominante des souches MPEC, présente dans 64 à 100 % des cas (Caroll et Jasper, 1977 ; Sanchez-Carlo *et al*., 1984 ; Valente *et al*., 1988 ; Barrow et Hill, 1989 ; Nemeth *et al*., 1991, 1994 ; Fang et Pyörälä, 1996 ; Kaipainen *et al*., 2002). Les facteurs connus des bactéries à Gram négatif permettant cette résistance au sérum sont la lipoprotéine de la membrane externe TraT (Sukupolvi et O'Connors, 1990) et l'antigène capsulaire K1 (Montenegro *et al*., 1985). Lorsque le gène codant la protéine TraT et l'antigène K1 sont recherchés par PCR et lysotypie, respectivement, chez les souches MPEC, les résultats obtenus sont loin de refléter les fréquences élevées de résistance au sérum acquises par le test

phénotypique de croissance en sérum bovin. Par exemple, sur les 95 souches MPEC étudiées par Nemeth *et al*. en 1991, 61 (64 %) étaient capables de croître dans un sérum bovin, mais seules 41 (43 %) et 3 (3 %) possédaient le gène *traT* et l'antigène K1, respectivement (Nemeth *et al*., 1991). En 2002, Kaipainen *et al*. ont analysé 273 souches dont 193 (70 %) étaient capables de croître dans un sérum bovin, mais seules 105 (38 %) possédaient le gène *traT* et aucune ne présentait l'antigène K1 (Kaipainen *et al*., 2002). Ces travaux montrent donc bien que les propriétés des MPEC ne peuvent pas toutes être expliquées par les connaissances actuelles des souches *E. coli* pathogènes. La même conclusion peut être formulée lorsque la résistance aux antibiotiques des MPEC est analysée. Une étude réalisée par Srinivasan *et al.* en 2007 a ainsi évalué la résistance de 135 souches aux 10 et 16 antibiotiques les plus couramment utilisés en médecines vétérinaire et humaine, respectivement (Srinivasan *et al*., 2007). Les résultats obtenus ont montré que d'un point de vue phénotypique, la plupart des souches étaient résistantes à au moins deux antibiotiques. Cependant, d'un point de vue génotypique, certaines de ces résistances ne s'expliquaient pas par la présence, dans le génome des souches, de gènes connus pour coder les facteurs de résistance correspondant.

4.2.2. Constance du lipopolysaccharide

Le seul facteur de virulence retrouvé de manière constante chez toutes les souches *E. coli* inductrices de mammites bovines est le lipopolysaccharide, ou LPS, constitué comme son nom l'indique uniquement de sucres et de lipides. Le LPS est un composant majeur de la membrane externe des bactéries à Gram négatif, pouvant être libéré lors de la croissance ou de la mort de ces dernières (Petsch et Anspach, 2000). Il est composé de trois entités synthétisées séparément, puis assemblées : le lipide A, le noyau et l'antigène O (Szalo *et al*., 2006). Le lipide A est ancré dans la membrane externe et représente la partie proximale du LPS. Il est fortement conservé chez les entérobactéries et porte une activité endotoxinique, étant capable d'induire une réponse immunitaire non spécifique chez les organismes hôtes. Le noyau correspond à la partie médiane du LPS et ne porte aucune fonction connue à l'heure actuelle. Enfin, l'antigène O, partie distale libre très variable du LPS, porte la spécificité antigénique des souches et leurs apporterait, selon le nombre et la nature des sucres qui le composent, une résistance plus ou moins forte à la molécule du complément. Au sein de l'espèce *E. coli*, plus de 170 antigènes O différents sont recensés (Nikaido, 1994). Parmi les LPS des souches MPEC, aucun antigène O ne semble prédominer et de nouveau une grande variabilité est observée (Correa et Marin, 2002 ; Wenz *et al*., 2006).

4.2.3. Apparition de souches persistantes et invasives

D'après la classification établie par Blowey et Edmondson en 1995, les souches *E. coli* inductrices de mammites bovines étaient considérées, encore récemment, comme des souches environnementales opportunistes, à l'origine de mammites cliniques temporaires (Blowey et Edmondson, 1995). Cependant, comme il l'a été annoncé précédemment, cette conception est désormais de plus en plus discutée et rejetée ; les termes "clinique" et "temporaire" semblent en effet ne plus être appropriés systématiquement. Depuis la fin des années 1990, plusieurs études ont découvert que certains évènements de mammites rapprochés dans le temps, d'un même quartier ou de quartiers différents de la même mamelle, étaient dus à la présence de la même souche bactérienne (Lipman *et al*., 1995 ; Döpfer *et al*., 1999 ; Bradley et Green, 2000, 2001b). Ces études analysaient le génotype des différentes souches bactériennes par des techniques de REP et ERIC-PCR. Etant donné la diversité des souches *E. coli* présentes dans l'environnement, les auteurs de ces différentes études en ont conclu que les cas répétés de mammites n'étaient pas dus à de nouvelles infections par la même souche, mais plutôt à la persistance de la même souche dans la mamelle. Ces mammites rapprochées dans le temps ont été appelées mammites récurrentes et les souches qui en sont responsables ont été qualifiées de persistantes. Il a également été constaté que ces mammites récurrentes à *E. coli* présentaient une sévérité beaucoup moins importante que celle des mammites non récurrentes, pouvant passer même inaperçues, donc être sub-cliniques (Bradley et Green, 2001b). Enfin, en comparant les génotypes de souches persistantes isolées dans différents endroits géographiques, une similitude a pu être observée (Bradley et Green, 2001b). La découverte et l'analyse des mammites récurrentes à *E. coli* suggèrent donc une adaptation de certaines souches MPEC à la glande mammaire bovine. Etant donné leur capacité à persister dans la mamelle, ces dernières doivent en effet posséder des systèmes pour se défendre contre le système immunitaire bovin et ne pas être éliminées lors de la traite. Un hypothèse émise et vérifiée *in vitro*, pour expliquer cette persistance, est l'internalisation des bactéries dans les cellules épithéliales mammaires. Une étude a effectivement montré que des souches *E. coli* isolées de mammites étaient capables *in vitro* d'adhérer à et d'envahir des cellules épithéliales mammaires (Döpfer *et al*., 2000). Les souches utilisées lors de cette étude étaient isolées de mammites récurrentes ou non et celles issues de mammites récurrentes présentaient des pouvoirs d'adhésion et d'invasion plus efficaces que les autres. La recherche par PCR de facteurs d'adhésion et d'invasion similaires à ceux des pathotypes EPEC et EIEC chez ces souches MPEC persistantes a été infructueuse (Döpfer *et al*., 2001 ; Dogan *et al*., 2006). Ces résultats montrent donc de nouveau l'existence très probable de facteurs de virulence typiques et inconnus chez les souches MPEC.

4.3. Problématique abordée

L'objectif de ces travaux a été de caractériser une souche *E. coli* inductrice de mammites bovines, la souche *E. coli* P4, isolée en 1976 d'un cas de mammite clinique (Bramley, 1976). Cette souche est très utilisée au niveau mondial comme outil d'infection pour étudier les répercussions physiologiques de cette pathologie chez le bovin. Cependant, son mode et ses facteurs de virulence restent peu renseignés. Seules deux études ont montré d'un point de vue phénotypique quelques propriétés de virulence de cette souche, à savoir ses capacités *in vivo* à induire l'apoptose de cellules épithéliales mammaires chez le bovin (Long *et al*., 2001) et à adhérer à et envahir des cellules épithéliales mammaires chez la souris (Gonen *et al*., 2007). L'une de ces études a également prouvé qu'*E. coli* P4 était capable de provoquer une réaction inflammatoire chez la souris selon un mécanisme indépendant du LPS, suggérant fortement l'existence de facteurs de virulence autres que ce dernier chez cette bactérie et, plus généralement, chez les MPEC. Deux autres informations concernant cette souche sont également disponibles dans la littérature, à savoir son expression de deux antigènes de surface : l'O-antigène de type 32 et le H-antigène de type 37.

Les premiers travaux réalisés ont concerné le typage moléculaire d'*E. coli* P4, afin de savoir à quel groupe phylogénétique de l'espèce *E. coli* elle appartient et de quel type de souche elle se rapproche le plus (commensale ou pathogène).

Dans un deuxième temps, le rôle cette souche dans la protéolyse importante des caséines du lait observée lors d'une mammite, rôle suggéré lors d'études antérieures effectuées au laboratoire, a été évalué. Pour cela la recherche d'enzymes à activité protéolytique sécrétées par la souche a été entreprise.

Enfin, des propriétés moléculaires spécifiques à cette souche ont été recherchées. Les présences dans son génome de gènes codant quelques uns des facteurs de virulence les plus connus chez les souches *E. coli* pathogènes pour le bovin et le veau et d'éléments génétiques mobiles ou mobilisables tels que des îlots génomiques ont été analysées.

Résultats

Les résultats obtenus ont fait l'objet d'une affiche présentée lors du séminaire annuel de l'école doctorale RP2E en 2006 et de deux publications : l'une vient d'être acceptée par Journal of Dairy Science et l'autre est en préparation pour une soumission à International Journal of Medical Microbiology. Le chapitre qui suit, dans lequel sont insérées cette affiche et ces deux publications, présente une synthèse de ces résultats. Le matériel et les méthodes employés, déjà décrits dans les publications ne sont pas détaillés.

1. Etude de la croissance d'*E. coli* P4 dans un lait issu de mammite bovine

Jusqu'à la fin du 20ème siècle, pour la majorité des microbiologistes, l'idée prédominante était que les souches d'*E. coli* inductrices de mammites bovines étaient des pathogènes opportunistes. Il était ainsi considéré qu'après avoir gagné la glande mammaire, ces souches étaient uniquement capables de déclencher une inflammation, ceci par le biais de leur LPS et qu'elles étaient ensuite tout à fait incapables de survivre et éliminées par les acteurs du système immunitaire bovin. Depuis cette même fin du 20ème siècle, plusieurs études ont, au contraire, rapporté l'existence de souches capables de persister dans la mamelle bovine et d'y induire des mammites récurrentes. Nous avons donc voulu dans un premier temps tester l'hypothèse du caractère opportuniste d'*E. coli* P4 en comparant sa croissance à celle d'une souche non pathogène bien connue, *E. coli* K12 MG1655, dans un lait modèle reproduisant les conditions environnementales d'inflammation créées lors d'une mammite. Ces conditions furent reproduites en infusant un inoculum de LPS (O26:B6) dans un quartier d'une mamelle d'une vache. Le lait modèle a ainsi été collecté huit heures après l'infusion intramammaire, temps correspondant au pic de l'inflammation, où une concentration maximale de cellules somatiques est enregistrée (Moussaoui *et al*., 2002). Les croissances des deux souches ont également été étudiées dans un lait sain, collecté chez le même animal juste avant l'infusion. Les propriétés de ces deux laits sont présentées dans le tableau 16.

Tableau 16 : Caractéristiques des deux laits utilisés pour comparer la croissance des souches *E. coli* P4 et *E. coli* K12 MG1655. (1) : les contaminations bactériennes ont été analysées après un étalement de 100 µL de chaque lait sur le milieu de culture microbiologique Luria-Bertani solide et une incubation de 24 et 48 heures à 37°C. (2) : la concentration en cellules somatiques a été déterminée à l'aide d'un compteur fluoro-opto-électronique utilisant la cytométrie en flux.

	Contamination bactérienne (1)	pH	Concentration en cellules somatiques (nombre.mL^{-1}) (2)
Lait modèle reproduisant une inflammation de type mammite	Aucune	6,98	$3,56.10^6$
Lait sain	Aucune	6,77	21.10^3

Les valeurs plus élevées de pH et de concentrations en cellules somatiques obtenues pour le lait modèle étaient bien caractéristiques d'une inflammation. Elles étaient en relation en effet avec la présence, notamment, dans un lait issu de mammite bovine, d'entités oxydantes, comme l'anion superoxyde (O_2^-), l'eau oxygénée (H_2O_2), le radical hydroxyl (OH^-), l'ion hypochloreux (OCl^-) et de cellules immunitaires, comme des neutrophiles (Rainard et Riollet, 2006). Avant toutes inoculations et suivis de croissance, les deux laits récupérés ont été écrémés afin de faciliter des analyses ultérieures. Il a été montré qu'une étape d'écrémage ne changeait pas les propriétés d'inflammation d'un lait issu de mammite bovine (Lohius *et* al., 1988). Dans le lait sain, une croissance identique des deux souches a pu être observée, débutant par une phase de latence de 1h30min environ, suivie d'une phase exponentielle de 3h avec un temps de génération de 30 minutes environ, puis d'une phase stationnaire (figure 20A).

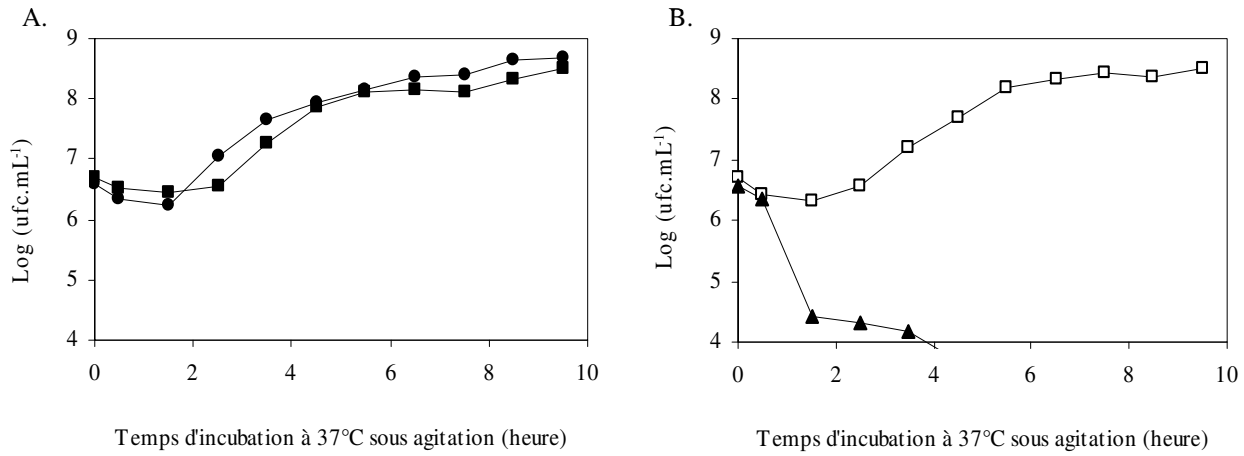


Figure 20 : Courbes de croissance d'*E. coli* P4 et d'*E. coli* K12 MG1655 dans un lait sain (A) (■ et ●, respectivement) et dans un lait modèle reproduisant une inflammation de type mammite (B) (□ et ▲, respectivement).

Dans le lait modèle mimant l'inflammation, des croissances très différentes des deux souches ont pu être observées (figure 20B). Celle de la souche *E. coli* P4 était quasiment identique à celle observée dans le lait sain : une phase de latence de 2h environ, suivie d'une phase exponentielle de 3h30min avec un temps de génération de 30 minutes environ, puis d'une phase stationnaire. Pour la souche *E. coli* K12 MG1655, aucune croissance bactérienne n'a pu être détectée. Au contraire, une diminution de la population bactérienne a été observée.

Ainsi, ces résultats suggéraient l'existence chez *E. coli* P4 de propriétés particulières, permettant son adaptation à la glande mammaire bovine et absentes chez *E. coli* K12 MG1655.

2. Typage moléculaire d'*E. coli* P4

Nous avons effectué un typage moléculaire basé sur l'analyse du génome core d'*E. coli* P4, afin de savoir de quel type de souches (pathogènes ou non pathogènes) elle se rapproche le plus d'un point de vue phylogénétique. En effet, bien que les souches MPEC soient toujours comparées aux souches non pathogènes commensales, aucune étude phylogénétique les concernant ne semble avoir été réellement menée jusqu'à présent. A notre connaissance, aucune publication rapportant ce type de travaux n'est disponible dans la littérature. Le matériel et les méthodes expérimentales utilisés ainsi que les résultats obtenus pour cette partie du travail sont exposés dans l'affiche et la publication insérées à la fin du paragraphe 4 de ce chapitre.

2.1. Détermination du groupe phylogénétique et séquençage d'un ITS 16S-23S

Les souches de l'espèce *E. coli* se répartissent dans quatre principaux groupes phylogénétiques : A, B1, B2 et D. Afin de déterminer à quel groupe phylogénétique la souche *E. coli* P4 appartient, la méthode développée par Clermont *et al*. (2000) a été employée. Les présences des gènes *chuA* et *yjaA* et du fragment d'ADN tspE4C2 au sein de son génome ont été recherchées par PCR, grâce aux couples d'oligonucléotides chuA-F/chuA-R, yjaA-F/yjaA-R et tspE4C2-F/tspE4C2-R, respectivement (Clermont *et al*., 2000). Les résultats obtenus ont montré une seule amplification positive et de taille attendue pour le couple yjaA-F/yjaA-R, indiquant la présence du gène *yjaA* et l'absence du gène *chuA* et du fragment d'ADN tspE4C2 au sein du génome d'*E. coli* P4, assignant ainsi cette souche au groupe phylogénétique A. Etant donné que le groupe A rassemble des souches aussi bien non pathogènes (comme les souches de la collection ECOR numéros 1, 4, 10, 13, 17, 20, 23 et 24) que pathogènes (comme les souches *E. coli* JM221, DAEC7, EDL1493, EIEC85b et DEC9a, entéroagrégative, à adhésion diffuse, entérotoxigénique, entéroinvasive et entérohémorragique, respectivement) (Escobar-Paramo *et al*., 2004), nous avons décidé d'analyser un second marqueur phylogénétique bactérien : la séquence nucléotidique de la région séparant les gènes codant les ARNr 16S et 23S (l'ITS 16S-23S). Cette région permet notamment au sein de l'espèce *E. coli* de différencier des souches du même groupe phylogénétique (Anton *et al*., 1998). C'est le cas par exemple, des souches *E. coli* K12 MG1655 et ECOR 10 appartenant toutes deux au groupe phylogénétique A, mais dont les séquences de l'ITS 16S-23S de l'opéron ribosomique *rrnE* présentent 99,4 % d'identité. En utilisant le couple d'oligonucléotides 16S14F/23S1R (Garcia-Martinez *et al*., 1996), deux produits d'amplification de 650 et 750 pb ont ainsi été obtenus par PCR à partir de l'ADN total d'*E. coli* P4. Le séquençage du produit de 650 pb a révélé la présence d'un ITS 16S-23S de 354 pb, présentant

100 % d'identité avec l'ITS 16S-23S de l'opéron ribosomique *rrnE* du génome de la souche non pathogène *E. coli* K12 MG1655 appartenant au groupe phylogénétique A.

2.2. Séquençage de multiples loci (MLST)

Afin de savoir si l'ensemble du génome core d'*E. coli* P4 se rapprochait fortement de celui de la souche non pathogène *E. coli* K12 MG1655, nous avons utilisé une autre méthode d'analyse phylogénétique bactérienne : la MLST (Multi Locus Sequence Typing). Cette dernière consiste à déterminer, au sein des génomes bactériens, les séquences nucléotidiques partielles de gènes constants ayant évolué plus vite notamment que les régions intergéniques des opérons ribosomiaux. Elle peut ainsi parfois permettre une distinction entre deux souches appartenant au même groupe phylogénétique et présentant une même séquence d'ITS 16S-23S. C'est le cas par exemple, des souches *E. coli* K12 MG1655 et HS, appartenant toutes deux au groupe phylogénétique A et possèdant un ITS 16S-23S de l'opéron ribosomique *rrnD* identique, mais dont les séquences du gène constant *icdA* présentent 99 % d'identité. Trois groupes de gènes constants ont été examinés. Le premier groupe, constitué de six gènes (*adk*, *gcl*, *gdh*, *mdh*, *metA*, *ppk*), a été décrit par Adiri *et al*. (2003) dans le but d'effectuer une analyse MLST de l'espèce *E. coli* en général. Les séquences nucléotidiques partielles de ces six gènes, déterminées chez *E. coli* P4, présentent 100 % d'identité avec les séquences correspondantes au sein du génome d'*E. coli* K12 MG1655. Le deuxième groupe était composé de quatorze gènes constants (*aroE*, *aspC*, *clpX*, *cstA*, *cyaA*, *dnaG*, *fadD*, *grpE*, *icdA*, *lysP*, *mtlD*, *mutS*, *pgi*, *rpoS*) identifiés par Reid *et al*. (2002) pour une analyse MLST des souches *E. coli* pathogènes cette fois-ci. Sur ces quatorze gènes, les séquences nucléotidiques partielles de onze d'entre eux (*aroE*, *aspC*, *clpX*, *cstA*, *dnaG*, *fadD*, *icdA*, *lysP*, *mutS*, *pgi* et *rpoS*), déterminées chez *E. coli* P4, sont identiques à 100 % aux séquences correspondantes chez *E. coli* K12 MG1655 et celles de trois d'entre eux (*cyaA*, *grpE* et *mtlD*) s'en éloignent (avec 99,80, 99,75 et 99,81 % d'identité, respectivement). Pour ces trois dernières, les pourcentages d'identité les plus élevés restent toutefois ceux obtenus avec les séquences correspondantes chez *E. coli* K12 MG1655. Le troisième groupe analysé, choisi par nous-même, était constitué de neuf gènes codant des protéases putatives (*degQ*, *elaD*, *pmbA*, *ptrB*, *sms*, *yaeL*, *ycaL*, *yhjJ*, *yhjS*) jamais utilisés jusqu'à présent pour une analyse MLST de l'espèce *E. coli*, mais retrouvés sur tous les génomes de cette espèce sequencés à l'heure actuelle et montrant des différences nucléotidiques entre différentes souches *E. coli*. Sur ces neuf gènes, les séquences nucléotidiques partielles de six d'entre eux (*degQ*, *ptrB*, *yaeL*, *ycaL*, *yhjJ* et *yhjS*), déterminées chez *E. coli* P4, sont identiques à 100 % aux séquences correspondantes chez *E. coli* K12 MG1655 et celles de trois d'entre eux (*elaD*, *pmbA*, et *sms*) s'en éloignent (avec 99,75, 98,72 et 96,85 % d'identité, respectivement). Pour les séquences partielles des gènes *elaD* et *pmbA*, les pourcentages

d'identité les plus élevés restent, comme précédemment, ceux obtenus avec les séquences correspondantes chez *E. coli* K12 MG1655. Pour la séquence partielle du gène *sms*, au contraire, un pourcentage d'identité plus élevé que 96,85 % a pu être obtenu : il s'agit d'un taux d'identité de 98,25 % avec la séquence correspondante chez *E. coli* O157:H7 Sakaï.

Ces résultats du séquençage de multiples loci ont donc montré qu'avec les marqueurs habituellement utilisés pour une analyse MLST globale de l'espèce *E. coli* le génome d'*E. coli* P4 se rapproche de celui d'*E. coli* K12 MG1655. A l'opposé, l'utilisation de marqueurs moins courants permet d'établir une distinction entre ces deux génomes. Les résultats de typage moléculaire obtenus suggèrent tout de même, dans leur ensemble, une plus forte proximité phylogénétique de la souche mammopathogène *E. coli* P4 avec la souche commensale non pathogène *E. coli* K12 MG1655.

3. Etude du rôle d'*E. coli* P4 dans la caséinolyse du lait observée lors d'une mammite

Des travaux antérieurs effectués au laboratoire avaient porté sur l'analyse de la protéolyse du lait observée lors d'une mammite (Michelutti *et al*., 1999). Les rôles des acteurs endogènes du lait responsables de cette caséinolyse, tels que le système plasmine-plasminogène et les protéases des cellules somatiques recrutées au cours de l'infection, avaient été bien étudiés lors de mammites induites par LPS et par *E. coli* P4 (Moussaoui *et al*., 2002 et 2004). Ces études avaient suggéré l'existence d'autres acteurs responsables de cette protéolyse, de nature exogène cette fois-ci, c'est-à-dire d'origine bactérienne. En effet, en comparant les conséquences de mammites expérimentales à LPS et à *E. coli*, induites par infusion intramammaire de 10 µg de LPS (LPS O26:B6) et de 10^6 ufc d'une souche *E. coli* (*E. coli* P4), respectivement, il avait été observé que l'intensité de la protéolyse était différente dans les deux cas : le taux de protéoses peptones (fraction peptidique dérivée de l'hydrolyse des caséines du lait) maximal enregistré était plus élevé lors d'une mammite expérimentale à *E. coli* (6,52 $g.L^{-1}$ contre 2,49 $g.L^{-1}$ pour le cas LPS) et le nombre de peptides produits dans la fraction protéoses peptones était également supérieur. Le plus fort taux de caséinolyse observé lors d'une mammite à *E. coli* ne semblait pas pouvoir être expliqué par une augmentation plus forte de l'activité de la plasmine ou par un recrutement plus massif de cellules somatiques. Les valeurs maximales atteintes pour ces deux paramètres étaient en effet plus faibles dans le cas de la mammite expérimentale à *E. coli* que dans le cas de la mammite expérimentale à LPS : une activité maximale de la plasmine de 41,4 et 75,8 µmol de ρ–nitroanilide.h^{-1}.L^{-1} et une concentration maximale de cellules somatiques de $3{,}98.10^6$ et 27.10^6 cellules.mL^{-1} étaient enregistrées lors des mammites expérimentales à *E. coli* et à LPS, respectivement. Deux hypothèses impliquant un rôle des bactéries *E. coli* avaient

alors été formulées. La première suggérait un rôle indirect, où les bactéries seraient capables de synthétiser des activateurs des protéases des cellules somatiques et/ou des facteurs induisant une maturation différente de ces cellules, donc de leurs protéases synthétisées. La seconde proposait un rôle direct, où les bactéries seraient capables de sécréter dans leur milieu environnant des protéases extracellulaires pouvant hydrolyser de façon notable les caséines du lait. Nous nous sommes intéressés à tester la deuxième hypothèse. Le rôle direct des bactéries *E. coli* dans la protéolyse du lait observée lors d'une mammite n'avait en effet jamais été étudié. Le matériel et les méthodes utilisés ainsi que les résultats obtenus pour cette partie du travail sont exposés dans la publication insérée à la fin de ce paragraphe.

3.1. Mise en évidence d'une sécrétion de protéases extracellulaires par la technique de zymographie

La sécrétion de protéases extracellulaires par *E. coli* P4 a été étudiée dans quatre milieux de culture différents et à différents temps de croissance. Un lait modèle reproduisant les conditions environnementales d'inflammation créées lors d'une mammite et un lait sain, dont les caractéristiques ont déjà été présentées dans la première partie de ce chapitre et dans le tableau 16, ont été utilisés. De même, deux milieux classiques de culture pour les bactéries *E. coli* ont été testés : l'un appelé milieu synthétique (MS), de composition définie, pauvre en nutriments, dont les seules sources de carbone et d'azote sont le glycérol et le sulfate d'ammonium, respectivement et l'autre appelé Luria-Bertani (LB), de composition complexe, riche en protéines, peptides et acides aminés libres. L'utilisation de ces quatre milieux permettait de tester si la sécrétion / activité des protéases était dépendante du milieu. *E. coli* P4 a ainsi été cultivée dans chacun de ces milieux, sous agitation, à une température de 37°C (température du lait *in vivo*) et en présence d'un tampon MOPS de pH 6,8 (pH du lait *in vivo*) évitant l'acidification des milieux *in vitro* qui accompagne la croissance bactérienne. Les croissances bactériennes obtenues sont présentées dans la figure 21.

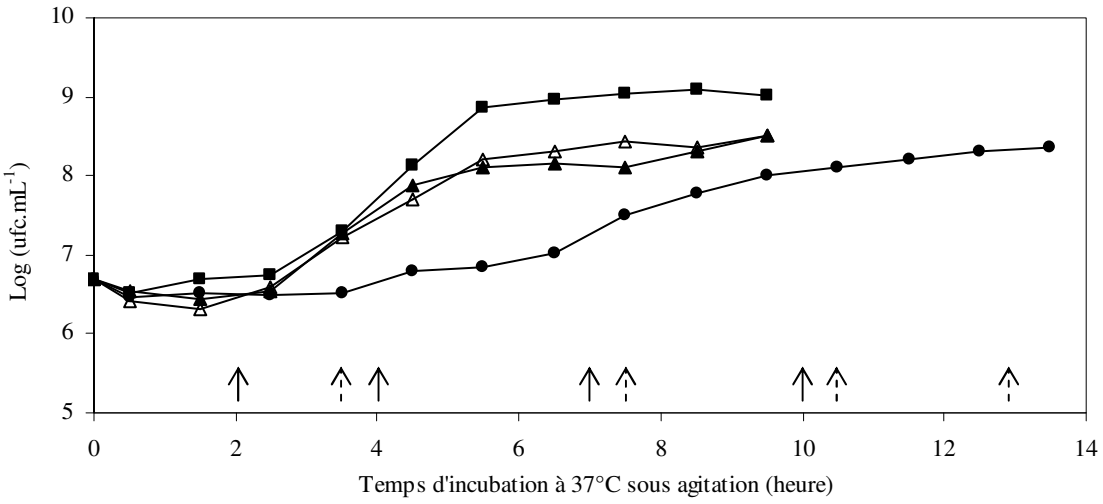


Figure 21 : Courbes de croissance d'*E. coli* P4 dans un lait modèle reproduisant une inflammation de type mammite (Δ), dans un lait sain (▲), dans un milieu LB (■) et dans un milieu MS (●). Sur l'échelle des abscisses, les flèches correspondent aux temps d'incubation auxquels les milieux extracellulaires de culture des milieux laits et LB (trait plein) et MS (trait pointillé) ont été prélevés pour l'analyse par zymographie.

Les milieux extracellulaires ont été récupérés après 2, 4, 7 et 10 heures d'incubation dans les deux milieux lait et le milieu LB et après 3,5, 7,5, 10,5 et 13 heures d'incubation dans le milieu MS. Ces quatre temps correspondaient à quatre stades différents de croissance : à une phase de latence, à une phase exponentielle, puis à un début de phase stationnaire et enfin à une phase stationnaire (figure 21). Avant toute analyse ultérieure, nous avons déterminé si les contenus extracellulaires obtenus ne pouvaient être contaminés par des constituants intracellulaires, suite à une lyse partielle ou totale des bactéries cultivées. Pour cela, la présence éventuelle d'acides nucléiques a été recherchée grâce à la méthode décrite par Moe *et al*. (1994), en utilisant comme agent intercalant fluorescent le bisBenzimide (Hoescht 33258). La démarche expérimentale suivie a consisté à ajouter aux échantillons de milieux extracellulaires du bisBenzimide à une concentration finale de 0,01 $mg.mL^{-1}$ et à évaluer la fluorescence de ces mélanges (λ excitation = 346 nm et λ émission = 460 nm). Une gamme étalon d'ADN de saumon, traitée de la même façon, a permis d'en déduire les concentrations d'ADN présentes dans les échantillons. Aucune trace d'acides nucléiques n'a pu être détectée dans l'ensemble des échantillons récupérés, ceci à tous les stades de croissance et dans tous les milieux testés (la concentration la plus faible d'ADN de saumon étalon utilisée comme seuil de détection étant de 5 $\mu g.mL^{-1}$). La recherche de protéases a ensuite été effectuée grâce à la technique de zymographie, en utilisant comme substrat les caséines. Cette technique consiste à faire migrer un échantillon de protéines sur un gel d'électrophorèse SDS-PAGE contenant, dans le gel de séparation, des caséines. Au terme de la migration électrophorétique, après une étape de renaturation des protéines séparées, le gel est incubé à une température et dans une solution favorisant l'attaque des caséines par les éventuelles protéases présentes dans l'échantillon déposé. Une étape de coloration finale du gel au bleu de Coomassie permet de révéler ces éventuelles protéases. En effet, de par la présence de

caséines, tout le gel se colore en bleu ; si une hydrolyse des caséines a eu lieu, des bandes incolores apparaissent témoignant de la présence de protéases. Aux quatre phases de croissance testées pour les échantillons extracellulaires des milieux lait sain, LB et MS, quatre mêmes plages de lyse des caséines ont pu être observées par zymographie, signalant la présence d'au moins quatre protéases de masses moléculaires apparentes proches de 96 kDa (protéase A), 54 kDa (protése B), 23 kDa (protéase C) et 21 kDa (protéase D) (figure 22).

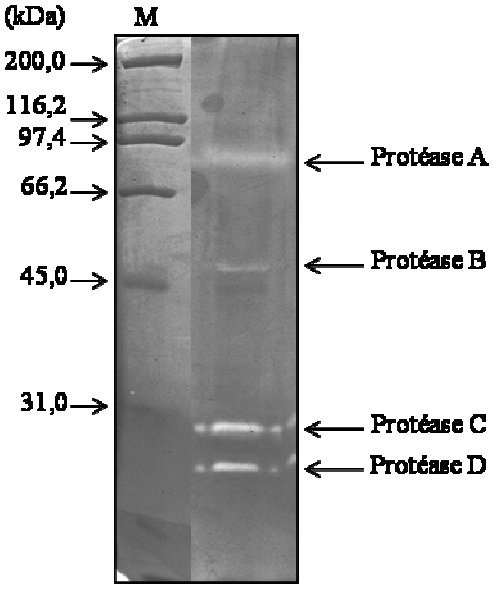


Figure 22 : Analyse par zymographie caséines des contenus extracellulaires de culture en milieu MS d'*E. coli* P4 aux phases de latence, exponentielle et stationnaire de croissance. Une quantité de 20 µg de protéines extracellulaires totales a été déposée. Les mêmes zymogrammes ont été obtenus en milieux lait sain et LB. M : marqueurs de masse moléculaire.

Ces plages de lyse étaient spécifiques des échantillons extracellulaires bactériens, c'est-à-dire absentes des mêmes milieux non inoculés avec la bactérie, suggérant donc la sécrétion d'au moins quatre protéines extracellulaires à activité caséinolytique par *E. coli* P4. Les zymogrammes obtenus avec les échantillons extracellulaires bactériens issus du milieu lait modèle reproduisant une inflammation de type mammite n'ont pas pu être interprétés. En effet, ce milieu seul est déjà caractérisé par un contenu en nombreuses protéases de masses moléculaires proches de 90, 50 et 25 kDa (Haddadi *et al*., 2006) et la distinction entre les protéases endogènes du lait et celles exogènes des bactéries n'a pu être ainsi faite par zymographie.

Les résultats obtenus ont donc montré qu'*E. coli* P4 est capable de sécréter au moins quatre protéases extracellulaires à activité caséinolytique dans trois milieux de culture et à quatre stades de croissance testés. Cette sécrétion ne semble donc ni milieu spécifique ni inductible à un stade de croissance particulier.

3.2. Evaluation de l'activité caséinolytique d'*E. coli* P4 et des quatre protéases extracellulaires mises en évidence

Dans un premier temps, l'activité caséinolytique globale d'*E. coli* P4 a été étudiée en incubant la bactérie à 37°C dans le milieu lait sain et en suivant au cours du temps les évolutions quantitatives par FPLC et qualitatives par électrophorèse SDS-PAGE des teneurs de chaque caséine et des teneurs en protéoses peptones, respectivement. Cinq temps d'incubation ont été retenus pour ces analyses : 2, 4, 7, 10 et 24 heures. Pour chacun de ces temps, la présence des quatre enzymes extracellulaires bactériennes a été vérifiée par zymographie caséines. Un témoin correspondant au même lait non inoculé a été analysé en parallèle. Aucune variation significative des teneurs en chaque caséine n'a pu être observée entre les différents temps d'incubation testés et entre les laits inoculés et non inoculés (figure 23).

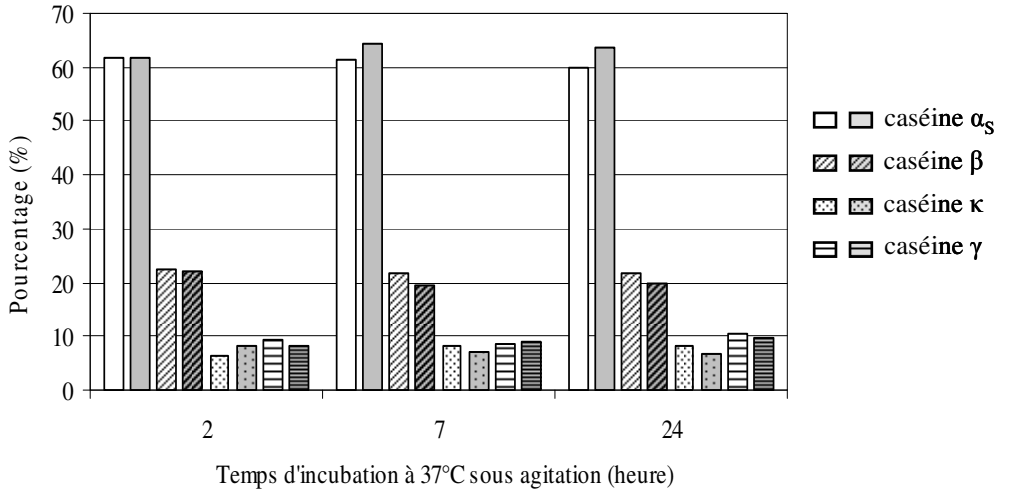


Figure 23 : Suivi quantitatif des teneurs de chaque caséine après 2, 7 et 24 heures d'incubation à 37°C sous agitation d'un lait inoculé avec *E. coli* P4 (histogrammes sur fond blanc) ou non (histogrammes sur fond gris).

De même, aucune évolution des profils électrophorétiques des contenus en protéoses peptones n'a pu être constatée entre les différents temps d'incubation testés et entre les laits inoculés et non inoculés (figure 24). Aucune activité caséinolytique globale de la souche n'a donc pu être détectée lors de cette incubation en milieu lait sain. Le milieu lait modèle mimant l'inflammation de type mammite n'a pas été utilisé pour cette expérience étant donné son contenu en enzymes protéolytiques endogènes déjà important et l'hydrolyse avancée des caséines qu'il renferme.

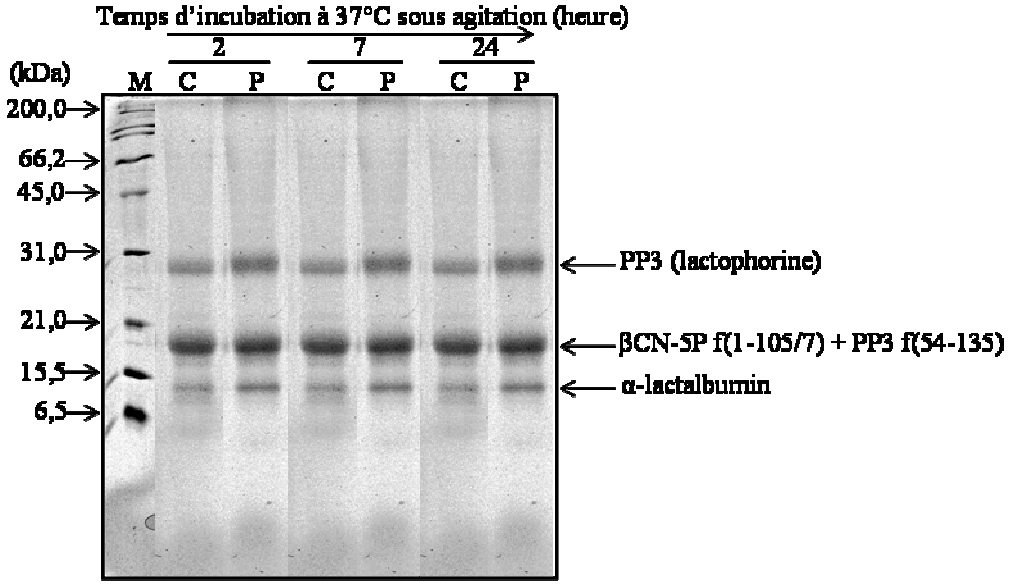


Figure 24 : Suivi qualitatif des teneurs en protéoses peptones après 2, 7 et 24 heures d'incubation à 37°C sous agitation d'un lait inoculé avec *E. coli* P4 (P) ou non (C). Des quantités de 60 µg de protéoses peptones ont été déposées par puits. M : marqueurs de masses moléculaires.

L'absence de détection d'activité caséinolytique lors de l'expérience précédente pouvait être expliquée par l'utilisation du lait entier comme milieu d'incubation. Ce dernier est en effet riche en urée et acides aminés libres, donc riche en sources d'azote et de carbone directement assimilables par la bactérie, ce qui peut avoir un effet inhibiteur sur l'activité des protéases, si celles-ci interviennent dans l'utilisation des caséines comme sources carbonée et azotée. L'activité caséinolytique globale d'*E. coli* P4 a donc été étudiée plus directement en incubant la bactérie dans un milieu dont les seules sources de carbone et d'azote sont des caséines. Ce milieu correspondait au milieu MS dépourvu de glycérol et de sulfate d'ammonium, mais complété avec une solution de caséines à une concentration similaire à celle retrouvée dans le lait. Dans ces conditions, pour se développer, la bactérie était obligée d'hydrolyser les caséines du milieu pour produire des sources disponibles et assimilables de carbone et d'azote. Après plus de 200 heures d'incubation dans ce milieu, aucune croissance bactérienne n'a pu être observée (figure 25). Les contenus de chaque caséine évalués qualitativement par électrophorèse SDS-PAGE sont restés inchangés (figure 25), malgré la détection dans le milieu de culture, dès cinq heures d'incubation, des quatre protéases extracellulaires bactériennes mises en évidence. Ce résultat suggérait donc que les enzymes détectées ne présentaient aucune activité caséinolytique significative pour permettre la croissance de la bactérie.

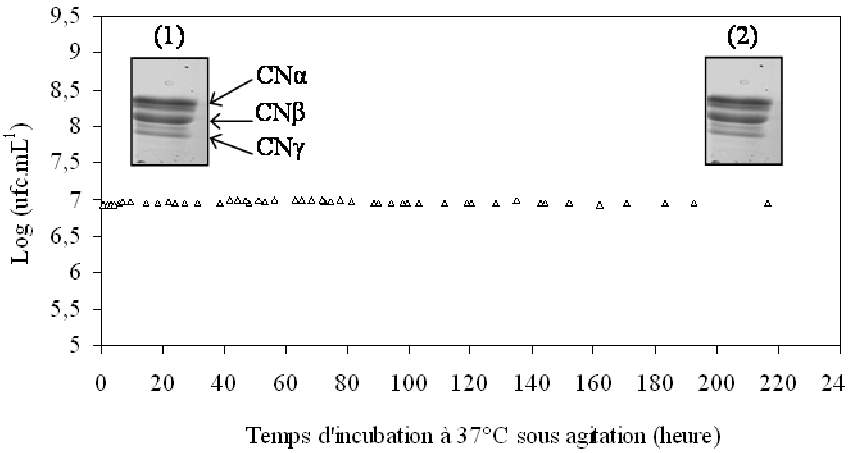


Figure 25 : Courbe de croissance d'*E. coli* P4 en milieu MS dépourvu de glycérol et de sulfate d'ammonium, mais complété avec une solution de caséinates à 0,15 % (m/v) ; et profils électrophorétiques SDS-PAGE du contenu en caséines du milieu après 0 (1) et 200 (2) heures d'incubation à 37°C (une quantité de 20 µg de protéines totales a été déposée pour les deux temps d'incubation).

Afin de confirmer l'absence d'activité caséinolytique des quatre protéases extracellulaires sécrétées par *E. coli* P4, une dernière expérience a été menée. Les milieux extracellulaires de la bactérie, récoltés après 10 et 13 heures d'incubation dans les milieux LB et MS, respectivement, ont été incubés avec des solutions de caséines totales, de caséine α_{S1}, de caséine α_{S2} et de caséines β+κ préparées à une concentration de 1 mg.mL^{-1} en tampon phosphate de sodium 0,05 M pH 6,8. Des mélanges de différents rapports « quantité de protéines extracellulaires bactériennes / quantité de caséines » (1/1, 10/1, 20/1, 50/1 et 100/1) ont ainsi été incubés pendant 24 et 48 heures à 37°C et analysés ensuite sur gel d'électrophorèse SDS-PAGE. Pour chaque solution de caséines (en mélange ou purifiées), chaque milieu, chaque rapport « quantité de protéines extracellulaires bactériennes / quantité de caséines » et chaque temps d'incubation testés, aucune hydrolyse des caséines par les protéases extracellulaires d'*E. coli* P4 n'a pu être détectée. Aucune évolution des profils électrophorétiques des différentes caséines n'a pu être en effet observée (figure 26). Cette expérience n'a pas été réalisée avec les milieux extracellulaires de la bactérie cultivée dans le lait modèle reproduisant l'inflammation de type mammite et le lait sain, étant donné le contenu déjà trop riche en caséines de ces milieux. Le résultat obtenu confirmait donc l'absence d'activité caséinolytique significative des quatre enzymes bactériennes extracellulaires mises en évidence.

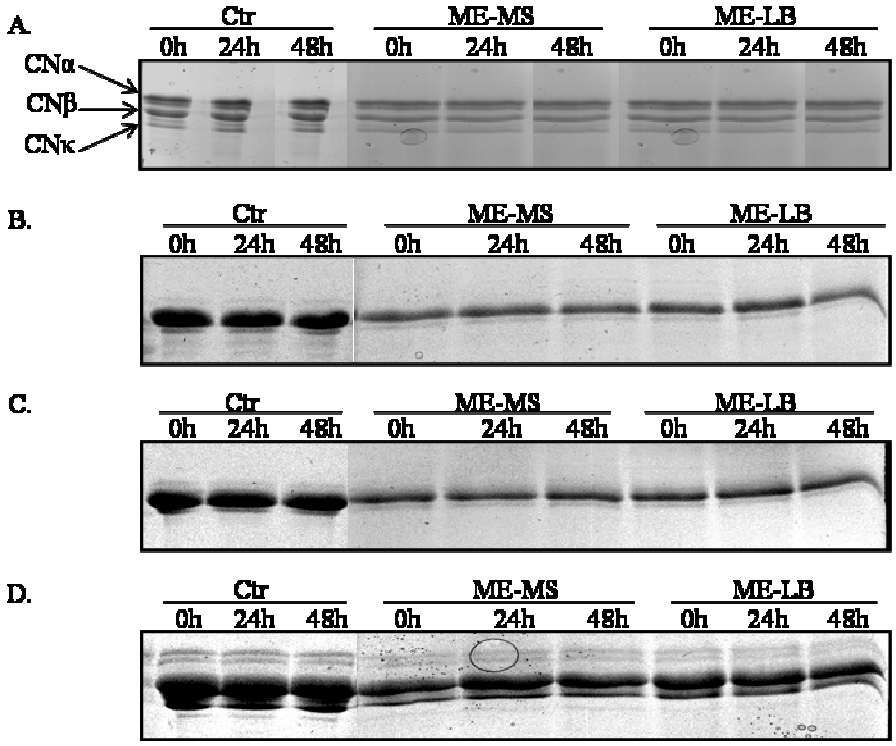


Figure 26 : Profils électrophorétiques SDS-PAGE obtenus après l'incubation, pendant 0, 24 et 48 heures à 37°C, de caséines totales (A), de caséines α_{s1} (B), de caséines α_{s2} (C) et de caséines β+κ (D), seules (Ctr) ou mélangées dans un rapport 1/1 avec des milieux extracellulaires d'*E. coli* P4 cultivée 13 heures en milieu MS (ME-MS) et 10 heures en milieu LB (ME-LB).

La conclusion de cette étude est que, bien qu'une sécrétion effective d'au moins quatre protéases extracellulaires capables d'hydrolyser les caséines ait été mise en évidence par zymographie, les différents protocoles expérimentaux mis en œuvre n'ont pas permis de révéler une activité caséinolytique significative d'*E. coli* P4. Ainsi, lors d'une mammite bovine, si le taux d'expression de ces quatre protéases n'est pas augmenté ou si ces quatre protéases ne présentent pas une activité protéolytique plus élevée sur les caséines (notamment les caséines oxydées) ou si d'autres protéases ne sont pas exprimées en supplément *in vivo*, un rôle direct d'*E. coli* P4 dans la caséinolyse du lait observée semble donc peu probable.

3.3. Essai de purification des protéases extracellulaires mises en évidence

Bien que les quatre enzymes extracellulaires mises en évidence chez *E. coli* P4 ne semblaient pas présenter une activité caséinolytique significative, le fait que cette bactérie puisse sécréter des protéases constituait un résultat intéressant. En effet, tout comme il l'a été démontré pour différentes souches *E. coli* pathogènes, ces protéases pourraient correspondre à des facteurs de virulence nécessaires à la bactérie pour attaquer les tissus hôtes et s'installer durablement au sein de la glande mammaire. Nous avons donc essayé de purifier les quatre enzymes extracellulaires mises en évidence

afin de les identifier. Le choix du milieu de culture à utiliser pour cette purification s'est porté sur le milieu synthétique. Ce dernier, dépourvu de protéine, n'amenait en effet aucun contaminant extérieur qui aurait pu rendre plus difficile la purification des protéines bactériennes. Un temps d'incubation de 13 heures de la bactérie dans ce milieu (à 37°C et sous agitation), avant la récupération de son milieu extracellulaire, a été choisi. Ce temps correspondait à un stade stationnaire de croissance de la bactérie (figure 21), où les quatre protéases avaient pu être détectées. Le choix de la phase stationnaire s'expliquait également par le fait qu'une concentration maximale de bactérie était obtenue à ce stade et donc, qu'une quantité maximale de protéases sécrétées dans le milieu pouvait être attendue. Après 13 heures d'incubation dans 18 litres de milieu MS, le milieu extracellulaire de culture d'*E. coli* P4 a ainsi été récupéré par centrifugation et filtré sur un filtre dont le diamètre des pores était de 0,45 µm. Le contenu protéique de ce milieu extracellulaire, analysé par électrophorèse SDS-PAGE monodimensionnelle, semblait être assez riche (figure 27). Sur l'électrophorégramme obtenu, plusieurs bandes, dont la migration électrophorétique pouvait correspondre à celle des protéases C et D, pouvaient être en effet observées. Par ailleurs, aucune bande nette de migration électrophorétique similaire à celle des protéases A et B ne pouvait être distinguée, du fait de la présence, aux faibles distances de migration, d'une traînée contenant probablement plusieurs bandes protéiques mal séparées. Un fractionnement du milieu extracellulaire semblait donc nécessaire pour la purification des quatre protéases.

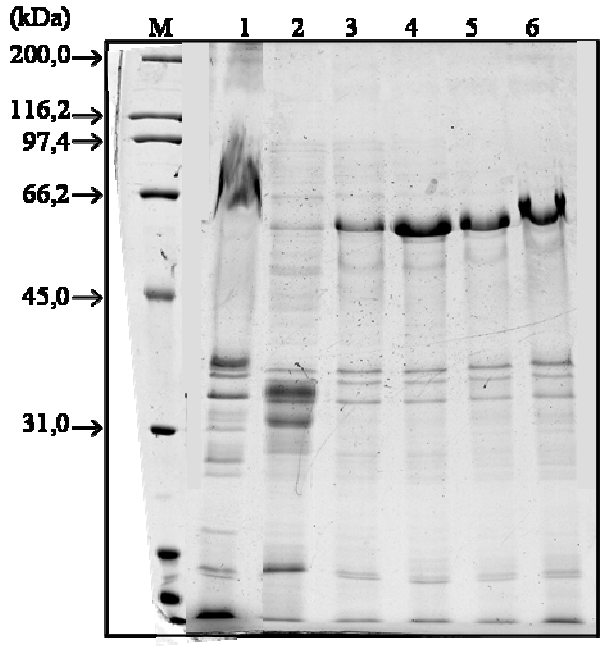


Figure 27 : Analyse par électrophorèse SDS-PAGE monodimensionnelle du milieu extracellulaire d'*E. coli* P4 incubée 13 heures à 37°C sous agitation en milieu synthétique. M : marqueurs de masse moléculaire. 1 : milieu extracellulaire brut. 2, 3, 4, 5, et 6 : milieu extracellulaire précipité avec 20, 40, 60, 80 et 100 % de sulfate d'ammonium, respectivement. Des quantités de 20 µg de protéines extracellulaires totales ont été déposées par puits.

La première technique de purification biochimique utilisée a été la précipitation au sulfate d'ammonium. Une précipitation fractionnée du milieu extracellulaire avec 20, 40, 60, 80 et 100 % de sulfate d'ammonium a été effectuée. Cette précipitation a permis d'améliorer la migration électrophorétique et donc la séparation des protéines de plus grandes masses moléculaires (figure 27).

Cependant, plusieurs bandes, de migration électrophorétique similaire à celle des quatre protéases mises en évidence, pouvaient être toujours observées dans chacun des culots de précipitation obtenus. Pour améliorer donc le fractionnement du milieu extracellulaire, une seconde technique de purification biochimique a été utilisée : la chromatographie échangeuse d'ions. Après avoir été, au préalable, dialysés, les différents culots de précipitation au sulfate d'ammonium obtenus ont ainsi été ajustés à un pH de 8,5 à l'aide d'un tampon Tris 0,05 M, puis déposés sur une résine échangeuse d'anions (Deae23, Whatman) équilibrée au même pH avec le même tampon. Ensuite, pour chaque culot, une élution fractionnée des constituants extracellulaires a été effectuée, en utilisant successivement les tampons suivants : un tampon phosphate de sodium 0,05 M pH 7, un tampon Mes 0,05 M pH 6, un tampon acétate de sodium 0,05 M pH 5,2, un tampon acide acétique 0,05 M pH 4,2 et un tampon acide acétique 0,05 M pH 4,2 contenant 1 M de NaCl. Les six fractions d'élution différentes obtenues pour chaque culot de précipitation ont, dans un premier temps, été analysées par zymographie caséines afin d'identifier les fractions contenant les protéases d'intérêt. Le zymogramme obtenu a ainsi permis de détecter les protéases A et B dans les fractions éluant à pH 7 et 6 des culots de précipitation à 60 et 80 % de sulfate d'ammonium et les protéases C et D dans les fractions éluant à pH 4,3 en présence de NaCl des culots de précipitation à 20 et 40 % de sulfate d'ammonium. Ces résultats montraient donc que les quatre protéases recherchées ne présentaient pas, d'une part, les mêmes propriétés hydrophobes / hydrophiles de surface et, d'autre part, les mêmes points isoélectriques : un point isoélectrique d'une valeur située entre 7 et 8,5 pouvait être suggéré pour les protéases A et B et un point isoélectrique d'une valeur inférieure à 4,3 pouvait être suggéré pour les protéases C et D. Dans un deuxième temps, le contenu protéique des fractions d'élution, dans lesquelles avaient pu être détectées les protéases, a été analysé par électrophorèse SDS-PAGE monodimensionnelle. Bien que les mêmes quantités de protéines que celles utilisées pour l'analyse par zymographie aient été employées, une coloration au nitrate d'argent a du être utilisée pour pouvoir visualiser des bandes protéiques sur le gel d'électrophorèse. Sur les profils électrophorétiques des fractions éluées à pH 7 et 6 et issus des précipités obtenus avec 60 et 80 % de sulfate d'ammonium, aucune bande correspondant à des protéines de masse moléculaire apparente proche de 90 kDa n'a pu être observée (figure 28A). L'absence de cette bande protéique, alors que ces fractions contenaient la protéase A, peut être expliquée par une trop faible quantité de protéase A présente. En effet, la sensibilité de la coloration au nitrate d'argent (de l'ordre du nanogramme) est plus faible que celle d'un zymogramme (de l'ordre du picogramme). Sur ces mêmes profils, une bande protéique de masse moléculaire apparente proche de 50 kDa et donc susceptible de correspondre à la protéase B, a été observée (figure 28A). Cependant, cette bande était également présente, avec la même intensité, sur le profil électrophorétique des fractions éluées à pH 7 et 6 et issus des précipités obtenus avec 20 et 40 % de sulfate d'ammonium (résultat non montré). Sachant que dans ces fractions, aucune activité caséinolytique n'avait pu être détectée par zymographie, il était donc peu probable que la bande en

question corresponde effectivement à la protéase B. Enfin, sur les profils électrophorétiques des fractions éluées à pH 4,3 en présence de NaCl et issus des précipités obtenus avec 20 et 40 % de sulfate d'ammonium, plusieurs bandes protéiques de masse moléculaire apparente proche de 20 kDa ont pu être visualisées (figure 28B), suggérant l'existence de plusieurs candidats encore pour les protéases C et D.

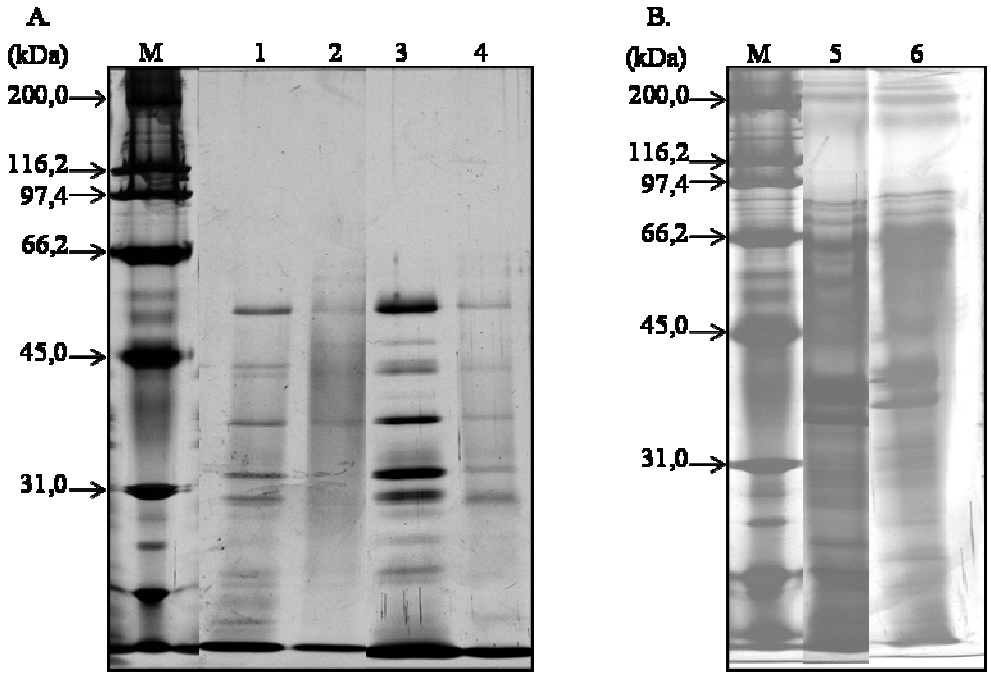


Figure 28 : Analyse par électrophorèse SDS-PAGE monodimensionnelle du milieu extracellulaire d'*E. coli* P4 (cultivée 13 heures à 37°C sous agitation en milieu synthétique) précipité au sulfate d'ammonium et fractionné par chromatographie échangeuse d'anions Deae23.

M : marqueurs de masse moléculaire.

1 et 2 : fractions chromatographiques éluées à pH 7 et 6, respectivement, du milieu extracellulaire précipité avec 60 % de sulfate d'ammonium.

3 et 4 : fractions chromatographiques éluées à pH 7 et 6, respectivement, du milieu extracellulaire précipité avec 80 % de sulfate d'ammonium.

5 et 6 : fractions chromatographiques éluées à pH 4,3 en présence de NaCl, du milieu extracellulaire précipité avec 20 et 40 % de sulfate d'ammonium, respectivement.

Des quantités de 20 µg de protéines totales ont été déposées par puits.

Ainsi, l'ensemble de ces résultats suggèrait que l'utilisation de 18 litres de culture n'était pas suffisante pour pouvoir obtenir des quantités suffisantes des quatre protéases et donc pour pouvoir les purifier aisément. En effet, bien que détectables, ces dernières semblaient présentes et donc excrétées, vraisemblablement en très faible quantité.

3.4. Publication

Influence of extracellular *Escherichia coli* proteases in caseinolysis of milk during a bovine mastitis, Dufour. Secretion and activity of extracellular caseinolytic proteases by the mammopathogen *E. coli* P4 strain were studied in order to support or not a direct bacterial role in the great caseinolysis of milk which occurs during bovine mastitis induced by *E. coli*.

SHORT COMMUNICATION: Can the mammopathogen *Escherichia coli* P4 strain have a direct role on the caseinolysis of milk observed during a bovine mastitis?

D. Dufour, N. Jameh, A. Dary[1], and Y. Le Roux

URAFPA, Nancy-Université, INRA, Equipe Protéolyse-Biofonctionnalité des Protéines et des Peptides, Boulevard des Aiguillettes, BP239, 54506 Vandoeuvre-lès-Nancy, France.

[1] Corresponding author: Annie Dary. Tel.: +330383684267 ; Fax: +330383684274; email: Annie.Dary@scbiol.uhp-nancy.fr.

ABSTRACT

During a bacterial bovine mastitis, quality of milk is altered because of an important caseinolysis. Endogenous potential actors of milk responsible for this caseinolysis have been well studied contrary to the exogenous bacterial ones. The aim of this study was to evaluate the direct role in caseinolysis of a mammopathogen strain, *Escherichia coli* P4. A secretion of at least four extracellular bacterial caseinolytic enzymes was highlighted by zymography, in three different growth media, and at each bacterial growth state, suggesting that their expression was constitutive. Different experimental schemes realized to evaluate caseinolytic potential did not show any significant caseinolytic activity of *E. coli* P4 and of the four extracellular proteases detected, suggesting that the high caseinolysis which is observed during an *E. coli* bovine mastitis does essentially result from endogenous milk actors.

Key words: mastitis, caseinolysis, *E. coli* P4, extracellular bacterial protease

Escherichia coli is recognized as the most encountered bacterial species in case of bovine mastitis in France (Seegers et al., 1997). Usually responsible for short-lasting clinical mastitis (Blowey and Edmondson, 1995), the mammopathogen *E. coli* strains are commonly associated with

recurrent mastitis which are rather sub-clinical (Bradley and Green, 2001). Whatever the kind of *E. coli* mastitis caused (clinical or sub-clinical), a great proteolysis of milk proteins, especially of caseins, is always observed during the disease (Michelutti et al., 1999), leading to important economical and industrial problems for the dairy cattle industry, particularly in case of subclinical mastitis. Endogenous potential actors responsible for this caseinolysis are the system plasminogen-plasmin and the large panel of proteases produced by the somatic cells recruited during infection, mainly by the polymorphonuclear neutrophils. A study of Moussaoui et al. (2004) reported the respective parts of these two endogenous actors, but also strongly suggested the existence of exogenous potential actors of caseinolysis, i.e. bacterial ones. Indeed, comparing the two experimental models currently employed to study *E. coli* mastitis, LPS (lipopolysaccharide) and *E. coli* mastitis models, these authors observed that caseinolysis is more important during an experimental *E. coli* mastitis than during an experimental LPS mastitis, both quantitatively and qualitatively (Moussaoui et al., 2002, 2004). Such a caseinolysis increase does not seem to result from a more important increase of plasmin activity nor from a more massive influx of somatic cells in milk during *E. coli* challenge (somatic cell counts of 3.98×10^6 and 27×10^6 cells/mL were registered during *E. coli* and LPS challenges, respectively, versus less than 10^5 cells per milliliter in an healthy milk of a non inflamed udder). Two hypotheses implying an *E. coli* role in milk proteolysis can account for such results. The first implies an indirect role of *E. coli* consisting in secretion of activators of proteases of somatic cells and / or in secretion of factors inducing a maturation of somatic cells and hence of their secreted proteases. The second, evaluated in this work, implies a direct role of *E. coli* which would secrete extracellular caseinolytic enzymes.

The bacterial strain chosen for this study was *E. coli* P4, isolated from a case of acute bovine mastitis (Bramley, 1976), and used in previous works of our unit research (Moussaoui et al., 2002, 2004). This strain was incubated at 37°C in four different growth media buffered at pH 6.8, i.e. milk pH, with 0.2 M pH 6.8 MOPS (3-(N-morpholino)propanesulfonic acid) buffer to prevent their acidification which usually occurs during the bacterial growth *in vitro* but not in the udder. Two classical microbial growth media were employed: minimum synthetic medium (called MS), containing 7 g/L K_2HPO_4, 2 g/L KH_2PO_4, 0.2 g/L Mg_2SO_4, 1 g/L $(NH_4)_2SO_4$ and 4 g/L glycerol; and Luria Bertani medium (called LB), containing 10 g/L bacto-tryptone (Difco^{TM}), 5 g/L bacto-yeast extract (Difco^{TM}) and 10 g/L NaCl. Two fresh milk media were also aseptically collected, one from an infused quarter cow 8 hours post intrammamary infusion of 10 µg of *E. coli* LPS (LPS O26:B6, Sigma Chemical Co, St. Louis, MO) (called LPSFM for LPS Fresh Milk), and one from the same cow just before infusion (called NIFM for Non Infused Fresh Milk). These two milks presented pH values of 6.98 and 6.77 and somatic cell counts (SCC) of 3.56×10^6 and 21×10^3 cells/mL (determined using a fluoro-opto electronic method), respectively. Such pH and SCC values corresponded to typical values

of inflamed and healthy milks, respectively. Both milks were skimmed by centrifugation at 4500 *g* for 25 min at 4°C, and absence of bacterial contamination was checked by dilution plating on LB solid medium containing bacto agar (15 g/L). As shown in figure 1A, bacterial growths in LB, LPSFM, and NIFM were similar, with a longer exponential growth phase and a higher bacterial count reached in LB. Bacterial growth in MS was slower, probably because this medium was poor in nutrient compared to the three other media. As *E. coli* P4 belongs to the phylogenetic group A of the *E. coli* species and as its core genome is very similar to the one of the commensal non pathogen *E. coli* K12 MG1655 strain (Dufour et al., in preparation), we determined whether the ability of *E. coli* P4 to grow in LPSFM as well as in NIFM was shared by *E. coli* K12 MG1655. Growth curves obtained with this strain are presented in figure 1B. In NIFM, growth of *E. coli* K12 MG1655 was similar to the one of *E. coli* P4, whereas in LPSFM, no growth of the bacterium could be observed. So, it appeared that these strains which belong to the same phylogenetic group did not present the same ability to grow in conditions miming inflammation, i.e. in antibacterial medium. *E. coli* P4 strain, contrary to *E. coli* K12 MG1655 strain, seems to be well adapted to grow in mastitic environmental conditions: its growth curves in LPSFM and NIFM are very similar. This ability of *E. coli* P4 is not always observed for all other mammopathogen *E. coli* strains. Indeed, several studies showed that some of them present a longer latent phase in LPSFM than in NIFM, sometimes with an initial decrease in numbers, and a lower bacterial count reached in LPSFM compared to in NIFM (Lohius et al., 1988, 1990).

Extracellular caseinolytic proteases secretion by *E. coli* P4 was examined in the four growth media at different times of incubation. Bacterial cultures were stopped by centrifugation at 12000 *g* for 15 min at 4°C, after 2, 4, 7 and 10 h of incubation in LPSFM, NIFM, and LB, and after 3.5, 7.5, 10.5, and 13 h of incubation in MS. These incubation times corresponded to four different bacterial growth states: latent state, exponential state, beginning of stationary state and stationary state (figure 1A). The supernatants with putative extracellular bacterial proteins were collected, filtered through a 0.45 µm filter, dialysed using a membrane with a 6-8 kDa molecular mass cutoff (Spectrum, Rancho Dominguez, US) for 48 h with NaN_3 0.01 % (w/v), and lyophilized (Hetosicc apparatus, Heto Birkeröd, Denmark). Before any analysis, absence of intracellular component, resulting from a total or partial cell lysis, was checked by testing the presence of nucleic acids in the extracellular contents, as described previously (Moe et al., 1994). Putative extracellular bacterial caseinolytic proteases were then searched by a zymography approach using caseins as substrat as described by Haddadi et al. (2006). This SDS-PAGE zymography was performed with 5 % polyacrylamide for the stacking gel in 0.125 M pH 6.8 Tris buffer and with 10 % polyacrylamide for the separating gel in 0.38 M pH 8.8 Tris buffer containing 0.26 mg/mL of sodium caseinate salt. About 20 µg extracellular bacterial proteins were loaded in the gel. For NIFM, MS, and LB media and at each growth time tested, four casein hydrolysis bands corresponding to at least four proteases of apparent molecular weight of 96 (called protease A), 54 (called protease B), 23 (called protease C), and 21 kDa (called protease D)

were detected specifically in the extracellular samples (figure 2, data only shown for MS medium), suggesting the bacterial secretion of at least four extracellular proteins with caseinolytic activity. Furthermore, the fact that these proteases were detected in the different bacterial growth media (NIFM, MS, and LB), and at each growth state (latent, exponential, and stationary) tested, indicated that their expression might be constitutive. Zymograms which were obtained with extracellular samples of *E. coli* P4 grown in LPSFM could not be interpreted because of the presence of several proteases of 90, 50, and 25 kDa which were present in the medium before its inoculation, as it has been previously reported (Haddadi et al., 2006). Hence, the distinction between endogenous proteases from milk and exogenous proteases from bacteria could not have been done.

In order to visualize the direct caseinolytic activity of the four detected extracellular proteases of *E. coli* P4, extracellular media of 10 and 13 h bacterial cultures realized in LB and MS respectively, were incubated for 24 and 48 h at 37°C with total casein solution (prepared in 0.05 M pH 6.8 sodium phosphate buffer at 1 mg/mL). Presence of the four proteases in these LB and MS extracellular samples was checked before by zymography (data not shown). Culture media realized with LPSFM and NIFM were not used for this study because of their high initial content in caseins. The reaction mixtures "LB or MS extracellular medium plus caseins" were analysed by SDS-PAGE according to the method of Laemmli and Favre (1973), using 5 % polyacrylamide for the stacking gel in 0.125 M pH 6.8 Tris buffer and 15 % polyacrylamide for the separating gel in 0.38 M pH 8.8 Tris buffer, and loading about 20 µg proteins per well. Mixtures with equal quantities of extracellular bacterial proteins and of caseins were firstly incubated and analysed. No qualitative evolution of caseins was detected on the electrophoregram, suggesting an absence of hydrolysis (figure 3A). Mixtures with higher quantities of extracellular bacterial proteins (ratio 10:1, 20:1, 50:1 and 100:1) were then incubated and analysed, but none hydrolysis could have been observed (data not shown). The same *E. coli* P4 extracellular protein samples were incubated for 24 and 48 h at 37°C with each casein fraction, i.e. with αs_1-, αs_2-, and β+κ-caseins solutions (also prepared in 0.05 M pH 6.8 sodium phosphate buffer at 1 mg/mL) separately, with equal quantities of extracellular bacterial proteins and of caseins, and analysed as above. Electrophoregram obtained showed no qualitative evolution in each casein fraction, suggesting again an absence of hydrolysis (figure 3B, 3C, 3D).

Two kinds of study were finally undertaken to evaluate the global caseinolytic potential of *E. coli* P4. In a first time, the global caseinolytic activity of *E. coli* P4 in milk, i.e. in a growth medium that it encounters when it contaminates the udder and induces an inflammation, was analysed. LPSFM was not used for such study because of its too great complexity (important endogenous proteolytic activity, and advanced hydrolysis of milk proteins). Hence, *E. coli* P4 was incubated in NIFM previously described. This NIFM presented a low somatic cell count (21×10^3 cells/mL), i.e. nearly no endogenous proteolytic activity, and thus enabled to evaluate only the bacterial proteolytic activity.

Indeed, with this value of somatic cell count (i.e. $<10^5$ cells/mL), a milk is considered as healthy (Burvenich et al., 2004), and contains mostly macrophages as somatic cells, which present a smaller proteolytic panel than polymorphonuclear neutrophils (Boutet et al., 2006). At different incubation times, *E. coli* P4 growth was determined and found to be similar to that presented in figure 1A. Evolutions of quantity of caseins and of the type of proteose-peptones were determined at five incubation times (2, 4, 7, 10 and 24 h), corresponding to four bacterial growth states (latent, exponential, beginning of stationary and two stationary states, respectively) (figure 1A, where the point 24 h is not shown, but is identical to the point 10 h). Casein content of the bacterial culture was recovered after precipitating caseins by setting the pH to 4.6, and percentage of each casein fraction was determined by Fast Protein Liquid Chromatography (FPLC) analysis on a MonoQ HR 5/5 anion-exchange column (Pharmacia, Uppsala, Sweden), according to the procedure of Collin et al. (1991). Proteose-peptone content of bacterial culture was recovered according to the method of Le Roux et al. (1995), and qualitative evolution of proteose-peptones was analysed by SDS-PAGE according to the method of Laemmli and Favre (1973), using 5 % polyacrylamide for the stacking gel in 0.125 M pH 6.8 Tris buffer and 15 % polyacrylamide for the separating gel in 0.38 M pH 8.8 Tris buffer, and loading about 60 µg proteose-peptones per well. At each of the incubation times tested, the presence of the four extracellular caseinolytic enzymes was checked by zymography. A control, corresponding to the same healthy milk without bacteria, was also analysed. No significant variation of the quantity of each casein fraction was observed between the different incubation times and between the inoculated milk and the control (figure 4). Likewise, no different type of proteose-peptones was detected between the different incubation times and between the two milks (figure 5). Thus, no obvious caseinolytic activity of the bacteria could be detected.

Secondly, the global caseinolytic activity of *E. coli* P4 was evaluated by incubating the bacteria in a growth medium with only caseins as nutrient sources. This medium corresponded to MS devoid of glycerol (carbon source) and $(NH_4)_2SO_4$ (nitrogen source), but supplied with a sodium caseinate (Sigma Chemical Co, St. Louis, MO) solution at concentration similar to milk one, i.e. 0.15 % (w/v), and initially prepared in a 0.05 M pH 7 sodium phosphate buffer. In these conditions, the bacteria had to develop, for their growth, a proteolytic activity to use caseins as nitrogen and carbon sources. After more than 200 h of incubation at 37°C in this medium, bacteria were still in a latent growth state, staying at an initial concentration of about 9.10^6 cfu/mL (figure 6). No increase of the bacterial concentration could be observed like those observed in LPSFM, NIFM, MS, and LB (figure 1), where, with the same initial inoculum, the bacterial concentrations reached about 10^9 cfu/mL. The four extracellular bacterial caseinolytic proteases were detected by zymography in the culture medium from 5 h to 200 h of incubation at 37°C. Despite, qualitative evolution of casein content of this medium, analysed by SDS-PAGE, did never change (figure 6). SDS-PAGE was performed according to the method of Laemmli and Favre (1973), using 5 % polyacrylamide for the stacking gel in 0.125

M pH 6.8 Tris buffer and 15 % polyacrylamide for the separating gel in 0.38 M pH 8.8 Tris buffer, and loading about 20 µg proteins in the gel. Thus, extracellular proteases detected did not enable the bacteria to use caseins as nutrient sources. These results suggest that caseins do not seem to be the specific substrate of the four extracellular enzymes secreted by *E. coli* P4. The no specific caseinolytic activity may be revealed by the zymography approach which is considered as a very sensitive method for protease detection and by the use of caseins as unique substrate. Caseins are effectively used as standard substrate (like insulin, gelatin, globin) for protease detection. Many intracellular proteases of *E. coli* have been identified *in vitro* using caseins as substrate, but they have another function *in vivo* (Sreedhara Swamy and Goldberg, 1982). It should be interesting to elucidate the exact function of the four extracellular enzymes secreted by *E. coli* P4 because it might result in the discovery of some virulence factors of this mammopathogen strain. Indeed, several extracellular secreted proteolytic enzymes with virulence properties have been identified in different enteropathogen and uropathogen *E. coli* strains (Dutta et al., 2002).

Though the results obtained in this study showed an effective extracellular secretion by *E. coli* P4 of four enzymes able to degrade caseins by zymography, the different experimental schemes realized to evaluate the intensity of caseinolytic potential did not support a significant caseinolytic activity of the bacteria or of the four enzymes detected. Thus, if these four proteases are not more expressed or if they do not present a higher proteolytic activity on caseins (notably on oxidized caseins) or if *E. coli* P4 does not express other proteases during a mastitis event, this work strongly suggest an absence of a direct role of *E. coli* P4 in the caseinolysis observed during a bovine mastitis. In perspective, the hypothesis of an indirect role of *E. coli* should be studied. Interests should focuse on the maturity stages of somatic cells present in milk during LPS and *E. coli* mastitis models. Indeed, such stages should be certainly different, and so be associated with different proteolytic panels. If this hypothesis was confirmed, it would be important for the dairy industry to reconsider the technique which implies only detection and count of somatic cells in milk to evaluate its protein degradation. Techniques which imply determination of somatic cell maturation should be rather consider.

References

Blowey, R. W., and P. W. Edmondson. 1995. Mastitis control in dairy herds. Ipswich, Farming Press, pp.29.

Boutet, P., Bureau, F., and Lekeux, P. 2006. La mammite bovine : de l'initiation à la résolution. Ann. Méd. Vet. 150:1-26.

Bradley, A. J., and M. J. Green. 2001. Adaptation of *Escherichia coli* to the bovine mammary gland. J. Clin. Microbiol. 39:1845-1849.

Bramley, A. J. 1976. Variations in the susceptibility of lactating and non-lactating bovine udders to infection when infused with *Escherichia coli*. J. Dairy Res. 43:205-211.

Burvenich, C., Monfardini, E., Mehrzad, J., Capuco, A.V., and Paape, M.J. 2004. Role of neutrophil polymorphonuclear leukocytes during bovine coliform mastitis: physiology or pathology ? Verh. K. Acad. Geneeskd. Belg. 66: 97-150.

Collin, J. C., A. Kokelaar, O. Rollet-Repecaud, and A. Delacroix-Buchet. 1991. Dosage des caséines du lait de vache par électrophorèse et par chromatographie liquide rapide d'échange d'ions (FPLC) : comparaison des résultats. Lait 71:339-350.

Dutta, P. R., R.Capello, F. Navarro-Garcia, and J. P. Nataro. 2002. Functional comparison of serine protease autotransporters of enterobacteriaceae. Infect. Immun. 70:7105-7113.

Haddadi, K., C. Prin-Mathieu, F. Moussaoui, G. C. Faure, F. Vangroenweghe, C. Burvenich, and Y. Le Roux. 2006. Polymorphonuclear neutrophils and *Escherichia coli* proteases involved in proteolysis of casein during experimental *E. coli* mastitis. Int. Dairy J. 16:639-647.

Laemmli, U. K., and M. Favre. 1973. Maturation of the head of bacteriophage T4. I. DNA packaging events. J. Mol. Biol. 80:575-599.

Le Roux, Y., O. Colin, and F. Laurent. 1995. Proteolysis in samples of quarter milk with varying somatic cell counts. 1. Comparison of some indicators of endogenous proteolysis in milk. J. Dairy Sci. 78:1289-1297.

Lohuis, J. A., W. Kremer, Y. H. Schukken, J. A. Smit, J. H. M. Verheijden, A. Brand, and A. S. Van Miert. 1990. Growth of *Escherichia coli* in milk from endotoxin-induced mastitic quarters and the course of subsequent experimental *Escherichia coli* mastitis in the cow. J. Dairy Sci. 73:1508-1514.

Lohuis, J. A., W. Van Leeuwen, J. H. Verheijden, J. A. Smit, A. Brand, and A. S. Van Miert. 1988. Growth of *Escherichia coli* in whole and skim milk from endotoxin-induced mastitic quarters: in vitro effects of deferoxamine, zinc, and iron supplementation. J. Dairy Sci. 71:2772-2781.

Michelutti, I., Y. Le Roux, P. Rainard, B. Poutrel, and F. Laurent. 1999. Sequential changes in milk protein composition after experimental *Escherichia coli* mastitis. Lait 79:535-549.

Moe, D., C. Garbarsch, and S. Kirkeby. 1994. The protein effect on determination of DNA with Hoechst 33258. J. Biochem. Biophys. Methods 28:263-276.

Moussaoui, F., F. Michelutti, Y. Le Roux, and F. Laurent. 2002. Mechanisms involved in milk endogenous proteolysis induced by a lipopolysaccharide experimental mastitis. J. Dairy Sci. 85:2562-2570.

Moussaoui, F., F. Vangroenweghe, K. Haddadi, Y. Le Roux, F. Laurent, L. Duchateau, and Burvenich C. 2004. Proteolysis in milk during experimental *Escherichia coli* mastitis. J. Dairy Sci. 87:2923-2931.

Seegers, H., J. L. Menard, and C. Fournichon. 1997. Mammites en élevage bovin laitier : importance actuelle, épidémiologie et plans de prévention. Renc. Rech. Ruminants 4:233-242.

Sreedhara Swamy, K. H., and A. L. Goldberg. 1982. Subcellular distribution of various proteases in *Escherichia coli.* J. Bacteriol. 149:1027-1033.

Figure 1.

A. Growth curves of *Escherichia coli* P4 in LPSFM (Δ), NIFM (▲), MS (●), and LB (■). Arrows indicate times at which extracellular bacterial protein samples from milks and LB (full line), and MS (dotted line) were recovered for zymogram analyses.

B. Growth curves of *E. coli* K12 MG1655 in LPSFM (Δ), and NIFM (▲).

Figure 2. SDS-PAGE zymogram analysis of *Escherichia coli* P4 extracellular content culture from MS. Clear zones indicate protease activities on Coomassie blue R250 stained gel. M: molecular mass standards.

Figure 3. SDS-PAGE electrophoregrams of total caseins (**A**), α_{s1}-casein (**B**), α_{s2}-casein (**C**) and β+κ-caseins (**D**) alone (Ctr) or mixed with an equal quantity of *Escherichia coli* P4 extracellular proteins from 13 h MS (EP-MS) and 10 h LB (EP-LB) cultures, and incubated for 0, 24 and 48 h at 37°C. Gel was stained with Coomassie blue R250.

Figure 4. Evolution of each casein fraction, determined by Fast Protein Liquid Chromatography (FPLC) analysis, from non infused fresh milk (NIFM) inoculated with *Escherichia coli* P4 (white background histogram) or not (grey background histogram), and incubated for 2, 7, and 24 h at 37°C.

Figure 5. SDS-PAGE electrophoregram of proteose-peptone fraction from non infused fresh milk (NIFM) inoculated with *Escherichia coli* P4 (P) or not (C), and incubated for 2, 7, and 24 h at 37°C. Gel was stained with Coomassie blue R250. M: molecular mass standards.

Figure 6. Growth curve of *Escherichia coli* P4 in MS devoid of $(NH_4)_2SO_4$ and glycerol, but supplied with 0.15 % (w/v) sodium caseinate (initially prepared in a 0.05 M pH 7 sodium phosphate buffer); and SDS-PAGE electrophoregram of caseins in culture medium after 0 (1), and 200 h (2) of incubation at 37°C. Gel was stained with Coomassie blue R250.

Figure 1

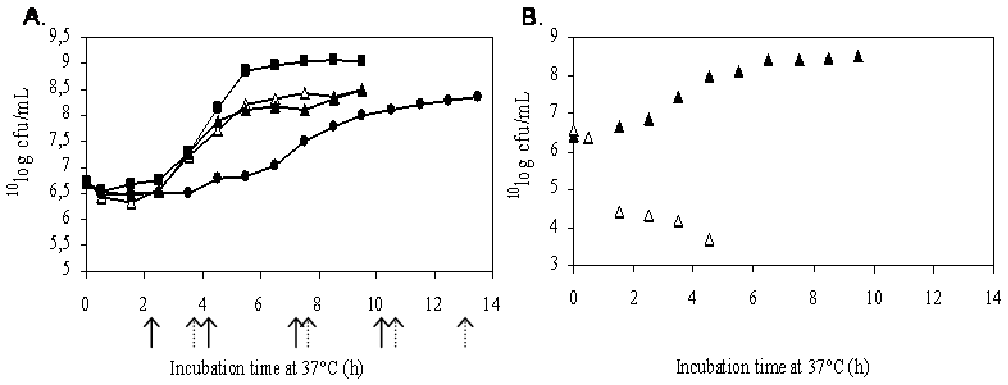


Figure 2

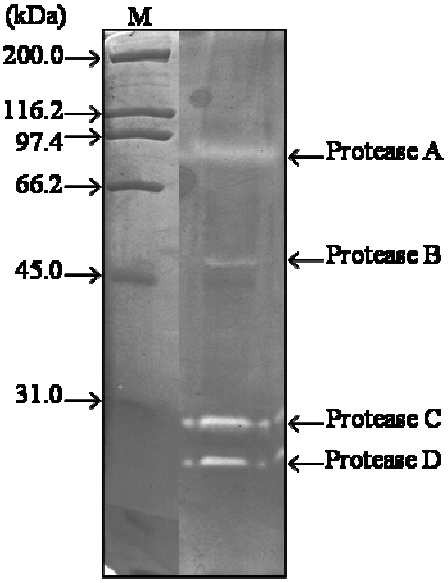


Figure 3

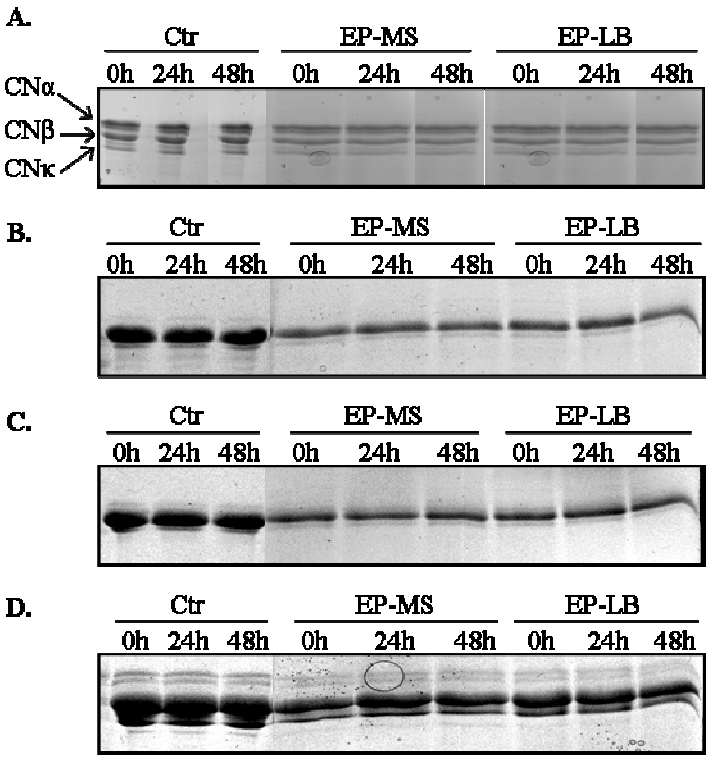


Figure 4

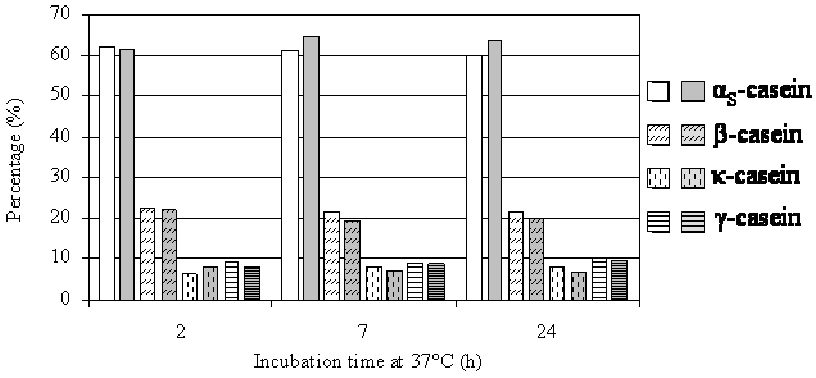


Figure 5

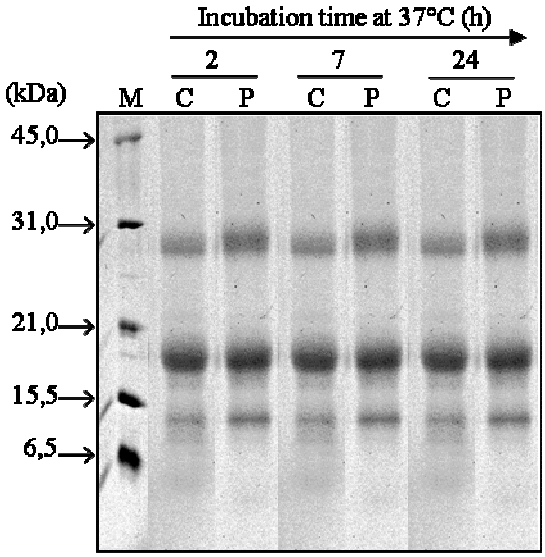
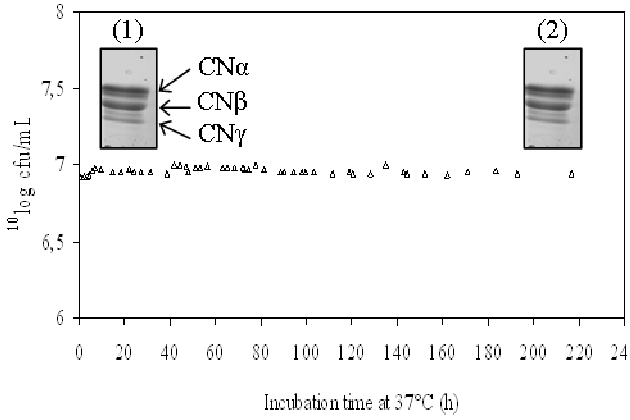


Figure 6



4. Recherche de séquences nucléotidiques particulières à *E. coli* P4

Au sein d'une espèce bactérienne, il est bien établi désormais que les propriétés spécifiques d'une souche sont portées non pas par son génome core mais par les parties variables de son génome. La seconde étude moléculaire entreprise a consisté à analyser non plus le génome core d'*E. coli* P4, mais les éventuelles parties variables de son génome. Le matériel et les méthodes expérimentales utilisés ainsi que les résultats obtenus pour cette partie du travail sont exposés dans l'affiche et la publication insérées à la fin de ce paragraphe.

4.1. Recherche de facteurs de virulence

Afin d'essayer d'expliquer la pathogénicité d'*E. coli* P4, nous avons pensé tout d'abord à rechercher dans son génome la présence de gènes codant différents facteurs de virulence, en particulier les plus connus et les plus fréquents chez les souches *E. coli* pathogènes pour le bovin et le veau, dont certains ont notamment été mis en évidence chez des souches MPEC. Vingt six gènes de virulence, dont les caractéristiques sont présentées dans le tableau 17, ont ainsi été recherchés dans le génome d'*E. coli* P4 par PCR en utilisant des couples d'oligonucléotides spécifiques décrits dans plusieurs travaux antérieurs. Malgré l'utilisation de différents gradients de températures d'hybridation pour les réactions de PCR, sur les 26 couples d'oligonucléotides testés un seul a permis d'obtenir un produit d'amplification de taille attendue, le couple spécifique du gène *traT* codant un facteur de résistance au sérum. Ces résultats sont en accord avec les travaux publiés dans la littérature qui montrent que, jusqu'à présent, la plupart des facteurs de virulence connus chez les autres pathotypes de l'espèce *E. coli* sont très rarement retrouvés par la technique de PCR au sein des génomes des souches MPEC (tableau 15) et que seul le facteur de virulence TraT semble être identifié de manière fréquente (Kaipainen *et al*., 2002).

Tableau 17 : Caractéristiques des 26 gènes de virulence recherchés par PCR au sein du génome d'*E. coli* P4.

Gène	Localisation génomique	Protéine et fonction codées	Etude de référence	Souches détentrices
traT	Plasmidique	TraT, facteur de résistance au sérum	Kaipainen *et al.*, 2002	MPEC
cnf2	Plasmidique	CNF2, toxine		
afa7-D	Chromosomique ou plasmidique	Afa7-D, invasine des adhésines de type Afa7	Lalioui *et al.*, 1999	*E. coli* inductrices de diarrhées et de septicémie chez le veau
afa7-E		Afa7-E, adhésine des adhésines de type Afa7		
afa8-D		Afa8-D, invasine des adhésines de type Afa8		
afa8-E		Afa8-E, adhésine des adhésines de type Afa8		
clpG	Plasmidique	CS31A, adhésine de type afimbriae	Bertin *et al.*, 1998	*E. coli* inductrices de diarrhées et de septicémie chez le bovin
eaeA	Chromosomique	EaeA, intimine	China *et al.*, 1996	EPEC chez le bovin
sltI/II	Chromosomique	Slt-I/II, toxines		
eltIA/IIA	Plasmidique	LT-IA/IIA, toxines	Kuhnert *et al.*, 1997	-
stIA/IB	Chromosomique	St-IA/IB, toxines		
neuC-neuA	Chromosomique	NeuC et NeuA, acide polysialique synthétase et acylneuraminate cytidylyltransferase, respectivement, enzymes nécessaires à la synthèse de capsule		
f17a-A	Chromosomique	F17a-A, sous-unité majeure des adhésines de type F17a	Bertin *et al.*, 1996	ETEC chez le bovin
f17b-A		F17b-A, sous-unité majeure des adhésines de type F17b		*E. coli* inductrices de diarrhées et de septicémie chez le veau
f17c-A		F17c-A, sous-unité majeure des adhésines de type F17c		*E. coli* inductrices de septicémie
f17d-A		F17d-A, sous-unité majeure des adhésines de type F17d		ETEC chez le bovin et *E. coli* inductrices de diarrhées chez le veau
cnf1	Chromosomique	CNF1, toxine	Yamamoto *et al.*, 1995	ExPEC
iucD	Chromosomique ou plasmidique	N-lysine monooxygenase, enzyme intervenant dans la synthèse d'aérobactine		Bactéries à Gram négatif
papE-papF	Chromosomique	PapE et PapF, sous-unités mineures des adhésines de type P		UPEC
sfaD-sfaE		SfaD et SfaE, sous-unité mineure et protéine chaperone, respectivement, des adhésines de type S		

4.2. Recherche d'îlots génomiques

Plusieurs études ont prouvé sur la base d'examens phénotypiques la virulence des MPEC, comme par exemple leur capacité d'adhérence et d'invasion des cellules épithéliales mammaires bovines (Döpfer *et* al., 2000). Ainsi, il est raisonnable de penser que les échecs de mise en évidence ne sont pas dus à une absence de facteurs de virulence, mais plutôt à la présence de facteurs différents encore inconnus. Afin d'essayer de mettre en évidence la présence d'éventuels facteurs de virulence encore inconnus chez *E. coli* P4, nous avons recherché la présence d'éléments génétiques mobiles ou mobilisables potentiels dans son génome. Ces derniers sont effectivement les vecteurs des propriétés nouvelles et spécifiques d'une souche au sein d'une espèce bactérienne. Parmi les différents types d'éléments génétiques mobiles ou mobilisables recensés dans le monde bactérien, nous nous sommes focalisés sur la recherche d'îlots génomiques au sein du génome d'*E. coli* P4. Ce choix peut notamment s'expliquer par le fait qu'il existe pour la détection des îlots génomiques une approche expérimentale facile et rapide à mettre en œuvre, basée sur la spécificité d'intégration de ces éléments dans les chromosomes bactériens.

4.2.1. Criblage des loci d'ARNt

Dans plus de 75 % des cas, du fait de la spécificité de leur intégrase, les îlots génomiques sont insérés en 3' des loci d'ARNt. Connaissant ainsi les séquences nucléotidiques des régions constantes amont et aval de ces loci au sein d'une espèce bactérienne, il est possible, par le biais de la PCR et de l'utilisation de couples d'oligonucléotides spécifiques à ces régions, de mettre en évidence des structures intactes ou altérées de ces loci (figure 29). Une structure altérée peut ensuite être effectivement expliquée par l'insertion d'un éventuel ADN étranger à proximité du locus d'ARNt, tel qu'un îlot génomique. Au sein de l'espèce *E.* coli, tous les loci d'ARNt ne semblent pas propices à l'insertion d'ADN étranger (Germon *et al*., 2007). Il est en effet considéré que sur les 85 loci d'ARNt communs aux souches *E. coli* K12 MG1655, *E. coli* EDL933 et *E. coli* CFT073 (représentants majeurs des souches non pathogènes, pathogènes intestinales et pathogènes extra intestinales, respectivement), seuls 36 peuvent être sélectionnés pour accueillir un ADN étranger, car n'étant pas situés à l'intérieur d'opérons ribosomiaux ou au sein de structures polycistroniques. De plus, il a été montré que ces 36 loci ne présentent pas tous la même probabilité d'accueil : 19 d'entre eux sont associés à une très faible probabilité et 17 à une plus forte probabilité (tableau 18).

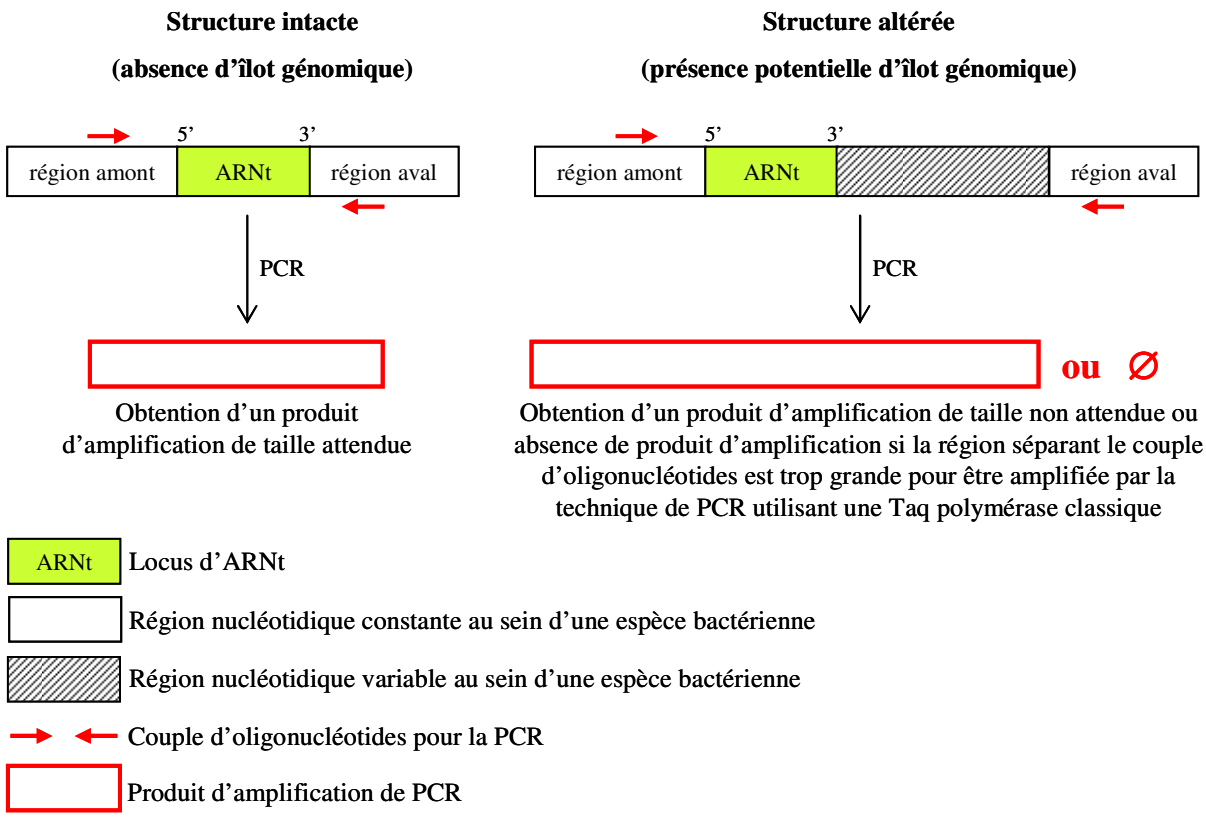


Figure 29 : Approche expérimentale permettant la détection de structures intactes ou altérées des loci d'ARNt bactériens.

Tableau 18 : Loci d'ARNt propices à l'insertion d'ADNs étrangers au sein de l'espèce *E. coli*, d'après Germon *et al.*, 2007. * : classement des loci par ordre de probabilité décroissante d'insertion.

Loci d'ARNt présentant une faible probabilité d'insertion *	**Loci d'ARNt présentant une forte probabilité d'insertion ***
leuZ	*aspV*
proL	*pheV*
serT	*leuX*
leuV	*thrW*
tyrV	*metV*
ileX	*asnT*
leuU	*argU*
glnX	*argW*
alaX	*glyU*
argQ	*asnU*
aspU	*asnV*
glyY	*asnW*
lysQ	*pheU*
lysV	*serU*
metY	*serX*
proK	*selC*
proM	*serW*
thrT	
valW	

Grâce à différents couples d'oligonucléotides décrits dans la littérature (Dezfulian *et al*., 2003 ; Germon *et al*., 2007) ou dessinés par nous-même (dont les séquences sont présentées dans le tableau 1 de la publication insérée à la fin de ce paragraphe), les structures de ces 17 loci au sein du génome d'*E. coli* P4 ont été examinées et comparées à celles présentes au sein du génome d'*E. coli* K12 MG1655. Pour quatre d'entre eux, les loci *asnT*, *aspV*, *pheV* et *serU*, un contexte nucléotidique différent de celui présent chez *E. coli* K12 MG1655 a pu être directement mis en évidence. Pour les loci *asnT*, *aspV* et *serU*, aucun produit d'amplification n'a pu être obtenu avec les couples d'oligonucléotides testés, alors qu'un produit d'amplification pouvait être obtenu chez *E. coli* K12 MG1655 et, pour le loci *pheV*, un produit d'amplification de taille non attendue (2208 pb chez *E. coli* P4, contre 623 pb chez *E. coli* K12 MG1655) a été obtenu. Ces résultats indiquaient donc l'existence, aux extrémités 3' de ces quatre loci d'ARNt du génome d'*E. coli* P4, de structures génomiques différentes de celles présentes dans le génome d'*E. coli* K12 MG1655 qui sont rappelées dans le tableau 19. Pour trois autres loci, les loci *argW*, *leuX* et *thrW*, aucun produit d'amplification n'a également pu être obtenu avec les couples d'oligonucléotides testés, tout comme dans le cas d'*E. coli* K12 MG1655. Ces résultats indiquaient donc soit, la présence de structures identiques à celles présentes dans le génome d'*E. coli* K12 MG1655 (tableau 19), soit, la présence de structures différentes.

Tableau 19 : Contextes génomiques de sept loci d'ARNt présents sur le génome de la souche *E. coli* K12 MG1655.

Locus	Insertion d'ADN étranger	Taille de l'ADN étranger inséré (kb)	Nature de l'ADN étranger inséré
argW	oui	12	KpLE1 (élément similaire à un prophage)
asnT	non	-	-
aspV	non	-	-
leuX	oui	40	KpLE2 (élément similaire à un prophage)
pheV	non	-	-
serU	non	-	-
thrW	oui	35	CP4-6 (prophage cryptique de type P4)

4.2.2. Caractérisation partielle des régions aval des loci *asnT*, *leuX*, *pheV* et *thrW* grâce à une technique de PCR arbitraire

Afin de déterminer les séquences nucléotidiques des régions situées immédiatement en 3' de quatre loci d'ARNt pour lesquels une structure altérée avait été mise en évidence, *asnT*, *leuX*, *pheV* et

thrW, une technique de PCR arbitraire a été utilisée (P. Germon, communication personnelle). Elle repose sur la réalisation de deux réactions de PCR successives (figure 30). La première utilise comme matrice, l'ADN total de la souche bactérienne étudiée et comme couple d'oligonucléotides, un oligonucléotide « forward » de séquence définie complémentaire à la région génomique d'intérêt et un oligonucléotide « reverse » de séquence partiellement définie, c'est-à-dire composée en 5' d'une suite de nucléotides correspondant à une séquence consensus fréquemment retrouvée au sein des génomes d'*E. coli* et en 3' d'une suite dégénérée de nucléotides. La deuxième réaction utilise comme matrice, le mélange réactionnel de PCR obtenu après la première réaction et comme couple d'oligonucléotides, un oligonucléotide « forward » de séquence définie également complémentaire à la région génomique d'intérêt mais à une position située légèrement en 3' de celle de l'oligonucléotide « forward » utilisé précédemment et un oligonucléotide « reverse » de séquence définie composée uniquement de la partie consensus de l'oligonucléotide « reverse » utilisé lors de la première réaction. Au terme de cette deuxième réaction de PCR, les produits d'amplification obtenus peuvent être visualisés après électrophorèse. Ceux de plus forte intensité, signe d'une bonne spécificité de la région amplifiée, peuvent être purifiés puis séquencés.

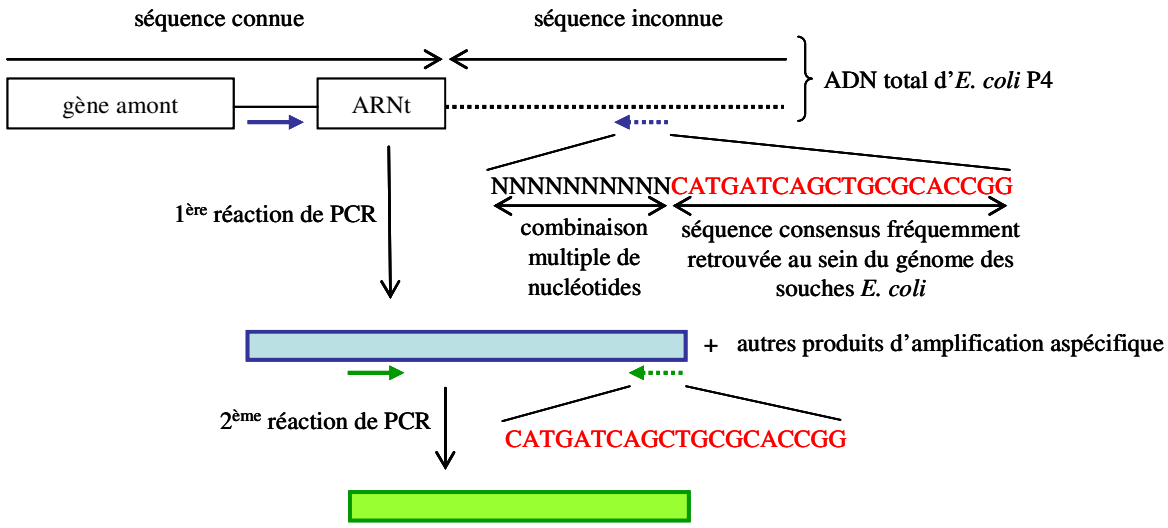


Figure 30 : Présentation de la technique de PCR arbitraire. →← : couple d'oligonucléotides utilisé pour la PCR.

La taille et le pourcentage en paires de bases guanine et cytosine des séquences nucléotidiques situées immédiatement en 3' des loci *asnT*, *leuX*, *pheV* et *thrW* sur le génome d'*E. coli* P4, amplifiées par PCR arbitraire puis séquencées sont présentés dans le tableau 20.

Tableau 20 : Caractéristiques des séquences nucléotidiques situées immédiatement en 3' des loci *asnT*, *leuX*, *pheV* et *thrW* sur le génome d'*E. coli* P4, amplifiées par PCR arbitraire et séquencées.

Locus	Taille (nucléotides)	GC (%)
asnT	202	40,1
leuX	1335	46,59
pheV	1677	46,57
thrW	696	47,56

Pour chacune de ces séquences, des pourcentages en paires de bases guanine et cytosine assez éloignés de celui du génome total d'*E. coli* K12 MG1655 (50,8 %) ont pu être calculés et des séquences partiellement identiques ont pu être retrouvées dans les banques de données grâce au programme blastn disponible sur le site NCBI. Sur certaines d'entre elles, des phases ouvertes de lecture (ORF) ainsi que des séquences de fixation des ribosomes (RBS) ont pu être détectées grâce au programme CDS Finder du site NCBI et aux programmes Glimmer et RBS Finder du site Nano + Bio Center. Aucun promoteur transcriptionnel en amont de ces ORFs n'a cependant pu être identifié en utilisant le programme BPROM du site SoftBerry, programme basé sur la recherche de séquences promotrices consensus déjà connues chez les procaryotes.

4.2.2.1. Cas du locus *asnT*

Les 202 nucléotides amplifiés et séquencés immédiatement en 3' du locus *asnT* sur le génome d'*E. coli* P4 présentent 96 % d'identité avec une région de 195 pb retrouvée sur le génome de la souche *E. coli* C ATCC 8739, décrite comme une souche commensale non pathogène, immédiatement en 3' de deux loci d'ARNt-Asn (correspondant probablement aux loci *asnT* et *asnV* du génome d'*E. coli* K12 MG1655). Aucune ORF complète n'a pu être détectée au sein de cette séquence d'*E. coli* P4, en raison de sa taille trop petite. Cependant, la séquence similaire retrouvée chez *E. coli* C ATCC 8739 contient les 36 premiers nucléotides d'un pseudogène d'une taille de 1263 nucléotides et présentant 82 % d'identité avec une ORF codant une intégrase putative de type P4 sur le génome de la souche *Sh. boydii* Sb 227.

4.2.2.2. Cas du locus *leuX*

Les 1335 nucléotides amplifiés et séquencés, localisés immédiatement en 3' du locus *leuX* sur le génome d'*E. coli* P4, présentent 75 % d'identité avec une région de 1369 pb retrouvée immédiatement en 3' du locus *leuX* sur le génome de la souche entéropathogène *E. coli* O157:H7 Sakaï. Une ORF, dont les caractéristiques sont présentées dans le tableau 21, a pu être détectée au sein de cette séquence. Cette ORF code une protéine similaire à la partie N-Terminale d'une intégrase

putative retrouvée immédiatement en 3' du locus *leuX* sur le génome d'*E. coli* O157:H7 Sakaï (figure 31). Une mutation créant un codon stop et amenant à la perte de la partie C-Terminale catalytique de l'enzyme semble s'être produite au sein de cette ORF, qui n'est, par conséquent, probablement plus fonctionnelle et peut plutôt être assimilée à un pseudogène.

Tableau 21 : Caractéristiques de l'ORF détectée immédiatement en 3' du locus *leuX* sur le génome d'*E. coli* P4.

ORF	Position	RBS séquence, position	Taille (acides aminés)	GC (%)	Masse moléculaire (kDa)	Recherche d'ORF similaire dans les banques de données
1-*leuX*	91-1065	AGGCG, 86-90 AGGGT, 76-80	324	48,1	36,51	69 % d'identité et 76 % de similarité avec les 336 1ers acides aminés d'une intégrase putative de 649 acides aminés identifiée chez *E. coli* O157:H7 Sakaï

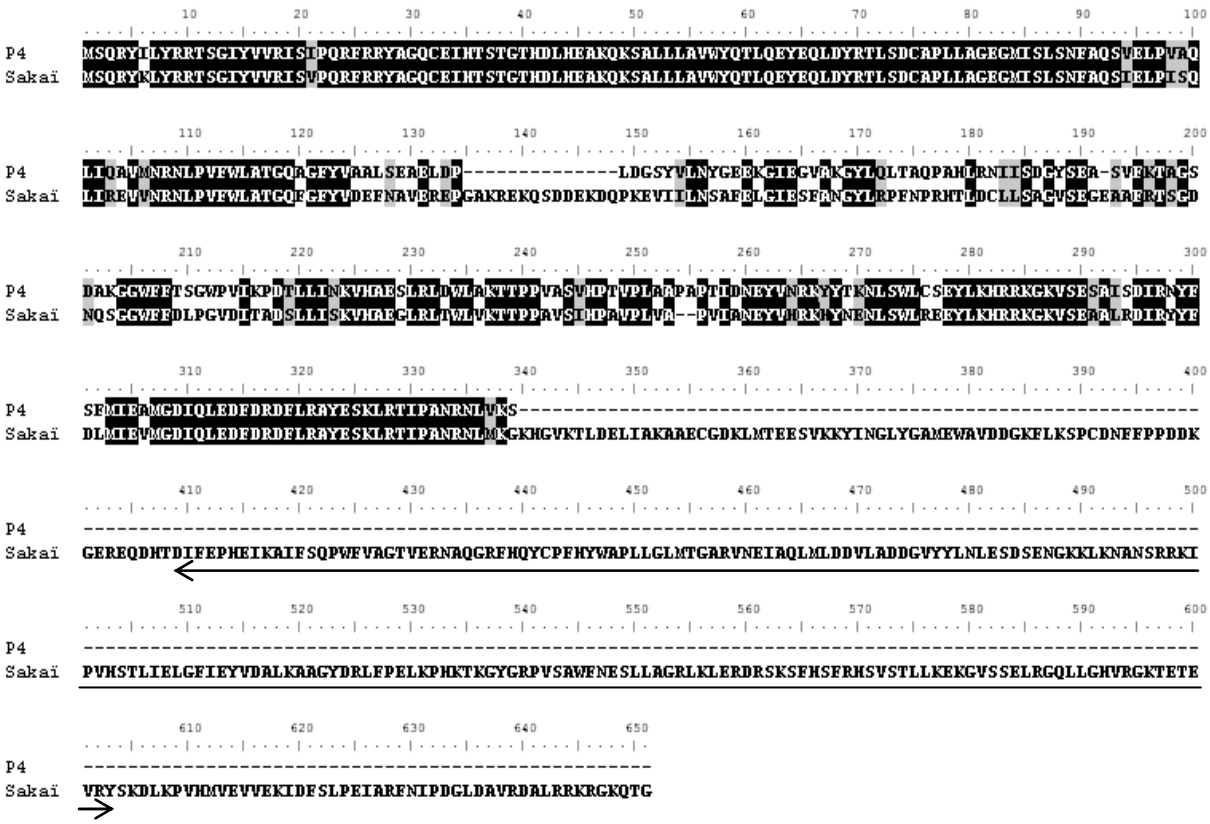


Figure 31: Alignement des structures primaires des protéines codées par les ORFs situées immédiatement en 3' du locus *leuX* sur les génomes d'*E. coli* P4 et d'*E. coli* O157:H7 Sakaï. Les acides aminés identiques entre les deux séquences sont inscrits sur fond noir, ceux qui sont similaires sur fond gris et ceux qui sont différents sur fond blanc. La double flèche indique le domaine C-Terminal catalytique de l'intégrase présente chez *E. coli* O157:H7 Sakaï.

Sur le génome d'*E. coli* O157:H7 Sakaï et en aval du locus *leuX*, se trouve une région appelée SpLE5, pour Sakaï prophage Like Element, ne contenant aucun gène d'origine phagique, excepté celui codant

l'intégrase situé en 1ère position. Etant donné les similitudes des régions 3' du locus *leuX* des génomes d'*E. coli* P4 et d'*E. coli* O157:H7 Sakaï, nous avons voulu savoir si l'intégralité de l'élément SpLE5 était présente également sur le génome d'*E. coli* P4. Pour cela, neuf couples d'oligonucléotides spécifiques de l'élément SpLE5 ont été dessinés d'après la séquence nucléotidique complète du génome d'*E. coli* O157:H7 Sakaï disponible dans les banques de données et testés par la technique de PCR sur l'ADN total d'*E. coli* P4. Deux témoins ont été utilisés, les ADNs totaux des souches *E. coli* K12 MG1655 (témoin négatif) et *E. coli* EDL933 (témoin positif, contenant également l'élément SpLE5). Aucun produit d'amplification de taille attendue n'a pu être obtenu avec les neuf couples à partir de l'ADN total d'*E. coli* P4 : soit les produits obtenus présentaient une taille inférieure à celle attendue (ce qui a été le cas avec sept des couples testés, spécifiques des extrémités 5' et 3' de l'élément SpLE5), soit aucun produit n'était obtenu (ce qui a été le cas pour deux des couples testés, spécifiques de la région centrale de l'élément SpLE5). Etant donné que les produits d'amplification obtenus n'ont pas été séquencés, aucune conclusion concernant la nature de la région présente en 3' du locus *leuX* sur le génome d'*E. coli* P4 ne peut être établie. Cependant, deux hypothèses peuvent être avancées : soit les produits d'amplification obtenus sont aspécifiques et cette région ne présente aucune similarité avec l'élément SpLE5 (excepté au niveau du premier gène codant l'intégrase), soit les produits d'amplification obtenus sont bien spécifiques et cette région n'est que très partiellement similaire à l'élément SpLE5.

4.2.2.3. Cas du locus *pheV*

Dans un premier temps, le produit d'amplification de 2208 pb obtenu lors du criblage du locus *pheV* a été séquencé. La séquence obtenue possède un pourcentage en paires de bases guanine et cytosine de 49,60 % et présente des identités supérieures à 80 % avec différents fragments d'un îlot de pathogénicité appelé LEE-PAI et identifié dans le génome de la souche *E. coli* RW1473, présentée comme pathogène pour le bovin et productrice de toxines Shiga. Les 333 derniers nucléotides de cette séquence présentent également 100 % d'identité avec les 333 premiers nucléotides du gène *yghD* d'*E. coli* K12 MG1655, premier gène constant retrouvé jusqu'à présent sur tous les génomes des souches *E. coli* en 3' du locus *pheV* (figure 32). Des nucléotides 1705 à 1727 de la séquence obtenue, une séquence de 23 pb identique à 100 % aux 23 derniers nucléotides du gène *pheV* est également retrouvée. Cette région de 2208 pb correspond donc vraisemblablement à l'extrémité 3' de l'îlot génomique présent en aval du locus *pheV* au sein du génome d'*E. coli* P4. Trois ORFs, dont les caractéristiques sont présentées dans le tableau 22 et la figure 32, ont pu être détectées au sein de la séquence obtenue. Ces trois ORFs codent des protéines similaires à des protéines hypothétiques de fonction inconnue identifiées sur l'îlot génomique GEI_{II} de la souche commensale *E. coli* Nissle 1917 (pour l'ORF 2-*pheV*), dans le génome de la souche entéropathogène *E. coli* B7A (pour l'ORF 3-*pheV*)

et sur l'îlot génomique LEE-PAI de la souche bovine *E. coli* RW1374 productrice de toxines shiga (pour l'ORF 4-*pheV*).

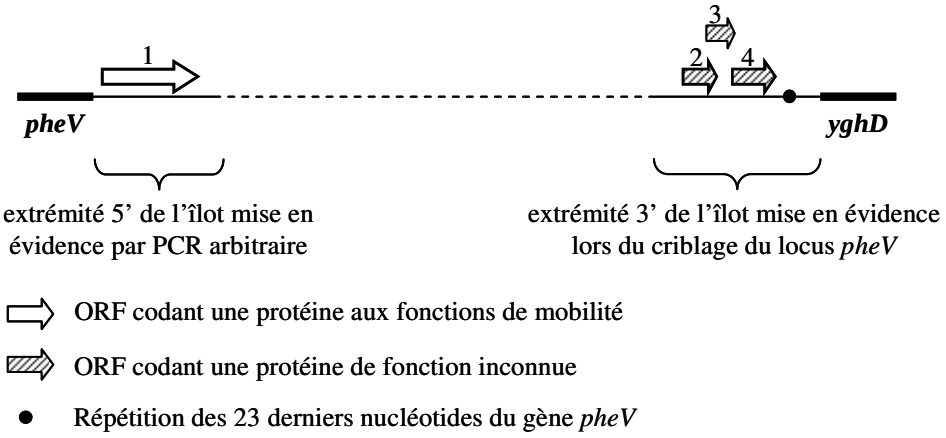


Figure 32 : Carte génétique partielle de l'îlot génomique situé en 3' du locus *pheV* sur le génome d'*E. coli* P4. Les chiffres indiquent les numéros des ORFs présentées dans les tableaux 22 et 23.

Tableau 22 : Caractéristiques des ORFs détectées dans la séquence nucléotidique du produit d'amplification de 2208 pb obtenu lors du criblage du locus *pheV* du génome d'*E. coli* P4.

ORF	Position	RBS séquence, position	Taille (acides aminés)	GC (%)	Masse moléculaire (kDa)	Recherche d'ORF similaire dans les banques de données
2-*pheV*	370-816	-	148	53,2	16,4	91 % d'identité et 94 % de similarité avec les 79 1ers acides aminés d'une protéine hypothétique de 143 acides aminés identifiée chez *E. coli* Nissle 1917
3-*pheV*	683-925	-	80	49,6	9,2	98 % d'identité et 100 % de similarité avec une protéine hypothétique de 80 acides aminés identifiée chez *E. coli* B7A
4-*pheV*	1022-1594	AGGAA, 1009-1013	190	45,8	21,82	98 % d'identité et 98 % de similarité avec une protéine hypothétique de 190 acides aminés identifiée chez *E. coli* RW1374

Les 1677 nucléotides amplifiés par PCR arbitraire et localisés immédiatement en 3' du locus *pheV* sur le génome d'*E. coli* P4, ont également été séquencés. La séquence obtenue présente 90 % d'identité avec une région de 1675 pb retrouvée sur le génome de la souche *Sh. flexneri* 2a strain 301 immédiatement en 3' d'un locus d'ARNt-Phe. Une ORF, dont les caractéristiques sont présentées dans le tableau 23 et la figure 32, a pu être détectée au sein de cette séquence. Cette ORF code une protéine similaire à une intégrase putative de type P4 retrouvée immédiatement en 3' d'un locus d'ARNt-Phe sur le génome de la souche *Sh. flexneri* 2a strain 301. Elle semble avoir ainsi gardé l'intégralité des motifs caractéristiques de l'intégrase du bactériophage P4 et est donc probablement toujours fonctionnelle.

Tableau 23 : Caractéristiques de l'ORF détectée immédiatement en 3' du locus *pheV* sur le génome d'*E. coli* P4.

ORF	Position	RBS séquence, position	Taille (acides aminés)	GC (%)	Masse moléculaire (kDa)	Recherche d'ORF similaire dans les banques de données
1-*pheV*	198-1460	-	420	47,7	47,99	90 % d'identité et 95 % de similarité avec une intégrase putative de type P4 de 420 acides aminés identifiée chez *Sh. flexneri* 2a 301

4.2.2.4. Cas du locus *thrW*

Les 696 nucléotides amplifiés, séquencés et localisés immédiatement en 3' du locus *thrW* sur le génome d'*E. coli* P4 présentent 98 % d'identité avec une région de 696 pb retrouvée 15640 nucléotides en 3' d'un locus d'ARNt-Phe sur le génome de la souche *Sh. flexneri* 2a strain 301 et immédiatement en 3' du locus *thrW* sur le génome de la souche *E. coli* O157:H7 Sakaï. Aucune ORF complète n'a pu être détectée chez *E. coli* P4, en raison de la petite taille de la séquence amplifiée. Cependant, les séquences similaires retrouvées chez les souches *Sh. flexneri* 2a strain 301 et *E. coli* O157:H7 Sakaï contiennent la dernière moitié et les derniers deux tiers, respectivement, d'ORFs codant des intégrases putatives présentant entre elles 82,5 % d'identité et localisées chacune au sein de régions génomiques variables et complètement différentes.

4.2.3. Caractérisation partielle de la région aval au locus *serU* grâce au séquençage partiel d'un cosmide recombinant

Dans le but d'obtenir de plus grandes informations sur les séquences nucléotidiques localisées en 3' des loci d'ARNt pour lesquels des structures altérées avaient été détectées, une banque de cosmides recombinants a été préparée à partir de l'ADN total d'*E. coli* P4. Les cosmides recombinants contenant les loci d'intérêt ont ensuite été recherchés par PCR en utilisant des couples d'oligonucléotides spécifiques des séquences situées en aval des loci (déterminées par PCR arbitraire) et confirmés par hybridation, en utilisant comme sondes les produits d'amplification obtenus par PCR avec les mêmes couples d'oligonucléotides à partir de l'ADN total d'*E. coli* P4. Un cosmide recombinant pouvant contenir l'ensemble ou une partie de la région variable localisée en aval du locus *serU* a été ainsi identifié. Le séquençage partiel de ce plasmide a permis d'obtenir une séquence de 13202 nucléotides présentant un pourcentage en paires de bases guanine et cytosine de 42,5 % et située immédiatement en 3' du locus *serU* sur le génome d'*E. coli* P4. Sept ORFs ont été détectées. Leurs caractéristiques sont présentées dans le tableau 24 et la figure 33. Les ORFs 3-*serU*, 5-*serU* et

7-*serU* codent des protéines qui possèdent l'intégralité des motifs caractéristiques des intégrases et des transposases et qui pourraient donc être toujours fonctionnelles. L'ORF 6-*serU* code une protéine qui ne possède que partiellement ces mêmes motifs et qui n'est donc probablement plus fonctionnelle. L'ORF 1-*serU* code une protéine qui pourrait être impliquée dans les phénomènes de glycosylation de la partie O-antigène des lipopolysaccharides. L'ORF 4-*serU* code une protéine de fonction inconnue. Enfin, l'ORF 2-*serU* code une protéine de séquence inconnue dans les banques de données et qui ne présente aucun domaine conservé connu.

Tableau 24 : Caractéristiques des ORFs détectées dans la séquence nucléotidique partielle du cosmide recombinant contenant le locus *serU* du génome d'*E. coli* P4.

ORF	Position	RBS séquence position	GC (%)	Taille (acides aminés)	Domaine conservé	Caractéristiques des protéines homologues			
						Fonction	Taille (acides aminés)	Appartenance	% d'identité / % de similarité
1-*serU*	168-1580	AGGGG 96-100	32,7	470	-	Protéine de conversion du O-antigène	485	Entérobactériophages (ST64T)	24 / 45
2-*serU*	2941-1625	AGGGG 2996-2992	39,6	438	-	-	-	-	-
3-*serU*	3886-3074	AGCGG 3973-3969	50,3	270	Intégrase/Transposase (complet)	Transposase d'IShe3	270	*E. coli* HS	99 / 99
4-*serU*	6702-4600	AGGGG 6794-6790	34,3	700	-	Protéine hypothétique TnpX	698	Plasmide F	100 / 100
5-*serU*	7427-10435	GGGAG 7416-7420	48,5	1002	Transposase (complet)	Transposase TnpA	1002	Plasmide F	100 / 100
6-*serU*	11117-10359	-	46,2	252	Transposase (partiel)	Transposase d'IS30	383	*E. coli* E24377A	100 / 100
7-*serU*	12098-13087	-	44,7	329	Intégrase (complet)	Intégrase	329	*S. typhimurium* phage ST64B	84 / 94

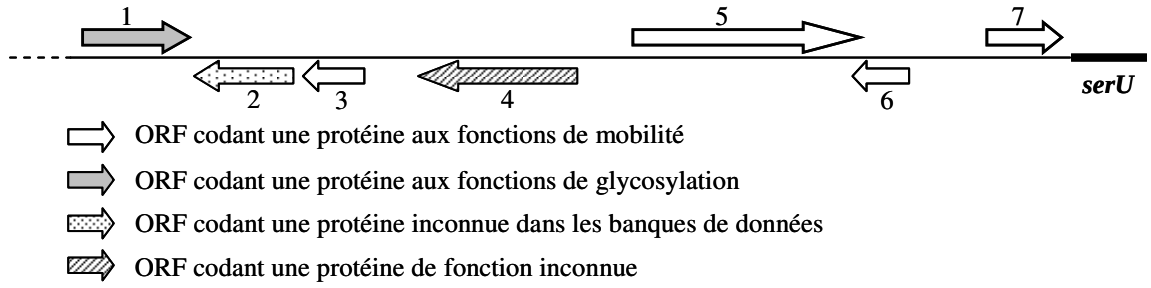


Figure 33 : Carte génétique partielle de l'îlot génomique situé en 3' du locus *serU* sur le génome d'*E. coli* P4. Les chiffres indiquent les numéros des ORFs présentées dans le tableau 24.

4.2.4. Caractérisation complète de la région aval au locus *argW*

La région située immédiatement en 3' du locus *argW* sur le génome d'*E. coli* P4 a également été amplifiée dans un premier temps par PCR arbitraire et séquencée. La séquence obtenue, d'une taille de 747 pb, présentait une seule séquence similaire (94 % d'identité) dans les banques de données. Il s'agissait d'une séquence identifiée immédiatement en 3' d'un locus d'ARNt-Arg sur le

génome de la souche aviaire pathogène *E. coli* APEC BEN1588. Etant donné qu'aucune autre information n'était disponible pour cette séquence et qu'elle ne présentait aucune homologie avec les séquences nucléotidiques déterminées jusque là pour les autres représentants de l'espèce *E. coli*, nous avons décidé de séquencer entièrement la région génomique variable située en 3' du locus *argW* sur le génome d'*E. coli* P4. Pour cela, cette région a tout d'abord été amplifiée par long-range PCR. En effet, cette région dont les tailles connues sur les génomes d'autres souches *E. coli* sont par exemple de 12819 pb chez *E. coli* K12 MG1655, 14357 pb chez *E. coli* O157 :H7 Sakaï et 14791 pb chez *E. coli* CFT073, ne peut être amplifiée par PCR classique. Un couple d'oligonucléotides spécifique a été dessiné pour cette amplification. L'oligonucléotide « forward » utilisé était complémentaire à un fragment de la séquence retrouvée immédiatement en 3' du locus *argW* sur le génome d'*E. coli* P4 déterminée par PCR arbitraire. L'oligonucléotide « reverse » utilisé était quant à lui complémentaire à un fragment de la séquence du gène *dsdA*, premier gène constant retrouvé jusqu'à présent sur tous les génomes des souches *E. coli* en 3' du locus *argW*. La réalisation de cette long-range PCR a permis d'obtenir un produit d'amplification d'une taille de 12 kb environ. Après le séquençage de ce produit et la soustraction des séquences correspondant aux gènes *argW* et *dsdA* à la séquence obtenue, une séquence de 11253 pb, présentant un pourcentage en paires de bases guanine et cytosine de 40,92 %, a pu être définie. Ce pourcentage global en paires de bases guanine et cytosine présente de grandes disparités à l'intérieur de la séquence (tableau 25). De plus, il se distingue nettement de ceux des régions génomiques *argW* des souches *E. coli* K12 MG1655 (44,67 %), *E. coli* O157:H7 Sakaï (45,10 %) et *E. coli* CFT073 (44,35 %) et de ceux des génomes complets de ces mêmes souche (50,8, 50,5 et 50,47 %, respectivement). Des nucléotides 8628 à 8650 de la séquence obtenue, une séquence de 23 pb identique à 100 % aux 23 derniers nucléotides du gène *argW* est retrouvée. Les 11253 pb ainsi amplifiés et séquencés en 3' du locus *argW* sur le génome d'*E. coli* P4 présentent donc plusieurs caractéristiques des îlots génomiques : un site d'intégration en aval d'un locus d'ARNt, une répétition de l'extrémité 3' de ce gène d'ARNt, une grande taille et un pourcentage en paires de bases guanine et cytosine atypique.

4.2.4.1. Contenu en ORFs

Huit ORFs, dont les caractéristiques sont présentées dans le tableau 25 et la figure 34, ont pu être détectées au sein de la région variable de 11253 pb située en 3' du locus *argW* sur le génome d'*E. coli* P4. Pour chacune des protéines codées par ces ORFs, des protéines similaires ont pu être retrouvées dans les banques de données : deux codent des protéines aux fonctions de mobilité (les ORFs 1-*argW* et 6-*argW*), trois codent des protéines aux fonctions de réplication / transcription (les ORFs 2-*argW*, 3-*argW* et 7-*argW*), une code une protéine à la fonction de transport (l'ORF 8-*argW*), une code une protéine de fonction inconnue (l'ORF 5-*argW*) et une enfin code une protéine à la

fonction de déacétylation (l'ORF 4-*argW*). Les régions génomiques séparant ces ORFs ne présentent aucune similarité avec les séquences des banques de données.

Tableau 25 : Caractéristiques des ORFs détectées sur l'îlot génomique situé en 3' du locus *argW* sur le génome d'*E. coli* P4.

ORF	Position	RBS séquence position	GC (%)	Taille (acides aminés)	Domaine conservé	Caractéristiques des protéines homologues			
						Fonction	Taille (acides aminés)	Appartenance	% d'identité / % de similarité
1-*argW*	165-1355	AGGGG (127-131)	43.07	396	P4 intégrase (partiel)	Intégrase	419	*Sa. enterica* ATCC 9150	68 / 82
2-*argW*	1512-1703	GAGG (1503-1506)	46.88	63	Régulateur transcriptionnel (complet)	Régulateur transcriptionnel putatif	65	*Y. intermedia* ATCC29909	73 / 91
3-*argW*	2671-3723	-	43.59	350	-	Protéine de réplication putative	231	*E. coli* O157:H7 Sakaï	97 / 99
4-*argW*	6910-4430	AAGAG (6921-6917)	31.20	826	Sir2 (complet)	Protéine hypothétique	792	*Rhodopseudomonas palustris* HaA2	23 / 41
5-*argW*	7564-7388	-	43.50	58	-	Protéine hypothétique	58	*E. coli* O157:H7 Sakaï	77 / 84
6-*argW*	7867-8493	-	47.69	208	Recombinase/Résolvase (complet)	Résolvase putative	208	*E. coli* O157:H7 Sakaï	98 / 99
7-*argW*	9681-8746	ATGAG (9692-9688)	48.08	311	Régulateur transcriptionnel (complet)	Régulateur transcriptionnel	311	*E. coli* K12 MG1655	100 / 100
8-*argW*	9899-11236	AGGGG (9884-9888)	48.73	445	Transporteur (complet)	Transporteur putatif	445	*E. coli* K12 MG1655	100 / 100

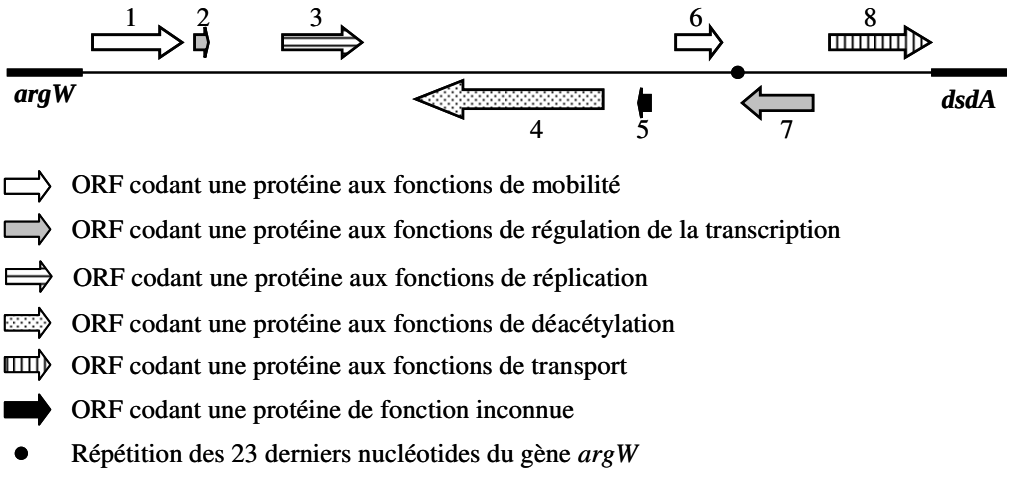


Figure 34 : Carte génétique de l'îlot génomique situé en 3' du locus *argW* sur le génome d'*E. coli* P4. Les chiffres indiquent les numéros des ORFs présentées dans le tableau 25.

La séquence en acides aminés de la protéine codée par l'ORF 1-*argW* est présentée dans la figure 35 et comparée à celle d'une intégrase de la souche *Salmonella enterica* ATCC 9150. Seules les premières moitiés des séquences sont similaires, montrant que la protéine codée par l'ORF 1-*argW* semble avoir conservé uniquement la partie N-Terminale de l'intégrase, c'est-à-dire le domaine de fixation à l'ADN et perdu la partie C-Terminale catalytique. Aucune homologie avec les séquences disponibles dans les banques de données n'a pu être trouvée pour la partie C-Terminale de l'ORF 1-

argW. Cette dernière peut donc être définie comme une protéine chimérique, résultant de l'association d'un domaine N-Terminal de recombinase à tyrosine et d'un fragment de protéine inconnue. Chez la souche *Salmonella enterica* ATCC 9150, l'ORF codant l'intégrase partiellement identique à la protéine codée par l'ORF 1-*argW* d'*E. coli* P4 est située immédiatement en 3' du locus *leuX* et constitue la première ORF d'un îlot de pathogénicité appelé SPI-10 d'une taille de 33 kb et contenant des gènes codant des protéines de type fimbriae. Au sein de l'espèce *Salmonella enterica*, le locus *leuX* est très souvent associé à l'accueil d'ADN étranger par transfert génétique horizontal et une grande variabilité dans la nature de cet ADN étranger est observée, même au sein d'un même sérovar (Bishop *et al*., 2005).

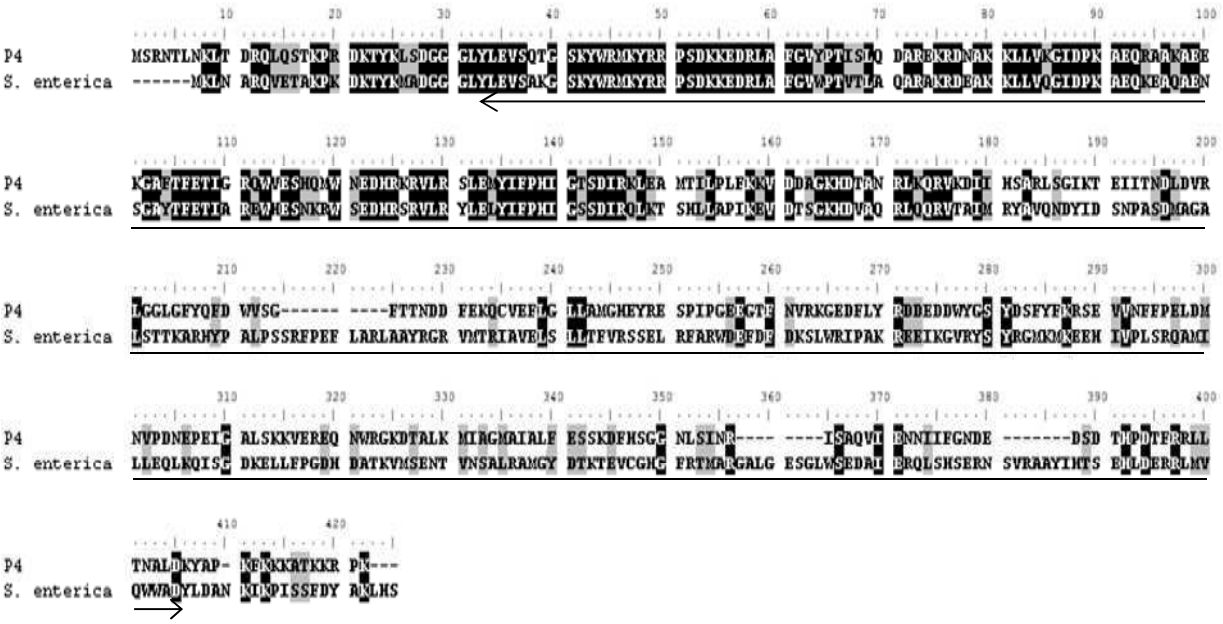


Figure 35 : Alignement des structures primaires de la protéine codée par l'ORF 1-*argW* d' *E. coli* P4 et d'une intégrase codée sur le génome de la souche *Salmonella enterica* ATCC 9150. Les acides aminés identiques entre les deux séquences sont inscrits sur fond noir, ceux qui sont similaires sur fond gris et ceux qui sont différents sur fond blanc. La double flèche indique l'intégralité du domaine conservé de l'intégrase présente chez *Salmonella enterica* ATCC 9150.

L'ORF 2-*argW* code une protéine similaire à un régulateur transcriptionnel de la souche *Yersinia intermedia* ATCC 29909. Sur le génome de cette dernière, l'ORF codant le régulateur transcriptionnel se situe en 3' d'un gène codant une intégrase ne se trouvant pas à proximité d'un locus d'ARNt.

La séquence en acides aminés de la protéine codée par l'ORF 3-*argW* est présentée dans la figure 36 et comparée à celle de la protéine de réplication putative de la souche *E. coli* O157:H7 Sakaï. Cent dix-neuf acides aminés supplémentaires sont présents à l'extrémité N-Terminale de l'ORF 3-*argW*. Chez *E. coli* O157:H7 Sakaï, la protéine de réplication putative est codée par une ORF située dans la région génomique variable *argW*.

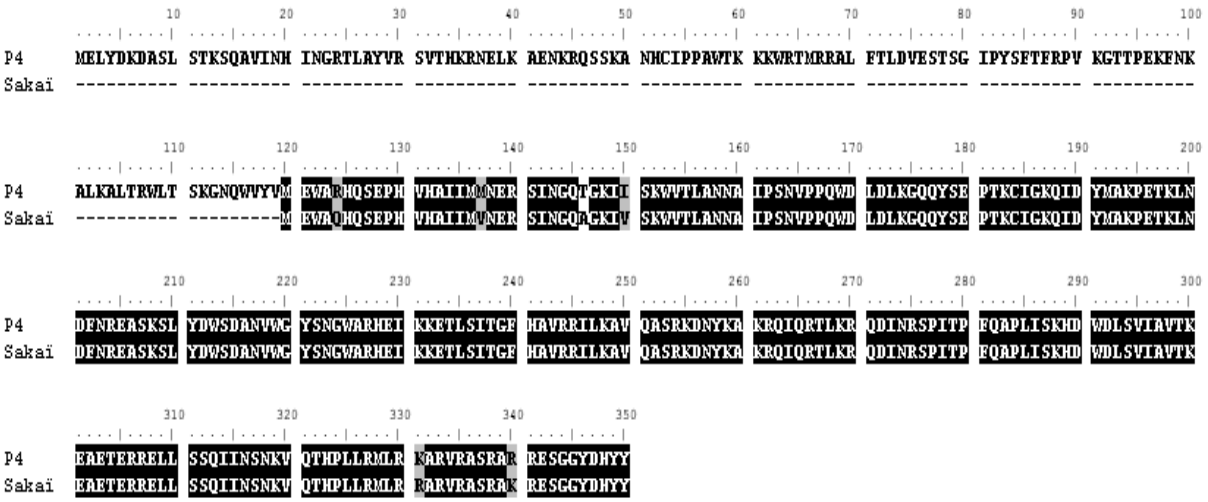


Figure 36 : Alignement des structures primaires de la protéine codée par l'ORF 3-*argW* d' *E. coli* P4 et de la protéine de réplication putative codée sur le génome de la souche *E. coli* O157:H7 Sakaï. Les acides aminés identiques entre les deux séquences sont inscrits sur fond noir, ceux qui sont similaires sur fond gris et ceux qui sont différents sur fond blanc.

L'ORF 5-*argW* code une protéine similaire à une protéine hypothétique de la souche *E. coli* O157:H7 Sakaï codée par une ORF située également dans la région génomique variable *argW*. Aucune fonction connue n'a pu être encore attribuée à cette protéine.

L'ORF 6-*argW* code une protéine similaire à une protéine aux fonctions de recombinase/résolvase de la famille des recombinases à sérine, toujours fonctionnelle, de la souche *E. coli* O157:H7 Sakaï codée par une ORF également située dans la région génomique variable *argW*.

Les deux dernières ORFs détectées, les ORF 7-*argW* et 8-*argW* présentent 100 % d'identité avec les gènes *dsdC* et *dsdX*, respectivement, situés immédiatement en 5' du gène *dsdA* sur le génome d'*E. coli* K12 MG1655. Chez *E. coli* K12 MG1655, la protéine DsdC, codée par le gène *dsdC*, a pour fonction la régulation de la transcription d'une enzyme appelée D-sérine déhydratase et impliquée dans la dégradation des petites molécules tels que les acides aminés. Sur les 10 génomes entièrement séquencés de l'espèce *E. coli* présentés dans le paragraphe 3.2 du chapitre « Introduction » de ce manuscrit, le gène *dsdC* est présent en entier sur les génomes des souches K12 W3110, HS, APEC01, UTI89, CFT073 et 536 et absent sur les génomes des souches O157:H7 Sakaï, EDL933 et E24377. Chez *E. coli* K12 MG1655, la protéine DsdX, codée par le gène *dsdX*, serait impliquée dans un système de transport membranaire encore mal défini. De la même façon, sur ces 10 mêmes génomes, le gène *dsdX* est retrouvé en entier sur les génomes des souches K12 W3110, HS, APEC01, UTI89, CFT073 et 536 et seuls les 220 derniers nucléotides de ce gène sont retrouvés sur les génomes des souches O157:H7 Sakaï, EDL933 et E24377.

4.2.4.2. Analyse de l'ORF 4-*argW*

L'ORF 4-*argW* est l'ORF de la plus grande taille détectée dans la région génomique variable située en 3' du locus *argW* du génome d'*E. coli* P4. Elle est composée de 2481 nucléotides et présente un pourcentage en paires de bases guanine et cytosine de 31,2 %, valeur nettement inférieure à la valeur moyenne des génomes de l'espèce *E. coli* (proche de 50 %). Une séquence RBS, de type AAGAG, est présente 11 pb en amont du codon d'initiation de la traduction, ainsi qu'un signal de terminaison de la transcription rho-indépendant 14 pb en aval du codon d'arrêt de la traduction. Cette ORF code une protéine de 826 acides aminés, présentant une masse moléculaire de 95,18 kDa et un domaine conservé appelé Sir2 (Silent Information Regulator 2). Les domaines Sir2 sont caractéristiques des enzymes catalysant des réactions de déacétylation NAD^+ dépendantes, appelées protéines Sir2 ou sirtuins. Ces dernières possèdent plusieurs motifs consensus tels que GAGISXXXGIPXXR, TQNID, GTS et (I/V)N retrouvés quelque peu modifiés dans la séquence de la protéine codée par l'ORF 4-*argW* (GAGFSRKTKNITNGEPPLA, TTNYD, GYS et IV, respectivement) et d'autres tels que PXXXH, HG, CXXC deux fois et FGE absents dans la séquence de la protéine codée par l'ORF 4-*argW* (figure 37). Les réactions de déacétylation NAD^+ dépendantes catalysées par les protéines Sir2 ont été très étudiées chez la levure et l'homme et il a été montré qu'elles intervenaient notamment dans la régulation du "gene silencing", de la réparation de l'ADN, des enzymes métaboliques et de la durée de vie des cellules (North et Verdin, 2004). Chez les bactéries, les protéines Sir2 sont peu étudiées et leurs fonctions restent encore non élucidées. La protéine hypothétique de la souche *Rhodopseudomonas palutris* HaA2 similaire à la protéine codée par l'ORF 4-*argW* d'*E. coli* P4 n'a pas fait l'objet d'une étude fonctionnelle et a simplement été découverte lors du séquençage du génome de cette souche.

Figure 37 : Structure primaire de la protéine codée par l'ORF 4-*argW* d'*E. coli* P4. Les doubles flèches indiquent des motifs conservés au sein des protéines Sir2.

4.2.5. Conclusion

La recherche d'îlots génomiques au sein du génome d'*E. coli* P4 a révélé l'existence de telles structures en 3' d'au moins six loci d'ARNt : *argW*, *asnT*, *leuX*, *pheV*, *serU* et *thrW*. Les caractéristiques partielles ou complètes de ces six loci ont montré dans chaque cas des séquences différentes de celles présentes chez la souche *E. coli* K12 MG1655. De plus, les données plus importantes obtenues pour les loci *pheV*, *serU* et *argW* ont permis de mettre en évidence des structures très chimériques. En effet, pour ces trois loci, de grandes variations en pourcentage en paires de bases guanine et cytosine, ainsi que des gènes d'origines très différentes ont été identifiés. Ces faits suggèrent que les îlots génomiques *pheV*, *serU* et *argW*, au sein du génome d'*E. coli* P4, ont été créés après plusieurs évènements de transfert génétique horizontal. Le cas de l'îlot génomique *serU* soutient bien cette hypothèse, étant donné la succession d'ORFs codant des enzymes de mobilité d'origines différentes qui a pu être mise en évidence à son extrémité 5'.

4.3. Affiche

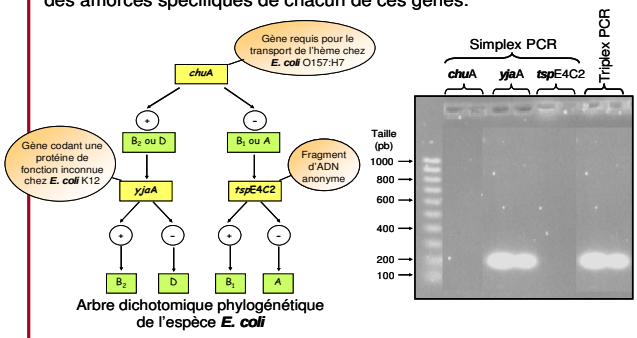
Recherche de facteurs de virulence chez *Escherichia coli* P4O32, bactérie inductrice de mammites bovines

Delphine Dufour, Yves Le Roux, Annie Dary.
Unité de Recherche Animal et Fonctionnalité des Produits Animaux, 54506 Vandoeuvre les Nancy.

Escherichia coli est l'une des espèces bactériennes de la flore intestinale anaérobie facultative des hommes et des animaux homéothermiques. Néanmoins, certaines souches de cette espèce sont responsables de pathologies intestinales et extra intestinales. ***E. coli*** P4O32 figure parmi ces souches pathogènes, étant à l'origine chez le bovin d'inflammations de la glande mammaire, encore appelées mammites. Si les mécanismes de l'infection mammaire chez le bovin sont bien renseignés, les propriétés de virulence permettant à ***E. coli*** P4O32 de coloniser la mamelle demeurent inconnues. Pourtant, cette bactérie est très utilisée pour l'induction de mammites expérimentales. L'objectif de ce travail est de caractériser les facteurs responsables de la pathogénicité d'***E. coli*** P4O32. La démarche expérimentale choisie est la suivante : détermination du groupe phylogénétique et recherche d'antigènes de surface et d'îlots de pathogénicité.

Détermination du groupe phylogénétique

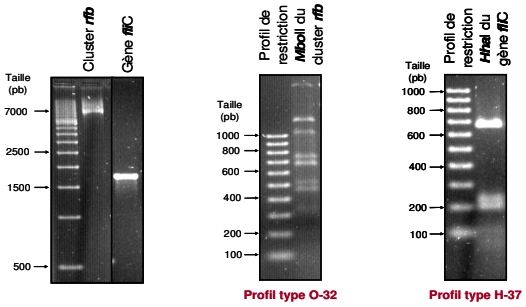
L'espèce ***E. coli*** est divisée en quatre groupes phylogénétiques : A, B_1, B_2 et D. L'appartenance d'une souche à l'un de ces groupes peut être déterminée grâce à l'utilisation de trois marqueurs phylogénétiques : ***chu***A, ***yja***A et ***tsp***E4C2. Ces trois marqueurs phylogénétiques ont été recherchés au sein du génome d'***E. coli*** P4O32 par des expériences de simplex et de triplex PCR en utilisant des amorces spécifiques de chacun de ces gènes.



Appartenance d'*E. coli* P4O32 au groupe phylogénétique A de l'espèce *E. coli*

Mise en évidence d'antigènes de surface

Le sérotypage des bactéries ***E. coli*** repose sur l'étude de trois antigènes potentiels : le O-Ag (extrémité du LPS), le H-Ag (flagelline) et le K-Ag (capsule). Un cluster de gènes ***rfb*** impliqué dans la synthèse d'un O-Ag et un gène ***fli***C codant un H-Ag ont été mis en évidence par long-range PCR dans le génome d'***E. coli*** P4O32 puis typés grâce à leurs profils de restriction ***Mbo***II et ***Hha***I respectivement.

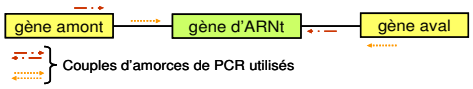


Expression potentielle par *E. coli* P4O32 d'un O-Ag 32 et d'un H-Ag 37

Recherche d'îlots de pathogénicité (PAI)

Les souches ***E. coli*** pathogènes se distinguent des souches non pathogènes par la présence dans leur génome d'îlots de pathogénicité. Ces régions chromosomiques, de 10 à 200 kpb, acquises lors de transferts horizontaux, codent divers facteurs de virulence et sont le plus souvent associés à des loci d'ARNt. Ces îlots ont été recherchés chez ***E. coli*** P4O32 par PCR en utilisant des amorces spécifiques de gènes d'ARNt du génome de la souche ***E. coli*** K12. L'absence d'amplification indique la présence de structures génomiques variables (PAI ?).

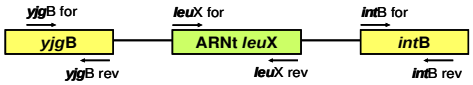
➢ **Amplification génomique par PCR de sept loci d'ARNt chez *E. coli* P4O32 :**



Loci d'ARNt	ARNt *metV*	ARNt *selC*	ARNt *phcU*	ARNt *pheU*	ARNt *pheV*	ARNt *leuX*	ARNt *thrW*
E. coli P4O32	+	+	+	+	+	–	–
E. coli K12 (témoin)	+	+	+	+	+	+	+

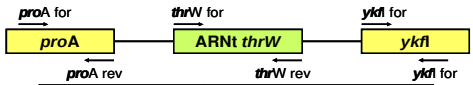
+ / - : présence / absence d'un amplicon de taille attendue

➢ **Etude détaillée de la structure génomique des loci d'ARNt *leu*X et *thr*W chez *E. coli* P4O32 :**



Couple d'amorces de PCR	*yjg*B for/rev	*leu*X for/rev	*int*B for/rev	*leu*X for / *int*B rev
E. coli P4O32	+	+	–	–
E. coli K12 (témoin)	+	+	+	+

+ / - : présence / absence d'un amplicon de taille attendue



Couple d'amorces de PCR	*pro*A for/rev	*thr*W for/rev	*ykf*I for/rev	*thr*W for / *ykf*I rev
E. coli P4O32	+	+	–	–
E. coli K12 (témoin)	+	+	+	+

+ / - : présence / absence d'un amplicon de taille attendue

Présence chez *E. coli* P4O32 de structures génomiques, en aval des gènes *leu*X et *thr*W, différentes de celles connues d'*E. coli* K12

Les résultats obtenus montrent qu'***E. coli*** P4O32 appartient au groupe A de l'espèce ***E. coli***, groupe qualifié de commensal et non pathogène (comportant la souche de laboratoire K12). La présence d'un cluster de gènes ***rfb*** et d'un gène ***fli***C dans son génome, à l'origine de l'expression potentielle de deux antigènes de surface de type O32 et H37, peut expliquer sa capacité à induire des mammites bovines contrairement aux autres souches de ce groupe phylogénétique qui en sont incapables. La variabilité génétique observée en aval des gènes ***leu***X et ***thr***W sera caractérisée davantage afin de déterminer s'il s'agit d'îlots de pathogénicité. Dans ce but, ces régions génomiques vont être recherchées dans une banque d'ADN génomique d'***E. coli*** P4O32 et séquencées.

4.4. Publication

First evidence of the presence of genomic islands in the bovine mammopathogen strain *Escherichia coli* P4

Delphine Dufour[1], Pierre Germon[2], Emilie Brusseaux[1], Yves Le Roux[1] and Annie Dary[1]*

[1] URAFPA, Nancy-Université, INRA, Equipe Protéolyse-Biofonctionnalité des Protéines et des Peptides, Boulevard des Aiguillettes, BP239, 54506 Vandoeuvre-lès-Nancy, France.
[2] INRA – Centre de Tours, UR 1282 IASP, Pathogénie Bactérienne, 37380 Nouzilly, France.

(*) corresponding author. Tel.: +3383684267; fax: +3383595889; email : Annie.Dary@scbiol.uhp-nancy.fr

Abstract

On the contrary to other pathotypes of *Escherichia coli*, the strains which are responsible for bovine mastitis, or MPEC strains (MammoPathogenic *Escherichia coli*), remain poorly characterized, perhaps because they were considered as simple opportunistic and environmental pathogens rapidly killed by the immune system of the cows. However, several studies have reported different virulence properties of some MPEC strains, such as the persistence in the udder for prolonged periods, or the adhesion and invasion of the bovine mammary epithelial cells. On the basis of the known virulence genes specific to the other pathotypes of the *E. coli* species, no specific genetic markers have yet been identified to explain these different virulence MPEC properties. The aim of this work was to research specific properties of a mammopathogen *E. coli* strain, *E. coli* P4, which is used worldwide to induce experimental mastitis, and for which bovine mammary epithelial cell adhesion and invasion properties have been demonstrated. In a first time, it was established that *E. coli* P4 belongs to the phylogenetic group A of the *E. coli* species, and that its core genome is very similar to that of the commensal no pathogen *E. coli* K12 MG1655 strain. In a second time, different virulence genes, specific of other *E. coli* pathotypes, and known to be implicated in properties of cell adhesion and invasion were researched, but only the presence of the *traT* gene, encoding a protein responsible for serum resistance was detected. Finally, seventeen tRNA loci, known to be frequently associated with genomic islands (GEIs), were screened, and an altered structure was detected for seven of them. The partial characterizations of five of them (*asnT*, *leuX*, *pheV*, *serU*, and *thrW*) and the complete characterization of one of them (*argW*) showed the presence of genomic islands which are different or absent from the *E. coli* K12 MG1655 genome, and from GEI already described in other pathotypes.

Keywords: mammopathogen *Escherichia coli*, bovine mastitis, phylogeny, genomic islands.

1. Introduction

Bovine mastitis is the most prevalent and economically important disease in dairy cattle in the world (Bradley, 2002). It is defined as an inflammation of the mammary gland, with various clinical signs, such as elevated rectal temperature, dullness, loss of appetite, appearance of diarrhoea, or a swelling and painful udder (Menzies *et al*., 1999). Sometimes it leads to the animal death. In addition, milk production and quality are affected (Michelutti *et al*., 1999). Production drops considerably (from 8.5 $L.day^{-1}$ for a healthy cow to 3.5 $L.day^{-1}$ for a sick cow), and never gets back to its initial level even after the cure. Quality is deeply spoiled with, especially, an important proteolysis of milk caseins, i.e. the major milk proteins, and the essential milk components used for the cheese manufacturing. In most cases, bovine mastitis results from an invasion of the udder by bacteria such as *Escherichia coli*, which became the most prevalent (Wenz *et al*., 2001).

E. coli strains which induce bovine mastitis, are called MammoPathogen Escherichia Coli (MPEC). They belong to the extraintestinal pathogenic group (ExPEC) which consists of several pathotypes such as the uropathogenic (UPEC) pathotype or the meningitis/septis pathotype (MNEC) (Johnson and Russo, 2002). Epidemiologic studies did not allow to assign particular *E. coli* serotypes to mastitis, nor to determine whether these strains share a common set of virulence factors. Hence, the MPEC strains were considered as simple opportunistic invaders of the mammary gland. As, virulence factors, which are known to be associated with invasive diarrheagenic and extraintestinal pathogenic *E. coli* strains, were not detected in the MPEC strains (Kaipainen *et al*., 2002), it was concluded that their LPS is sufficient for their virulence. Finally, it was proposed that severity of a mastitis caused by a MPEC strain only depends on cow factors (Burvenich *et al*., 2003).

However, several publications showed that other factor, different from the LPS, could be involved in the MPEC pathogenesis. Indeed, Dopfer *et al*. (2000 and 2001) and Dogan *et al*. (2006) highlighted adhesion and invasion properties of some MPEC strains to mammary epithelial cells, while Bradley and Green (2000 and 2001) showed that some MPEC strains are able to persist in the udder for prolonged periods (more than 100 days), and to cause recurrent clinical mastitis. Such recurrent mastitis imply that these MPEC strains possess specific gene allowing them to persist and survive in the mammary gland, which contains many anti-bacterial entities such as somatic cells, immunoglobulins, lactoferrin, complement, lysozyme, lactoperoxidase, or cytokines (Sordillo *et al*., 1997).

It is clearly established now that specific properties of some strains belonging to a same species are confered by genomic sequences called genomic islands (GEIs) which have been acquired by horizontal gene transfer (Dobrindt *et al*., 2004). These regions encode accessory functions such as

additional metabolitic activities, antibiotic resistance, symbiosis or pathogenesis properties. They play an important role in the adaptative evolution of bacteria to their environment. For a decade, GEIs have been widely researched and analysed in intestinal and extraintestinal pathogenic *E. coli* strains (Hacker and Kaper, 2000) as well as in commensal ones (Grozdanov *et al*., 2004; Hejnova *et al*., 2006), but not yet in the MPEC strains, whereas they could allow to explain their virulence.

The aim of this work was to research potential GEIs in the genome of the bovine mammopathogen *E. coli* P4 strain, which was isolated from an acute bovine mastitis (Bramley, 1976), and which is worldwide used to induce experimental bovine mastitis. In a first time, phylogenetic analyses were carried out to assign *E. coli* P4 to one of the four main phylogenetic groups of the *E. coli* species. Secondly, typical virulence genes, known to be present in the genome of calf or cow pathogenic *E. coli* strains, were researched, as well as GEIs.

2. Materials and methods

2.1. Bacterial strains, plasmid, and culture conditions

E. coli P4 was provided by the Department of Physiology, Biochemistry and Biometrics (Professor Burvenich, Gent, Belgium). *E. coli* K12 MG1655, used as a control, was purchased from Institut Pasteur (Paris, France). *E. coli* EPI100^{TM} (EPICENTRE, Madison, Wisconsin) was used as the recipient strain for the construction of a cosmid library of *E. coli* P4 genome and the cosmid pWEB:Tnc^{TM} (EPICENTRE, Madison, Wisconsin) was used as vector.

The *E. coli* strains were grown in Luria-Bertani (LB) medium (Sambrook *et al*., 1989) with or without 1.5 % (wt.vol^{-1}) Bacto Agar (Difco Laboratories, Detroit, Michigan) at 37°C. Recombinant strains were grown in LB medium supplemented with ampicillin (100 µg.mL^{-1}) (EUROMEDEX, Souffelweyersheim, France). All strains were stored at -70°C in LB medium supplemented with glycerol (20 % (wt.vol^{-1})).

2.2. DNA isolation, electrophoresis and sourthern hybridization

Genomic DNAs of *E. coli* P4 and *E. coli* K12 MG1655 were isolated as described previously (Leblond *et al*., 1996). Genomic DNA of *E. coli* O157:H7 EDL933 was provided by Pierre Germon (INRA, Tours, France). For the construction of the cosmid library of *E. coli* P4 genome, genomic DNA of *E. coli* P4 was isolated as described previously (Sambrook *et al*., 1989). Recombinant cosmids were extracted by the alkaline lysis method (Sambrook *et al*., 1989). Purification of DNA fragments and extraction from agarose gel slices were performed with NucleoSpin® Extract II system (MACHEREY-NAGEL, Hoerdt, France). The DNA fragments were separated on agarose gels

according to Sambrook *et al*. (1989). Both 1 kb and 100 bp DNA ladders (Bio-Rad, Hercules, CA, USA) were used as molecular weight markers. Southern blotting, prehybridization and hybridization conditions were used as previously described by Demuyter *et al*. (1988). Labeling of probes and detection of specific hybrids were performed using the DIG DNA labeling and detection kit, under the conditions recommended by the supplier (Roche Molecular Biochemicals, Mannheim, Germany).

2.3. Cosmid library construction and screening

Genomic DNA of *E. coli* P4 was partially digested with the restriction endonuclease *Sau*3AI (New England Biolaps, Beverley, MA) and ligated to the *Bam*HI (New England Biolaps, Beverley, MA) digested cosmid vector pWEB:Tnc™, treated before with alkaline phosphatase (New England Biolaps, Beverley, MA). The recombinant cosmids were packaged in phage heads with the Gigapack III XL system (Stratagene, Heidelberg, Germany), according to the instructions of the manufacturer, and used to infect *E. coli* EPI100™. Recombinant cosmids containing the six altered tRNA loci identified were firstly screened by PCR, using six pairs of primers designed according to the nucleotide sequences of the region located downstream from these six tRNA loci and determined by arbitrary PCR. PCR products obtained using these pairs of primers and the *E. coli* P4 genomic DNA as template, were also used as probes in Sourthern hybridization experiments to confirm the identified recombinant cosmids. These probes were labelled and hybridized against *Eco*RI (New England Biolaps, Beverley, MA) restriction digest of the identified recombinant cosmids.

2.4. PCR primers

All PCR primers used in this study were purchased from Eurogentec Society (Belgium).
Primers used for the determination of phylogenetic group were firstly described by Clermont *et al*. (2000). They enable the partial amplification of *chuA*, a gene required for heme transport in enterohemorrhagic *E. coli* O157:H7*, yjaA*, a gene of unknown function identified in the complete genome sequence of *E. coli* K12 MG1655 (Accession number: U00096.2), and tspE4C2, an anonymous DNA fragment associated with neonatal meningitis strains. These three amplifications enable the assignment of a strain to a phylogentic group as follows: *chuA* positive and *yjaA* positive, group B2; *chuA* positive and *yjaA* negative, group D; *chuA* negative and tspE4C2 positive, group B1; *chuA* negative and tspE4C2 negative, group A.
Primers used for the amplification of the 16S-23S rDNA intergenic spacer (16S-23S ITS) region correspond to those already described by Garcia-Martinez *et al*. (1996). They target the rRNA operons of *E. coli* K12 MG1655 from 16S rRNA position 1389 to 23S rRNA position 130.

Primers used for MLST study were firstly described by Adiri *et al.* (2003), and Reid *et al*. (2002). They enable the partial amplification of different housekeeping genes in the genome of *E. coli* K12 MG1655: *adk* (adenylate kinase), *gcl* (glyoxylate carboligase), *gdh* (glucose-6-phosphate dehydrogenase), *mdh* (malate dehydrogenase), *metA* (homoserine transsuccinylase), *ppk* (polyphosphate kinase), *aspC* (aspartate aminotransferase), *clpX* (ATPase and specificity subunit of ClpX-ClpP ATP-dependent serine protease), *fadD* (acyl-CoA synthetase), *icdA* (isocitrate dehydrogenase), *lysP* (lysine transporter), *aroE* (shikimate dehydrogenase), *cstA* (carbon starvation protein), *cya*A (adenylate cyclase), *dnaG* (DNA primase), *grpE* (heat shock protein), *mtlD* (mannitol 1-phosphate dehydrogenase), *mutS* (methyl-directed mismatch repair protein), *rpoS* (sigma fator 38), *pgi* (phosphoglucose isomerase). Additional primers were designed for the partial amplification of genes encoding the serine endoprotease Do and the following putative proteases of unknown function in the genome of *E. coli* K12 MG1655: *ptrB* (proteinase II), *ycaL* (putative protease), *yhjS* (putative protease), *yaeL* (zinc metallopeptidase), *sms* (probable ATP-dependent protease), *yhjJ* (predicted zinc-dependent peptidase), *elaD* (predicted enzyme), *pmbA* (predicted peptidase required for the maturation and secretion of the antibiotic peptide MccB17) (table 1).

Virulence factors were researched using primers already described and allowing the detection of genes encoding: (i) major sub-unit A of fimbrial adhesins F17a, F17b, F17c, and F17d (Bertin *et al*., 1996); (ii) minor sub-units E and F of fimbrial adhesins P, minor sub-unit D and chaperon protein E of fimbrial adhesins S, N-lysin monooxygenase required for aerobactin synthesis, and cytotoxic necrotizing factor 1 (CNF1) (Yamamoto *et al*., 1995); (iii) cytotoxic necrotizing factor 2 (CNF2), and serum resistance factor TraT (Kaipainen *et al*., 2002); (iv) surface antigen CS31A (Bertin *et al*., 1998); (v) adhesin and invasin sub-units of afimbrial adhesins Afa7 and Afa8 (Lalioui *et al*., 1999); (vi) intimin EaeA, shiga like toxin SltI and SltII (China *et al*., 1996); (vii) heat labile enterotoxin EltIA, EltIIA, heat stable enterotoxin StIA or StIB, and polysialic acid synthetase NeuC and acylneuraminate cytidyldyltransferase NeuA, both required for capsule synthesis (Kuhnert *et al*., 1997).

The primers of Dezfulian *et al*. (2003) were used to research the presence of potential GEIs at the *leuX, metV, selC, thrW* tRNA loci, while the primers of Germon *et al*. (2007) were used to research GEIs at the *asnT, asnUVW, aspV, glyU, pheU, pheV, serU, serX, serW* tRNA loci. For the *argU* and *argW* tRNA loci, GEIs were researched using the primers indicated in table 1 and designed from the *E. coli* K12 MG1655 genome.

To amplify the first nucleotides downstream from the interesting tRNA loci, arbitrary PCRs consisting of two steps were used. The first step corresponded to a PCR using genomic DNA of the strain as template, a first forward primer targetting the tRNA gene (argW1ter, asnT-F, aspV-F, leuX1, pheV3bis, serU3, and thrW1) (table 1), and the degenerated reverse primers ARB1, or ARB3, or ARB4 (table 1). The second step corresponded to a PCR using an aliquote (2 µL) of the products of

the first PCR as template, a second forward primer targetting a sequence of the same tRNA gene but located downstream from the region targetted with the first forward primer (argW-F, asnT3bis, aspV3, leuX3, pheV-F, serU-F, and thrW3), and the reverse primer ARB2 (table 1).

Nine pairs of primers were designed to cover the prophage like element SpLE5 localized downstream from the *leuX* tRNA locus in *E. coli* O157:H7 Sakaï genome (Accession number: BA000007.2) (table 1).

Six pairs of primers were designed to screen recombinant cosmids containing the six altered tRNA loci identified (*argW*, *asnT*, *leuX*, *pheV*, *serU*, and *thrW*). These primers were designed according to the nucleotide sequences of the region located downstream from the interesting loci and determined by arbitrary PCR (table 1). PCR products obtained using these pairs of primers and the *E. coli* P4 genomic DNA as template, were also used as probes in hybridization experiments.

Finally, the entire genomic island located downstream from the *argW* tRNA locus was amplified with the forward primer argW3bis, designed according to the first nucleotide sequence of the region located downstream from the *argW* locus which had been amplified by arbitrary PCR and sequenced, and with the reverse primer dsdA-R, designed from the complete genome sequence of *E. coli* K12 MG1655 (table 1).

2.5. PCR amplification

A thermocycler My Cycler BioRad was used for the different amplification procedures which are presented in table 2. All the PCRs were performed in a volume of 50 µL with about 200 ng of genomic DNA. Between 1 and 1.25 U of Taq DNA Polymerase (BioRad) were used for each PCR, except for the arbitrary PCRs where 5 U were used, for the screening of the *asnUVW*, *aspV*, *glyU*, *serX* tRNA loci and of the prophage like element SpLE5 PCRs where 1 U of Accu Taq DNA polymerase (Sigma) was used, and for the *argW* tRNA locus long PCR where 2.5 U of Accu Taq DNA polymerase (Sigma) were used. 200 µM of each deoxyribonucleoside triphosphate (BioRad) were used for each PCR, except for all the virulence genes PCRs where 100 µM of each dNTP were used, and for the *argW* tRNA locus long PCR where 500 µM of each dNTP were used. 1 µM of each primer was used for each PCR, except for the screening of the major sub-unit A of fimbrial adhesins F17a, F17b, F17c and F17d multiplex PCRs and for the screening of EaeA, SltI, SltII, EltIA, EltIIA, NeuA/C and StI simplex PCR where 0.6 µM were used, for the screening of fimbrial adhesins P and S, aerobactine, and CNF1 multiplex PCR and for the *argW* tRNA locus long PCR where 0.4 µM were used, for the screening of TraT, ClpG, and Cnf2 simplex PCR where 0.3 µM were used, for the Afa PCRs where 0.45 µM were used, and for the arbitrary PCRs where 0.8 µM were used.

2.6. DNA sequence analysis, sequence annotation, and nucleotide sequence accession numbers

DNA sequence were carried out by the Genome Express Society (Grenoble, France) according to the method of Sanger.

Homology researches as well as researches for conserved protein domains were performed using the BLAST programs of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/). Putative CDSs were identified using the CDS Finder program of the National Center for Biotechnology Information (http:///www.ncbi.nlm.nih.gov/gotf/gorf.htlm), and the Glimmer program of the Nano + Bio Center (http://www.nbz.uni-kl.de/). Conserved domains of putative CDSs were researched using the CDD program of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). RBS and rho-independent terminator were researched using the RBS Finder and Trans Term programs of the Nano + Bio Center (http://www.nbz.uni-kl.de/).

E. coli P4 specific sequences of the *cyaA*, *grpE*, *mtlD*, *elaD*, *pmbA* and *sms* genes, and of the *argW*, *asnT*, *leuX*, *pheV* (two sequences), *serU*, and *thrW* tRNA loci determined in this study were submitted to the EMBL nucleotide sequence database, and are available under accession numbers EU876896, EU876897, EU876898, EU876899, EU876900, EU876901, and EU876907, EU876902, EU876904, EU876905-EU876906, EU876908, and EU876903, respectively.

3. Results

3.1. *E. coli* P4 is phylogenetically closed to *E. coli* K12 MG1655

Ribotyping and multi locus enzyme electrophoresis (MLEE) analyses have shown that *E. coli* strains fall into four main phylogenetic groups (A, B1, B2, and D) (Herzer *et al.*, 1990 ; Selander *et al.*, 1987), and that pathogenic strains mainly belong to groups B2 and D, whereas most of the commensal ones belong to the groups A and B1. As ribotyping and MLEE tools are too complex and time-consuming, Clermont *et al.* (2000) proposed a simple and rapid phylogenetic grouping technique consisting in researching in the genome of an *E. coli* strain the presence or absence of the two genes *chuA* and *yjaA*, and of the anonymous DNA fragment tspE4-C2. Therefore, the *chuA* and *yjaA* genes as well as the tspE4-C2 DNA fragment were researched by PCR in *E. coli* P4 using *E. coli* K12 MG1655 as control. Only the *yjaA* gene was detected in *E. coli* P4 as in the control *E. coli* K12 MG1655, suggesting that *E. coli* P4 belonged to the commensal phylogenetic group A.

Other phylogenetic markers, which potentially would link *E. coli* P4 to other pathogenic *E. coli* genomes or at least which would distance it from commensal ones, were analyzed. In a first time, the

16S-23S rDNA intergenic spacer (16S-23S ITS), which allows to distinguish strains belonging to the same *E. coli* phylogenetic group (Anton *et al*., 1998) was amplified. Two amplicons of about 650 and 750 bp were obtained. These two amplicons corresponded to the two types of 16S-23S ITS of *E. coli*, i.e. to 16S-23S ITS containing only one or two tRNA genes. Only the amplicon of 650 bp was sequenced. It contains a 16S-23S ITS sequence of 354 bp, sharing 100 %, 98 % and 99 % identity with the 16S-23S ITS sequence of the rRNA operon *rrnE* of *E. coli* K12 MG1655 (phylogenetic group A) (Accession number: EU00096.2), *E. coli* O157:H7 Sakaï (phylogenetic group D) (Accession number: BA000007.2) and *E. coli* CFT073 (phylogenetic group B2) (Accession number: AE014075.1), respectively.

In a second time, a MLST study, which is commonly used for typing bacterial strains (Enright and Spratt, 1999), was employed to try to distinguish *E. coli* P4 from others strains of the phylogenetic group A. Three defined groups of genes were examined. The first enclosed 6 housekeeping genes (*adk*, *gcl, gdh, mdh, metA*, *ppk*) described by Adiri *et al*. (2003) for a multilocus sequence typing of the *E. coli* strains. An internal fragment of each of these genes was amplified and sequenced. In each case, the nucleotide sequence shared 100 % identity with the corresponding one of the commensal *E. coli* K12 MG1655 strain. The second group of genes which was examined enclosed 14 housekeeping genes (*aroE, aspC, clpX, cstA, cyaA, dnaG, fadD, grpE, icdA, lysP, mtlD, mutS, pgi, rpoS*) described by Reid *et al*. (2002) for a multilocus sequence typing of pathogenic *E. coli* strains. The nucleotide sequences of an internal fragment of 11 out of these 14 genes present 100 % identity with the corresponding ones of the commensal *E. coli* K12 MG1655 strain. For the 3 others (*cyaA, grpE, mtlD*), the sequences obtained were different from those of *E. coli* K12 MG1655, since they shared 99.8 %, 99.75 %, and 99.81 % identity, respectively, but remained closer to those of *E. coli* K12 MG1655 than to those of the other *E. coli* strains. Finally, the third group of genes enclosed 9 genes encoding putative proteases (*degQ, elaD, pmbA, ptrB, sms, yaeL, ycaL, yhjJ, yhjS*) which have been never used for *E. coli* MLST study until now, but which are shared by the different *E. coli* strains and which show nucleotide variations between the different *E. coli* strains. The nucleotide sequences of an internal fragment of 6 out of these 9 genes display 100 % identity with the corresponding ones of *E. coli* K12 MG1655. For the 3 others (*elaD*, *pmbA*, *sms*) the sequences obtained shared 99.75 %, 98.72 %, and 96.85 % identity, respectively, with the corresponding ones of *E. coli* K12 MG1655, and 97.69 %, 98.17 %, and 98.25 % identity, respectively, with the corresponding ones of *E. coli* O157:H7 Sakaï.

To conclude, these study revealed that the core genome of *E. coli* P4 is very similar to that of *E. coli* K12 MG1655, a commensal and non pathogen strain.

3.2. Virulence factors profile of *E. coli* P4

To attempt to explain the pathogenesis of *E. coli* P4, virulence factors known to be present in pathogenic *E. coli* strains isolated from calf or cow were firstly researched by PCR using primers already described in the litterature (Yamamoto *et al*., 1995; Bertin *et al*., 1996, 1998; China *et al*., 1996; Kuhnert *et al*., 1997; Lalioui *et al*., 1999; Kaipainen *et al*., 2002). On the 26 virulence genes examined, and in spite of using different PCR hybridization temperatures, only the gene *traT* was detected in the genome of *E. coli* P4, which encodes resistance to the killing action of serum. This result agrees with previous studies which often observed that mammopathogen *E. coli* strains do not possess the known virulence factors of calf and cow pathogenic intestinal and extra-intestinal *E. coli* strains (Kaipainen *et al*., 2002).

3.3. Detection of genomic islands in *E. coli* P4 genome by tRNA screening

Although *E. coli* P4 seemed to present a core genome close to that of *E. coli* K12 MG1655 and to possess none of virulence factor tested, it is quite able to grow in milk collected from an inflammed mammary gland, i.e. in antibacterial milk medium containing many immune components, whereas *E. coli* K12 MG1655 can not grow in this same medium (Dufour *et al*., submitted). Besides, published studies reported the ability of *E. coli* P4 to induce apoptosis of mammary cells (Long *et al*., 2001), and to adhere/invade mammary cells (Gonen *et al*., 2007). Such a behaviour might result from DNA regions encoding specific properties, i.e. from genomic islands. As tRNA genes are common sites for integration of foreign DNA elements (Hou, 1999), a PCR-based tRNA screening of the *E. coli* P4 genome was performed to detect sequence context alterations around the tRNA genes, using the *E. coli* K12 MG1655 genome as a control. Size differences between the PCR products obtained from the two genomes or the absence of the expected amplicon from the genome of *E. coli* P4, indicate a genomic variation and possibly a chromosomal integration of foreign DNA at the 3' end of the tested tRNA gene. Seventeen tRNA loci were screened: *argU, argW, asnT, asnUVW, aspV, glyU, leuX, metV, pheU, pheV, selC, serU, serW, serX*, and *thrW*. They corresponded to the 17 loci which are the most frequently associated with genomic islands (Germon *et al.*, 2007). Sequence context variations were detected in 7 (*argW, asnT, aspV, leuX, pheV, serU*, and *thrW*). For *argW, asnT, aspV, leuX, serU*, and *thrW* loci, no amplicon was obtained, while one amplicon of unexpected size (2208 bp, versus 623 bp for K12) was obtained for the *pheV* locus. Therefore, these results suggested the presence of genomic variation at the 3' end of these 7 tRNA loci in the *E. coli* P4 genome.

3.4. Sequence analysis of the region located downstream from the *argW*, *asnT*, *leuX*, *pheV*, *serU*, and *thrW* tRNA loci

The regions located immediately downstream from 6 (*argW*, *asnT*, *leuX*, *pheV*, *serU* and *thrW*) out of the 7 altered tRNA loci were amplified from the *E. coli* P4 genome by arbitrary PCR. Fragments of 747-bp, 202-bp, 1335-bp, 1677-bp, 755-bp, and 696-bp were obtained for the *argW*, *asnT*, *leuX*, *pheV*, *serU* and *thrW* loci, respectively, and sequenced. In the wake of this work, the sequences thus obtained are called ArgW-P4, AsnT-P4, LeuX-P4, PheV-P4, SerU-P4, and ThrW-P4 sequences, respectively.

(i) *argW* tRNA locus: the ArgW-P4 sequence had a G+C content of 43.11 %, and shared 94 % identity with a 627-bp fragment located downstream from a tRNA-Arg gene in an avian pathogenic *E. coli* (strain APEC BEN1588) (Accession number: DQ023206).

(ii) *asnT* tRNA locus: the AsnT-P4 sequence had a G+C content of 40.1 %, and displayed 96 % identity with a 195-bp fragment of *E. coli* C ATCC 8739 (Accession number: CP000946), located immediately downstream from two tRNA-Asn loci which probably correspond to *asnV* and *asnT* loci in *E. coli* K12 MG1655. Computer analysis did not reveal any ORF, but the corresponding fragment of *E. coli* C ATCC 8739 contain the first 36 nucleotides of a 1263-bp pseudogene presenting 82 % identity with an ORF encoding a putative P4-type integrase in *Shigella boydii* Sb 227 complete genome (Accession number: CP000036.1).

(iii) *leuX* tRNA locus: the LeuX-P4 sequence had a G+C content of 46.59 %. It contained one ORF of 975 nucleotides from the positions 91 to 1065, with a G+C content of 48.1 %. This ORF encoded a putative protein of 324 amino acids, with a deduced molecular mass of 36.51 kDa. Two predicted Shine-Dalgarno sites (AGGCG and AGGGT, respectively) were identified 5 bp and 15 bp upstream from the proposed start codon. The deduced amino acid sequence of this putative protein displayed 69 % identity and 76 % similarity with the first 336 amino acids of a putative integrase of 649 amino acids (Accession number: BAB38665) encoding by the ORF located immediately downstream from the *leuX* tRNA locus in *E. coli* O157:H7 Sakaï (figure 1). The ORF of *E. coli* P4 was truncated, and has lost the C-terminal catalytic domain (region 409-603). As the genomic context of *E. coli* P4 located downstream from the *leuX* tRNA locus was similar to that of *E. coli* O157:H7 Sakaï which contains the prophage like element SpLE5, nine pairs of primers specific to this prophage were designed and used in different PCR reactions to determine whether such an element was present in the genome of *E. coli* P4. Genomic DNAs of *E. coli* O157:H7 EDL933 and *E. coli* K12 MG1655 were used as positive and negative controls, respectively. No amplicon with the expected size was obtained.

They were either smaller than expected, as it was the case for products obtained using the primers leuX1/X2, leuX3/X4, leuX5/X6, leuX13/X14, leuX15/X16, and leuX17/X18, or absent, as it was the case using the primers leuX7/X8, leuX9/10, and leuX11/X12. Thus, it appeared that an altered structure of the prophage like element SpLE5 is present in the *E. coli* P4 genome.

(iv) *pheV* tRNA locus: the PheV-P4 sequence had a G+C content of 46.57 %. It contained one ORF of 1263 nucleotides from the positions 198 to 1460, with a G+C content of 47.70 %. This ORF encoded a putative protein of 420 amino acids, with a deduced molecular mass of 47.99 kDa. No predicted Shine-Dalgarno site with *E. coli* consensus sequence could be identified upstream from the proposed start codon. The deduced amino acids sequence of this putative protein displayed 90 % identity and 95 % similarity with a putative P4-type integrase of 420 amino acids (Accession number: AAN44445) identified in *Shigella flexneri* 2a strain 301 genome located downstream from a phe-tRNA locus. Thus, *E. coli* P4 ORF conserved an entire domain of bacteriophage P4 integrase, and seemed to encode a functional protein.

The 2208-bp PCR product which had been obtained during the *pheV*-tRNA screening (see above) was also sequenced. This sequence had a G+C content of 49.60 % and displayed some partial identities (> 80 %) with different genomic fragments of the LEE-PAI of the bovine shiga toxin-producing *E. coli* RW1473 strain (Accession number: AJ303141). Its last 333 nucleotides also displayed 100 % identity with the first 333 nucleotides of the *yghD* gene of *E. coli* K12 MG1655 genome which is the first conserved gene found in the *E. coli* genomes downstream from the *pheV* tRNA locus. From the positions 1705 to 1727, a 23-bp sequence that was 100 % identical to the last 23 nucleotides of *pheV* gene, was identified. Consequently, this 2208-bp region seems to represent the 3' end of the genomic island located downstream from the *pheV* locus in the *E. coli* P4 genome. Computer analysis revealed the presence of three ORFs (ORF1-*pheV*, 2-*pheV*, and 3-*pheV*, table 3). They encoded hypothetical proteins similar to proteins with unknown function of the commensal strain *E. coli* Nissle 1917, the enteropathogen strain *E. coli* B7A, and the bovine enterohemorragic strain *E. coli* RW1374, respectively. Thus, the region located downstream the *pheV*-tRNA locus in the *E. coli* P4 genome seemed to have a composite structure.

(v) *serU* tRNA locus: the SerU-P4 sequence had a G+C content of 44.11 %. It displayed 79 % identity with a fragment of 716-bp located immediately downstream from a Ser-tRNA locus in the *E. coli* HS genome (Accession number: CP000802).

(vi) *thrW* tRNA locus: the ThrW-P4 sequence had a G+C content of 47.56 %, and presented 98 % identity with a 696-bp fragment of *Shigella flexneri* 2a strain 301 genome located 15640 nucleotides downstream a *thr*-tRNA locus, and with a 696-pb fragment of *E. coli* O157:H7 Sakaï genome located

immediately downstream the *thrW*-tRNA locus. In *S. flexneri* 2a 301 and *E. coli* O157:H7 Sakaï, the corresponding sequences respectively contain the last moiety and the last 2/3 of an ORF encoding a putative integrase. The primary structures of the putative integrases of the two strains share 82.5 % identity, and the nucleotidic sequences located around the two integrase genes are quite different.

(vii) Conclusion: the analyses of the regions located immediately downstream from the *argW*, *asnT, leuX, pheV, serU*, and *thrW* loci in the *E. coli* P4 genome showed in each case the presence of nucleotide sequences different from those present at the corresponding loci in the *E. coli* K12 MG1655 genome. For the *pheV* locus, the analysis of this sequence strongly suggested the presence of a GEI at this locus in *E. coli* P4, whereas no foreign DNA is inserted at this locus in *E. coli* K12 MG1655 genome. For the *leuX* locus, an altered structure of *E. coli* O157:H7 Sakaï SpLE5 seemed to be present in *E. coli* P4 genome, whereas an other prophage like element (KpLE2) with different genomic content is identified at this locus in *E. coli* K12 MG1655 genome. Finally, for the *argW*, *asnT*, *serU*, and *thrW* loci, the nucleotide sequences available for *E. coli* P4 genome were quite different from those of the corresponding loci in the *E. coli* K12 MG1655 genome. Indeed, at the *argW* and *thrW* loci of this strain, a prophage like element and a P4 cryptic prophage are present, respectively, whereas no insertion of foreign DNA are detected in the *asnT* and *serU* loci.

3.5. Partial sequence analysis of the genomic island located downstream from the *serU* tRNA locus in *E. coli* P4 genome

In order to characterize the regions located downstream from the identified altered tRNA loci in *E. coli* P4, a recombinant cosmid library was constructed with genomic DNA of this strain. Recombinant cosmids containing the regions of interest were selected by PCR and reconfirmed by hybridization, using appropriate PCR products specific to the targetted regions, as probes. One recombinant cosmid containing the *serU* locus was identified and partially sequenced. A nucleotide sequence of a 13,202-bp region downstream from the *serU* locus was determined. This sequence presented a G+C content of 42.5 %. Seven ORFs (table 4 and figure 2) were detected. Four of them (ORF3-*serU*, 5-*serU*, 6-*serU*, and 7-*serU*) encoded putative proteins with mobility function, like transposase and integrase. The ORF1-*serU* encoded a protein which was related to glycosylation function of the O-antigen part of lipopolysaccharides, the ORF4-*serU* encoded a protein of unknown function, and the protein encoded by the ORF2-*serU* presented no significant similarity with any protein of the databases. Proteins similar to the proteins encoded by the 7 ORFs-*serU* of *E. coli* P4 have different origin such as bacteriophage DNA (ORF1-*serU*, and 7-*serU*), bacterial chromosome DNA (ORF3-*serU*, and 6-*serU*), and bacterial plasmid (ORF4-*serU*, and 5-*serU*). Thus, this partial sequence suggested that the *serU* region of *E. coli* P4 genome presented a chimerical structure.

3.6. Sequence analysis of the genomic island located downstream from the *argW* tRNA locus in *E. coli* P4 genome

(i) General features: as the ArgW-P4 sequence presented no homology with any sequence of the sequenced *E. coli* genomes, and that no information was available about the similar sequence of *E. coli* APEC BEN 1588, we decided to sequence the entire variable genomic region located downstream from the *argW* locus. A forward primer specific to the ArgW-P4 sequence, and a reverse primer specific to the sequence of the *dsdA* gene (first conserved gene found in the *E. coli* genomes downstream from the *argW* tRNA locus) were designed, and used to carry out long range PCR. A PCR product of about 12 Kb was obtained, and sequenced. After removing the nucleotide fragments corresponding to the *argW* and *dsdA* genes, a genomic region of 11,253 bp was defined, with a G+C content of 40.92 %. This G+C content shows high variability across the genomic region (table 5). It appeared to be very different from the *argW* genomic region of *E. coli* K12 MG1655 as well as its complete genome (44.67 % and 50.80 % respectively). From the positions 8,628 to 8,650 of the genomic region, a 23-bp sequence that was 100 % identical to the last 23 nucleotides of *argW* gene, was identified. Thus, this genomic region exhibited several features which are typical of GEIs: chromosomal integration site downstream from a tRNA locus, direct repeat of the 3' end of this tRNA locus, large size (>10,000 bp), and atypical G+C content.

(ii) ORFs content: eight ORFs were detected in this 11,253 nucleotides sequence (table 5 and figure 3). The ORF1-*argW* and ORF6-*argW* encoded proteins with mobility function; the ORF2-*argW*, ORF3-*argW*, and ORF7-*argW* encoded proteins with replication / transcription functions; the ORF8-*argW* encoded a protein with transport function; the ORF5-*argW* encoded a protein of unknown function; and the ORF4-*argW* encoded a protein with deacetylation function. Genomic regions between all these ORFs presented no similarity to any nucleotide sequence in the databases.

The deduced amino acids sequence of the putative protein encoded by the ORF1-*argW* presented 68 % identity and 82 % similarity with the amino acids sequence of an integrase of the *Salmonella enterica* ATCC 9150 strain (figure 4). Only the first moieties of the amino acids sequence of the two proteins are similar, showing that putative protein encoded by the ORF1-*argW* conserved only the N-terminal domain of integrase, i.e. the DNA-binding domain, and lost the catalytic C-terminal domain. The C-terminal domain of the ORF1-*argW* presented no significant degree of similarity with any protein sequence of the database. Consequently, the protein encoded by the ORF1-*argW* could be considered as a chimerical protein, resulting from the association of a N-terminal domain of a tyrosine recombinase with a fragment of unknown origin. The two last ORFs detected presented 100 % identity with the two ORFs located upstream from the *dsdA* gene in *E. coli* K12 MG1655 genome.

(iii) Sequence analysis of the ORF4-*argW*: the ORF4-*argW* is the largest ORF detected in the *argW* genomic island of *E. coli* P4. It consisted of 2,481 nucleotides, and presented a G+C content of 31.20 %, which was significantly lower than the average in *E. coli* genome. A predicted Shine-Dalgarno site (AAGAG) was identified 11-bp upstream from the proposed start codon. A possible rho-independent transcriptional termination signal was also identified 14-bp downstream from the proposed termination codon. This ORF4-*argW* encoded a putative protein of 826 amino acids, with a deduced molecular mass of 95.18 kDa. From the position 27 to 223 of the amino acids sequence of this putative protein, a conserved domain of Silent Information Regulator 2 (Sir2) was detected. Sir2 domain is characteristic of enzymes which catalyze NAD^+-dependent protein/histone deacetylation. Such proteins have been identified in organisms ranging from bacteria to humans. They contain a series of consensus motifs grouped in a region of about 200 amino acids. The position of this region is highly variable in the total sequence of the Sir2 proteins, and amino acids sequences around this region can be highly variable too. Prokaryotic Sir2 proteins lack certain consensus motifs, as it is the case for the protein encoded by the ORF4-*argW* of *E. coli* P4, in which the consensus motifs GAGISXXXGIPXXR, TQNID, GTS, and (I/V)N were effectively identified with an altered composition (GAGFSRKTKNITNGEPPLA, TTNYD, GYS, and IV, respectively) (figure 5), and the consensus motifs PXXXH, HG, two sets of CXXC, and FGE were not identified.

4. Discussion

Although mastitis caused by bacteria and most particularly by *E. coli* represent a worldwide problem in dairy cattle industry, leading to multibillion dollar losses (Sordillo and Streicher, 2002), the MPEC strains remain poorly characterized and the study presented here is, at our knowledge, the first one dealing with the molecular typing of a bovine mammopathogen strain and the identification of GEIs.

The phylogenetic studies we carried out showed that the bovine MPEC strain *E. coli* P4 belonged to the phylogenetic group A of the *E. coli* species, as the famous commensal *E. coli* K12 MG1655 strain. The phylogenetic group A, as the group B1, mainly contains commensal *E. coli* isolates, whereas the groups B2 and D get together extra-intestinal pathogenic *E. coli* isolates (Picard *et al*., 1999). By using the 16S-23S ITS of the rRNA operon *rrnE*, or by establishing partial nucleotide sequences of various genes, we showed that the core genome of *E. coli* P4 was close to that of *E. coli* K12 MG1655, since the 16S-23S ITS used as well as 23 out of 29 genes studied shared 100 % identity with the corresponding ones of *E. coli* K12 MG1655, and 5 out of the 6 other genes studied showed the higher percentage of identity with the corresponding ones of *E. coli* K12 MG1655.

However, in spite of their similar core genomes, *E. coli* P4 and *E. coli* K12 MG1655 exhibit opposite behaviours when grown in a milk reproducing the enflamed environment of a bovine

mammary gland during a mastitis, since the first one was able to grow rapidly whereas the second one was not and even did not survive in this high antibacterial environment (Dufour *et al*., submitted). Such observations suggest that *E. coli* P4 might possess specific properties.

Attempts to identify some specific virulence factors in *E. coli* P4 by PCR led to the only detection of the TraT factor, which is involved in the serum resistance (Montenegro *et* al., 1985). This component is the most frequently identified in the *E. coli* strains inducing bovine mastitis (Kaipainen *et al*, 2002), and it can be responsible for the survival and the growth of *E. coli* P4 in the milk miming inflammation, by protecting it from the action of the phagocytes. All other virulence factors tested, in particular the 12 factors which are known to be involved in the cell adhesion process, appeared to be absent. Though, recent studies have demonstrated that *E. coli* P4 strain is able to adhere, invade the mammary gland epithelial cells (Gonen *et al*., 2007), and induce cell death via apoptosis in lactating mammary gland (Long *et al*., 2001). Such results suggest that MPEC, and most particularly the strain *E. coli* P4, might exhibit specific properties which have been not yet identified.

It is well know that *E. coli* is a highly adaptative organism, and that this species contains both commensal and pathogenic strains which mainly differ by their GEIs, acquired by horizontal genetic transfers, and which confer them specific properties (Dozois and Curtiss, 1999). In most cases, GEIs are located downstream from the genes encoding tRNA (Hacker and Kaper, 2000). Over the 85 tRNA loci which are present in the sequenced *E. coli* genomes, and after elimination of the unlikely acceptor loci (polycistronic tRNA or tRNA present in ribosomial operon), we screened 17 out of 36 likely acceptor tRNA, and found that 7 of them displayed genomic context different from those present in the *E. coli* K12 MG1655 genome. These altered loci are *argW, asnT, aspV, leuX, pheV, serU*, and *thrW*, and correspond to loci where genomic islands are frequently detected. For example, pathogenicity islands (PAIs) II, III, IV, and V of the uropathogenic *E. coli* strain 536 are localized at *leuX, thrW, asnT*, and *pheV* loci, respectively (Oelschlaeger *et al*., 2002); Vat-PAI of the avian pathogenic *E. coli* Ec222 strain is localized at *thrW* locus (Parreira and Gyles, 2003); LEE-PAI of the bovine shiga toxin-producing *E. coli* RW1473 strain is localized at *pheV* locus (Jores *et al*., 2001); and GEIs II, III, and IV of the non-pathogenic probiotic *E. coli* strain *Nissle* 1917 are localized at *pheV, argW*, and *asnT* loci, respectively (Grozdanov *et al*., 2004).

Partial analysis of five of these altered tRNA loci showed that the regions located immediatly downstream the tRNA encoding gene contain ORF which appeared to be related to an integrase, one of the typical features of a GEI. Such integrase appeared to be functional as it was the case for the *serU* and *pheV* loci, or not as at the *leuX* locus. In spite of the high similarities observed between these integrases and integrases present in other *E. coli* strains, this does not imply that the regions located downstream are also similar. Indeed, similar integrase encoding genes are frequently located downstream a tRNA locus in different *E. coli* strains, though the DNA contents of the corresponding

islands are not identical. The partial characterization of the regions located downstream from the *serU* and *pheV* loci revealed the presence of ORFs with very different origin.

The complete characterization of the GEI located at *argW* locus in *E. coli* P4 revealed several features which are typical of GEIs. Its borders are the *argW* tRNA gene at one end, and the *dsdA* conserved gene at the other; its size exceeds 10,000-bp; it contains two ORFs encoding mobile genetic elements (one partially conserved integrase, and one putative resolvase); and its G+C content (40.92 %) is very different from that of the complete *E. coli* genomes (about 50 %). Analysis of the sequence of this GEI reveals the presence of genes encoding proteins involved in four cellular functions: replication, transcription, transport, and regulation. The ORF4-*argW* is possibly involved in the regulation of gene expression since it encodes a putative protein which contains a Sir2-like conserved domain, i.e. a putative protein which catalyzes NAD^+ dependent protein/histone deacetylation. Acetylation / deacetylation is one of the important regulation processes of cellular metabolism (Kouzarides, 2000). The Sir2 proteins, also known as sirtuins, are found in all eukaryotes and many archaea and prokaryotes, and have been shown to regulate gene silencing, DNA repair, metabolic enzymes, and life span (North and Verdin, 2004). Although sirtuins of yeast and human are intensively studied, the bacterial ones remain unknown. Considering the different role which are played by these proteins, it would be interesting to determine whether the protein encoded by the ORF4-*argW* of the *argW* GEI of *E. coli* P4 could be involved in a transcriptional silencing, a phenomenon which has been shown to play a critical role in bacterial virulence (Dagberg and Uhlin, 1992). Such a phenomenon could allow *E. coli* P4 to survive in the mammary gland and cause persistent infection without alerting the immun system.

Aknowledgments

This work was supported by "Le Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche".

References

Adiri, R.S., Gophna, U., Ron, E.Z., 2003. Multilocus sequence typing (MLST) of *Escherichia coli* O78 strains. FEMS Microbiol. Lett. 222, 199-203.

Anton, A.I., Martinez-Murcia, A.J., Rodriguez-Valera, F., 1998. Sequence diversity in the 16S-23S intergenic spacer region (ISR) of the rRNA operons in representatives of the Escherichia coli ECOR collection. J. Mol. Evol. 47, 62-72.

Bertin, Y., Martin, C., Girardeau, J.P., Pohl, P., Contrepois, M., 1998. Association of genes encoding P fimbriae, CS31A antigen and EAST 1 toxin among CNF1-producing *Escherichia coli* strains from cattle with septicaemia and diarrhea. FEMS Microbiol Lett. 162, 235-239.

Bertin, Y., Martin, C., Oswald, E., Girardeau, J.P., 1996. Rapid and specific detection of F17-related pilin and adhesion genes in diarrheic and septicemic *Escherichia coli* strains by multiplex PCR. J. Clin. Microbio. 34, 2921-2928.

Bradley, A.J., 2002. Bovine mastitis: an evolving disease. Vet. J. 164, 116-128.

Bradley, A.J., Green, M.J., 2001. Adaptation of *Escherichia coli* to the bovine mammary gland. J. Clin. Microbiol. 39, 1845-1849.

Bradley, A.J., Green, M.J., 2000. A study of the incidence and significance of intramammary enterobacterial infections acquired during the dry period. J. Dairy Sci. 83, 1957-1965.

Bramley, A.J., 1976. Variations in the susceptibility of lactating and non-lactating bovine udders to infection when infused with *Escherichia coli*. J. Dairy Res. 43, 205-211.

Burvenich, C., Van Merris, V., Merhzad, J., Diez-Fraile, A., Duchateau, L., 2003. Severity of *E. coli* mastitis is mainly determined by cow factors. Vet. Res. 34, 521-564.

China, B., Pirson, V., Mainil, J., 1996. Typing of bovine attaching and effacing *Escherichia coli* by multiplex in vitro amplification of virulence-associated genes. Appl. Environ. Microbiol. 62, 3462-3465.

Clermont, O., Bonacorsi, S., Bingen, E., 2000. Rapid and simple determination of the *Escherichia coli* phylogenetic group. Appl. Environ. Microbiol. 66, 4555-4558.

Dagberg, B., Uhlin, B.E., 1992. Regulation of virulence associated plasmid genes in enteroinvasive *Escherichia coli*. J. Bac. 174, 7606-7612.

Demuyter, P., Leblond, P., Decaris, B., Simonet, J.M., 1988. Characterization of two families of spontaneously amplifiable units of DNA in *Streptomyces ambofaciens*. J. Gen. Microbiol. 134, 2001-2007.

Dezfulian, H., Batisson, I., Fairbrother, J.M., Lau, P.C.K., Nassar, A., Szatmari, G., Harel, J., 2003. Presence and characterization of extraintestinal pathogenic *Escherichia coli* virulence genes in F165-positive *E. coli* strains isolated from diseased calves and pigs. J. Clin. Microbiol. 41, 1375-1385.

Dobrindt, U., Hochhut, B., Hentschel, U., Hacker, J., 2004. Genomic islands in pathogenic and environmental microorganisms. Nat. Rev. Microbiol. 2, 414-424.

Dogan, B., Klaessig, S., Rishniw, M., Almeida, R.A., Oliver, S.P., Simpson, K., Schukken, Y.H., 2006. Adherent and invasive *Escherichia coli* are associated with persistent bovine mastitis. Vet. Microbiol. 116, 270-282.

Döpfer, D., Nederbragt, H., Almeida, R.A., Gaastra, W., 2001. Studies about the mechanism of internalization by mammary epithelial cells of *Escherichia coli* isolated from persistent bovine mastitis. Vet. Microbiol. 80, 285-296.

Döpfer, D., Almeida, R.A., Lam, T.J.G.M., Nederbragt, H., Oliver, S.P., Gaastra, W., 2000. Adhesion and invasion of *Escherichia coli* from single and recurrent clinical cases of bovine mastitis in vitro. Vet. Microbiol. 74, 331-343.

Dozois, C.M., Curtiss, R., 1999. Pathogenic diversity of *Escherichia coli* and the emergence of exotic islands in the gene stream. Vet. Res. 30, 157-179.

Enright, M.C., Spratt, B.G., 1999. Multilocus sequence typing. Trends Microbiol. 7, 482-487.

Garcia-Martinez, J., Martinez-Murcia, A.J., Rodriguez-Valera, F., Zorraquino, A., 1996. Molecular evidence supporting the existence of two major groups in uropathogenic *Escherichia coli*. FEMS Immunol. Med. Microbiol. 14, 231-244.

Germon, P., Roche, D., Melo, S., Mignon-Grasteau, S., Dobrindt, U., Hacker, J., Schouler, C., Moulin-Schouleur, M., 2007. tDNA loci polymorphism and ecto-chromosomal DNA insertion hot-spots are related to the phylogenetic group of *E. coli* strains. Microbiology 153, 826-837.

Gonen, E., Vallon-Eberhard, A., Elazar, S., Harmelin, A., Brenner, O., Rosenshine, I., Jung, S., Shpigel, N.Y., 2007. Toll-like receptor 4 is needed to restrict the invasion of *Escherichia coli* P4 into mammary gland epithelial cells in a murine model of acute mastitis. Cell Microbiol. 9, 2826-2838.

Grozdanov, L., Raasch, C., Schulze, J., Sonnenborn, U., Gottschalk, G., Hacker, J., Dobrindt, U., 2004. Analysis of the genome structure of the non-pathogenic probiotic *Escherichia coli* strain Nissle 1917. J. Bacteriol. 186, 5432-5441.

Hacker, J., Kaper, J.B., 2000. Pathogenicity islands and the evolution of microbes. Annu. Rev. Microbiol. 54, 641-679.

Hejnova, J., Pages, D., Rusniok, C., Glaser, P., Sebo, P., Buchrieser, C., 2006. Specific regions of genome plasticity and genetic diversity of the commensal *Escherichia coli* A0 34/86. Int. J. Med. Microbiol. 296, 541-546.

Herzer, P.J., Inouye, S., Inouye, M., Whittam, T.S., 1990. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. J. Bacteriol. 172, 6175-6181.

Hou, Y.M., 1999. Transfer RNAs and pathogenicity islands. Trends Biochem. 24, 295-298.

Johnson, J.R., Russo, T.A., 2002. Extraintestinal pathogenic *Escherichia coli*: "The other bad *E. coli*". J. Lab. Clin. Med. 139, 155-162.

Jores, J., Rumer, L., Kiebling, S., Kaper, J.B., Wieler, L.H., 2001. A novel locus of enterocyte effacement (LEE) pathogenicity island inserted at *phe*V in bovine shiga toxin-producing *Escherichia coli* O103:H2. FEMS Microbiol. Lett. 204, 75-79.

Kaipainen, T., Pohjanvirta, T., Shpigel, N.Y., Shwimmer, A., Pyörälä, S., Pelkonen, S., 2002. Virulence factors of *Escherichia coli* isolated from bovine clinical mastitis. Vet. Microbiol. 85, 37-46.

Kouzarides, T., 2000. Acetylation: a regulatory modification to rival phosphorylation? EMBO J. 19, 1176-1179.

Kuhnert, P., Hacker, J., Mühldorfer, I., Burnens, A.P., Nicolet, J., Frey, J., 1997. Detection system for *Escherichia coli*-specific virulence genes: absence of virulence determinants in B and C strains. Appl. Environ. Microbiol. 63, 703-709.

Lalioui, L., Jouve, M., Gounon, P., Le Bouguenec, C., 1999. Molecular cloning and characterization of the afa-7 and afa-8 gene clusters encoding afimbrial adhesins in *Escherichia coli* strains associayed with diarrhea or septicaemia in calves. Infect. Immun. 67, 5048-5059.

Leblond, P., Fischer, G., Francou, F.X., Berger, F., Guérineau, M., Decaris B., 1996. The unstable region of *Streptomyces ambofaciens* includes 210 kb terminal inverted repeats flanking the extremities of the linear chromosomal DNA. Mol. Microbiol. 19, 261-271.

Long, E., Capuco, A.V., Wood, D.L., Sonstegard, T., Tomita, G., Paape, M.J., Zhao, X., 2001. *Escherichia coli* induces apoptosis and proliferation of mammary cells. Cell Death Differ. 8, 808-816.

Menzies, F. D., Gordon, A. W., Mc Bride, H. Mc., 2003. An epidemiological study of bovine toxic mastitis. Proceedings of the British Mastitis Conference, Garstang, 1-13.

Michelutti, I., Le Roux, Y., Rainard, P., Poutrel, B. Laurent, F., 1999. Sequential changes in milk protein composition after experimental *Escherichia coli* mastitis. Lait 79, 535-549.

Montenegro, M.A., Bitter-Suermann, D., Timmis, J.K., Agüero, M.E., Cabello, F.C., Sanyal, S.C., Timmis, K.N., 1985. *tra*T gene sequences, serum resistance and pathogenicity-related factors in clinical isolates of Escherichia coli and other gram-negative bacteria. J. Gen. Microbiol. 131, 1511-21.

North, B.J., Verdin, E., 2004. Sirtuins: Sir2-related NAD-dependent protein deacetylases. Genome Biol. 5, 224.

Oelschlaeger, T.A., Dobrindt, U., Hacker, J., 2002. Pathogenicity islands of uropathogenic *E. coli* and the evolution of virulence. Int. J. Antimicrob. Agents 19, 517-521.

Parreira, V.R., Gyles, C.L., 2003. A novel pathogenicity island integrated adjacent to the *thr*W tRNA gene of avian pathogenic *Escherichia coli* encodes a vacuolating autotransporter toxin. Infect. Immun. 71, 5087-5096.

Picard, B., Garcia, J.S., Gouriou, S., Duriez, P., Brahimi, N., Bingen, E., Elion, J., Denamur, E., 1999. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. Infect. Immun. 67, 546-553.

Reid, S.D., Herbelin, C., Bumbaugh, A.C., Selander, R.K., Whittam, T.S., 2002. Multilocus sequence typing (MLST) of pathogenic *Escherichia coli*. http://www.shigatox.net/stec/mlst-new/

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular cloning: a laboratory manual, second ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Seegers, H., Menard, J.L., Fournichon, C., 1997. Mammites en élevage bovin laitier : importance actuelle, épidémiologie et plans de prévention. Renc. Rech. Ruminants 4, 233-242.

Selander, R.K., Caugant, D.A., Whittam, T.S., 1987. Genetic structure and variation in natural populations of *Escherichia coli*. In Neidhart F. C., Ingraham J. L., Low K. B., Magasanik B., Schaechter M. and Umbarger H. E. (ed), *Escherichia coli* and *Salmonella typhimurium* : cellular and molecular biology, American Society for Microbiology, Washington, D.C., pp.1625-1648.

Sordillo, L.M., Streicher, K.L., 2002. Mammary gland immunity and mastitis susceptibility. J. Mammary Gland Biol. Neoplasia 7, 135-146.

Sordillo, L.M., Shafer-Weaver, K., DeRosa, D., 1997. Immunology of the mammary gland. J.Dairy Sci. 80, 1851-1865.

Wenz, J.R., Barrington, G.M., Garry, F.B., McSweeney, K.D., Dinsmore, R.P., Goodell, G., Callan, R.J., 2001. Bacteremia associated with naturally occurring acute coliform mastitis in dairy cows. JAVMA 219, 976-981.

Yamamoto, S., Terai, A., Yuri, K., Kurazono, H., Takeda, Y., Yoshida, O., 1995. Detection of urovirulence factors in *Escherichia coli* by multiplex polymerase chain reaction. FEMS Immunol. Med. Microbiol. 12, 85-90.

Figure 1: Comparaison of the predicted amino acids sequences of the proteins encoded by the ORFs located immediately downstream from the *leuX* tRNA locus in *E. coli* P4 and *E. coli* O157:H7 Sakaï genomes. Identical residues are shaded in dark, similar ones in grey, and different ones are in white background. The double arrow indicates the C-terminal catalytic domain of *E. coli* O157:H7 Sakaï integrase.

Figure 2: Partial genetic map of the *serU* genomic region of *E. coli* P4. Numbers indicate ORFs described in table 4.

Figure 3: Genetic map of the *argW* genomic region of *E. coli* P4. Numbers indicate ORFs described in table 5.

Figure 4: Comparaison of the predicted amino acids sequences of the protein encoded by the ORF1-*argW* of *E. coli* P4 and the integrase of *Salmonella enterica* ATCC9150. Identical residues are shaded in dark, similar ones in grey, and different ones are in white background. The double arrow indicates the complete Bacteriophage P4 integrase domain.

Figure 5: Predicted amino acids sequence of the protein encoded by the ORF4-*argW* of *E. coli* P4. The double arrows indicate some of the conserved consensus motifs specific to the Sir2 proteins.

Figure 1

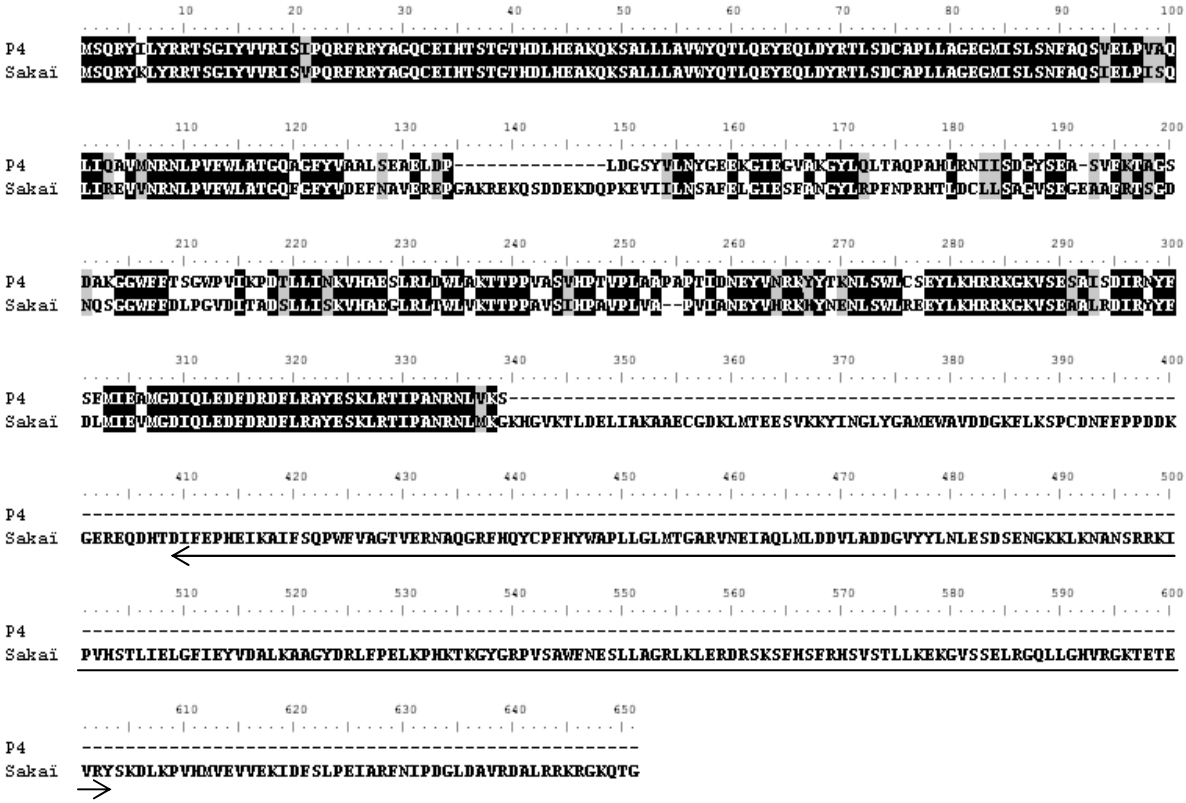


Figure 2

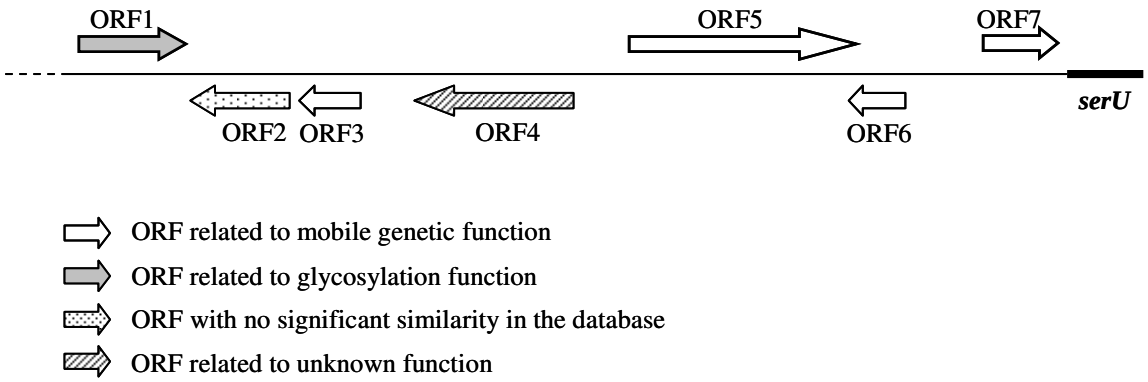
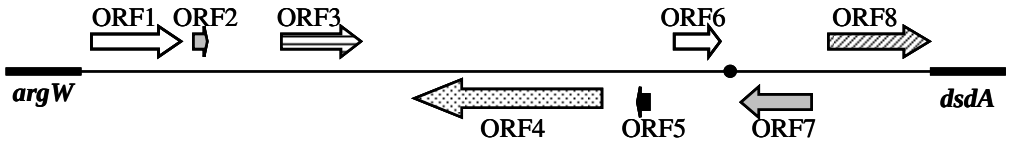


Figure 3



ORF related to mobile genetic function

ORF related to transcription DNA function

ORF related to replication DNA function

ORF related to deacetylation function

ORF related to transporter function

ORF related to unknown function

● Direct repeat of the last 23 nucleotides of the *argW* gene

Figure 4

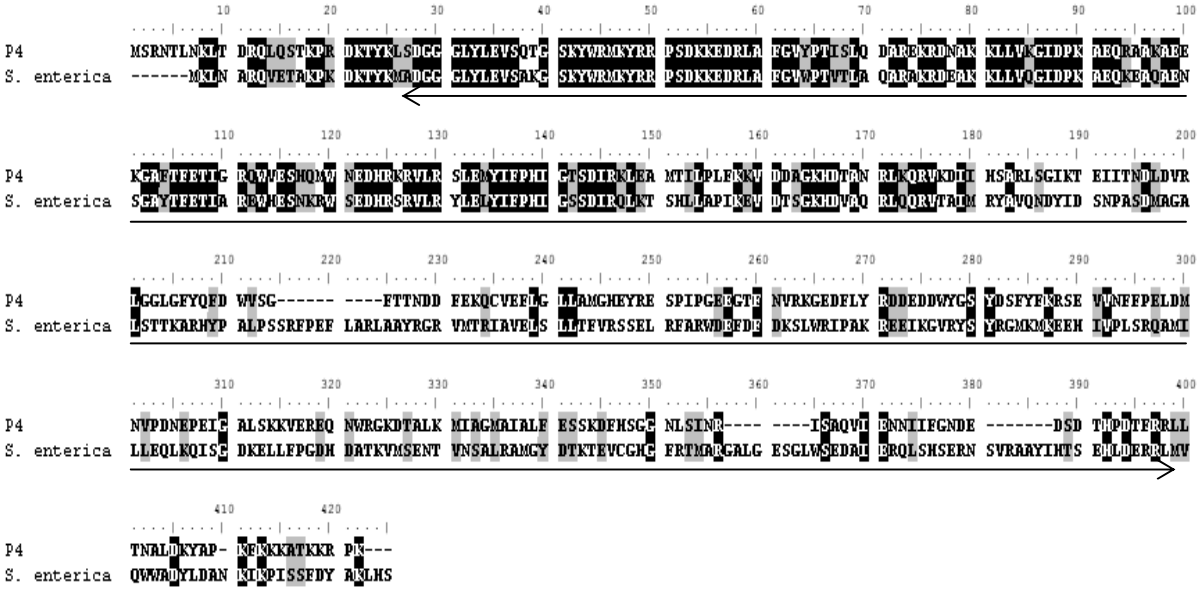


Figure 5

10 20 30 40 50 60 70 80 90 100
MFNKMDVTKL NLPKVDIEEM LRRISSGKAI LFTGAGFSRK TKNITNGEPP LAKALSKKIS QLAGLGSDNE DLMFTSRFFL KHGNKSELLK LLKDNFVLRE
110 120 130 140 150 160 170 180 190 200
VSEYHEKICS LPWRRFYTTN YDNSIELACL KTGRRIESLD IDDLPKDHIS QGDLCVHLNG AIDKAIISDL DSKIKLTNSS YLSPQSFISS QWNYVFKRDL
210 220 230 240 250 260 270 280 290 300
ETASAVIFVG YSMYDMDVQR LLYQTDSLVE KTYFIVHENA SFQETFFLTD FGHVLPIGVE GFSNLINDIQ FQSNENREEI CLDCFELKEI SYNHETITDT
310 320 330 340 350 360 370 380 390 400
EIKDFLLFGK HDDNQINTSV SNDFLDNFLI NRDLLKETIR LIQSKNNILI HSELGNGKTV FLKMITYLLA REGYNVYTFS EKSEYDDELS EIDWIVKNKK
410 420 430 440 450 460 470 480 490 500
NVVIVIEGYN KAERLLNHIN INYPDEISII ITDRSAIALR TAYFINSLDI EFSEVSLDQL TESEISDFVD LIENQGLWSE LTSLSKINKI KKVKEDYNGQ
510 520 530 540 550 560 570 580 590 600
ISGILLGLLK SPSIQERIKY LTDELFKSQE YKDTVFAIAL CDIIDVRKTS SIISEMAGND SIYKMSLRSS DQFKSLYRFT ESGTAIETKS SLMSLAIINN
610 620 630 640 650 660 670 680 690 700
SFNVSYVRHK LLSIVETFNN LKNVSYDANK IFKSLLRFHV LEVLLPQKQK ALDSYYMELK RVCPWLTDSP MYWVQYAMCR LSIGDLDSAQ TYLNDAYSLA
710 720 730 740 750 760 770 780 790 800
RSKEDYHTEN IDTQQARLLI MKCLKEPDSG KCFDSFIEAH NILIGLPDDG YKYRQIIPYK DVYELKYKHF NKGNKVQFEH ACKALLAQVS SVDYHQDDLT
810 820
LIKRVSFINR SKVILEEIVK SITSNR

Table 1: New PCR primers used in this study (*F : forward, R : reverse).

	Target gene	Primer*	Primer sequence (5'→3')	Annealing temperature (°C)	PCR product size (bp)
MLST simplex PCRs	*deg*Q	degQ-F	ATCAACCTGCACAACCCTTC	58	650
		degQ-R	CAACACTTCGCTGACAAACG		
	*yae*L	yaeL-F	GTCCGGTTGCAAACTTCATT	58	657
		yaeL-R	GCCCATACTGGCGTACAACTT		
	sms	sms-F	CTGCCGACTGACAATCTCAA	58	652
		sms-R	CCACGTTCACAAACACATCC		
	*yhj*J	yhjJ-F	CGGCAGTCAAATCCAAGACC	56	647
		yhjJ-R	GGTAATTGTCTCTTATGA		
	*ela*D	elaD-F	TTCAGAAGATCGCCGCTAAT	56	676
		elaD-R	TTGCCGGATGCTCTATTTTT		
	*pmb*A	pmbA-F	ATCACCAATACCGAAGGTGG	60	646
		pmbA-R	CCCGGTGCCCATCTCTTTGAG		
	*ptr*B	ptrB-F	CCAGTTTGCGCCAAATTAAC	58	648
		ptrB-R	CGATAACGCCGTGGAATAAT		
	ycaL	ycaL-F	GCGACCTCTGCAGCATTACT	60	644
		ycaL-R	TCCAGGCTAGCCAGTGTTTC		
	*yhj*S	yhjS-F	GAAGAGGCGGAGATCCAAC	58	648
		yhjS-R	AAACAGCACCAGCCGATTAC		
tRNA screening simplex PCRs	*arg*U	argU-F	CGCCCTTAGCTCAGTTGGA	55	483
		argU-R	TGTTGCGTGATCAAATAGGC		
	*arg*W	argW-F	TAGTTAAATGGATATAACGAGC	55	600
		argW-R	TTTTCGCCTTTACAGGATCG		
*arg*W locus long PCR	*arg*W	argW3bis	TTGGGATCTATCCCCTTGAC	62	/
	*dsd*A	dsdA-R	GCTGCTTTACGTCCTTCCTC		
Screening of recombinant cosmids PCRs	*arg*W	argW3	TGCAAAGCAGCTAAGCCACG	55	250
		argW3bis	TTGGGATCTATCCCCTTGAC		
	*asn*T	asnT5bis	ATGTCACTGGTTCGAGTCC	55	191
		asnT6bis	GGGGTCACTCCAGAATCG		
	*leu*X	leuX1	GCAGGACAGTGTGAGATTCAC	55	267
		leuX19	CCGTCGCCAGCCAGAATAC		
	*phe*V	pheV5	GATGATAGATTGTGCAGTCTGC	55	500
		pheV6	ACGAGACGAGGCGAATCAG		
	*ser*U	serU4	GTCGCGTACAGGCTAACC	55	592
		serU5	CTGTCGTATACCAGCACC		
	*thr*W	thrW11	ACACATCATTAGTACCTTCC	55	256
		thrW12	AAAGCGGCGCTGAGACGG		

Table 1: Continued.

	Target position in *E. coli* O157:H7 Sakaï genome	Primer*	Primer sequence (5'→3')	Annealing temperature (°C)	PCR product size (bp)
Screening of SpLE5 PCRs	5347257	leuX1	GCAGGACAGTGTGAGATTCAC	58	5001
	5352052	leuX2	AAGCTGCCGTTTTTACCTCA		
	5352173	leuX3	CAACCCTAAGGCCATCTGAG	60	5001
	5356975	leuX4	AAGCCATTCAGCGTCAAAGTAA		
	5357133	leuX5	AAGGCAGGTATGGCACGTTA	58	4803
	5361933	leuX6	CTGGGCTGATGATGGTTTTT		
	5362177	leuX7	TTGCTTGCTCAAAAGCATGTA	60	4790
	5366967	leuX8	AGCTGTCGGGGTTCCTTAAT		
	5367145	leuX9	AACTGTCTGCGCTTTTTGAAT	58	4805
	5371948	leuX10	TTTATTGCTCCCCTGACGAT		
	5372276	leuX11	CTCGATTGCAGTTGCGATAA	58	5001
	5377073	leuX12	CTTCCTGATCGTGTGCTTCA		
	5377199	leuX13	CTCTCTACGGCGGTCTTCAC	60	5001
	5381993	leuX14	GCGTTCATCGGGTAGATGAT		
	5382112	leuX15	ATGTCGGCAGAAAAGGTCATT	60	5001
	5386910	leuX16	TCTGCGACGCTATTTTTGTGGA		
	5387111	leuX17	AACGCAACAGTTTTGCACTG	58	5001
	5391906	leuX18	TGGTTGAAGCCTTTCGTAGG		
	Target gene or factor	Primer*	Primer sequence (5'→3')		
Arbitrary PCRs		ARB1	GGCCACGGCGTCGACTAGTACNNNNNNNNNNGACTG		
		ARB2	GGCCACGCGTCGACTAGTAC		
		ARB3	GGCCACGCGTCGACTAGTACNNNNNNNNNNGAATG		
		ARB4	GGCCACGCGTCGACTAGTACNNNNNNNNNNGATTCG		
	*arg*W	argW1ter	CTCCATTAGCGGGTATACTCATGCC		
		argW-F	TAGTTAAATGGATATAACGAGC		
	*asn*T	asnT-F	TAAGTAATATGCGCCCCGTTC		
		asnT3bis	TAGTTCAGTCGGTAGAACGG		
	*asp*V	aspV-F	GGCGGAAGCTGCCTCTGGCG		
		aspV3	TTCCCAACGGAACACAACGC		
	*leu*X	leuX1	GTGGCGTGCGACAGGTATA		
		leuX3	CGTTTTCCGCATACCTCTTC		
	*phe*V	pheV3bis	GATGATAGATTGTGCAGTCTGC		
		pheV-F	ACGAGACGAGGCGAATCAG		
	*ser*U	serU3	GAACAAAATTGCACGGACAG		
		serU-F	TCCTGGCATCATGGCAACC		
	*thr*W	thrW1	AAGGCCATTGACGCATCGC		
		thrW3	TGAACATGTCCTTTCAGGGC		

Table 2: PCR amplification procedures used in this study.

	Initial denaturation step	Amplification cycles	Final elongation step
Phylogenetic group, MLST, virulence factors, and screening of recombinant plasmids PCRs	94°C-5min	(94°C-30s, T-30s, 72°C-30s) *30	72°C-7min
Screening tRNA of *asn*T, *arg*U, *arg*W, *leu*X, *met*V, *phe*U, *phe*V, *sel*C, *ser*U, *ser*W, *thr*W loci PCRs	94°C-5min	(95°C-1min, T-1min, 72°C-1min) *30	72°C-7min
Screening tRNA of *asn*UVW, *asp*V, *gly*U, *ser*X loci and SpLE5 PCRs	94°C-5min	(95°C-1min, T-1min, 72°C-5min) *30	72°C-7min
ITS 16S-23S PCR	95°C-3min	(95°C-1min, T-50s, 72°C-2min) *35	72°C-10min
*arg*W tRNA locus long PCR	94°C-1min	(98°C-10s, T-30s, 72°C-20min) *30	72°C-10min
First step of arbitrary PCRs	95°C-5min	(95°C-30s, 30°C-30s, 72°C-2min) *6 (95°C-30s, 45°C-30s, 72°C-2min) *30	None
Second step of arbitrary PCRs	95°C-5min	(95°C-30s, 52°C-30s, 72°C-2min) *30	None

Table 3: ORFs detected within the 1875-bp *E. coli* P4 fragment obtained in the *pheV*-tRNA screening approach. RBS[a]: Ribosome Binding Site, sequence and position. (aa)[b]: amino acids. CD[c]: conserved domain.

ORF	Position	RBS[a]	G+C content (%)	Predicted size (aa)[b]	Accession number	CD[c]	Characteristics of homologous protein			
							Function (Accession number)	Species (strain)	Lenght (aa)[b]	%identity/ %similarity
ORF1 *-pheV*	370-816	/	53.2	148	ACG50150	/	Hypothetical protein (CAE55810)	*E. coli* (Nissle 1917)	143	91/94
ORF2 *-pheV*	683-925	/	49.6	80	ACG50151	/	Hypothetical protein (ZP_00716238.1)	*E. coli* (B7A)	80	98/100
ORF3 *-pheV*	1022-1594	AGGAA, 1009-1013	45.8	190	ACG50152	/	Hypothetical protein (CAI43851)	*E. coli* (RW1374)	190	98/98

Table 4: ORFs detected within the *E. coli* P4 *serU* genomic island partially characterized. RBS[a] : Ribosome Binding Site, sequence and position. (aa)[b] : amino acids. CD[c]: conserved domain.

ORF	Position	RBS[a]	G+C content (%)	Predicted size (aa)[b]	Accession number	CD[c]	Characteristic of homologous protein			
							Function (Accession number)	Species (strain)	Lenght (aa)[b]	%Identity/ %similarity
ORF1-*serU*	168-1580	AGGGG (96-100)	32.7	470	ACG50163	/	O-antigen conversion protein (NP_720275)	Entero-bacteriophage (ST64T)	485	24/45
ORF2-*serU*	2941-1625	AGGGG (2996-2992)	39.6	438	ACG50164	/	/	/	/	/
ORF3-*serU*	3886-3074	AGCGG (3973-3969)	50.3	270	ACG50165	Integrase/ Transposase (entire)	IShe3 transposase (YP_001456923)	*E. coli* (HS)	270	99/99
ORF4-*serU*	6702-4600	AGGGG (6794-6790)	34.3	700	ACG50166	/	TnpX hypothetical protein (NP_061387)	Plasmid F	698	100/100
ORF5-*serU*	7427-10435	GGGAG (7416-7420)	48.5	1002	ACG50167	Transposase (entire)	TnpA transposase (NP_061389)	Plasmid F	1002	100/100
ORF6-*serU*	11117-10359	/	46.2	252	ACG50168	Transposase (partial)	IS30 transposase (YP_001451392)	*E. coli* (E24377A)	383	100/100
ORF7-*serU*	12098-13087	/	44.7	329	ACG50169	Integrase (entire)	Integrase (NP_700401)	*S. typhimurium* phage (ST64B)	329	84/94

Table 5: ORFs detected within the *E. coli* P4 *argW* genomic island. RBS[a] : Ribosome Binding Site, sequence and position. (aa)[b] : amino acids. CD[c]: conserved domain.

ORF	Position	RBS[a]	G+C content (%)	Predicted size (aa)[b]	Accession number	CD[c]	Characteristic of homologous protein			
							Function (Accession number)	Species (strain)	Lenght (aa)[b]	%Identity/ %similarity
ORF1 -*argW*	165-1355	AGGGG (127-131)	43.07	396	ACG50154	P4 integrase (partial)	Integrase (YP_153333)	*Salmonella enterica* (ATCC9150)	419	68/82
ORF2 -*argW*	1512-1703	GAGG (1503-1506)	46.88	63	ACG50155	Transcriptional regulator (entire)	Predicted transcriptional regulator (ZP_00834742)	*Yersinia intermedia* (ATCC29909)	65	73/91
ORF3 -*argW*	2671-3723		43.59	350	ACG50156	/	Putative replication protein (NP_311263)	*E. coli* (O157:H7 Sakaï)	231	97/99
ORF4 -*argW*	6910-4430	AAGAG (6921-6917)	31.20	826	ACG50157	Sir2 (entire)	Hypothetical protein (YP_487138)	*Rhodopseudomonas palustris* (HaA2)	792	23/41
ORF5 -*argW*	7564-7388		43.50	58	ACG50158	/	Hypothetical protein (NP_311266)	*E. coli* (O157:H7 Sakaï)	58	77/84
ORF6 -*argW*	7867-8493		47.69	208	ACG50159	Recombinase/ Resolvase (entire)	Putative resolvase (NP_311267)	*E. coli* (O157:H7 Sakaï)	208	98/99
ORF7 -*argW*	9681-8746	ATGAG (9692-9688)	48.08	311	ACG50160	Transcriptional regulator (entire)	Transcriptional regulator (AAC75424)	*E. coli* (K12 MG1655)	311	100/100
ORF8 -*argW*	9899-11236	AGGGG (9884-9888)	48.73	445	ACG50161	Transporter (entire)	Predicted transporter (AAC75425)	*E. coli* (K12 MG1655)	445	100/100

Discussion et perspectives

La thématique de ces travaux de thèse a été de rechercher des déterminants génétiques pouvant expliquer l'adaptation de la souche mammopathogène *E. coli* P4 à la glande mammaire bovine. La pathogénicité des souches MPEC, longtemps sous-estimée, est très peu, voire non élucidée. De manière unanime, il est admis à l'heure actuelle que le manque de connaissances sur les facteurs de virulence de ces souches s'explique par l'utilisation de méthodes expérimentales non adéquates pour leur mise en évidence.

Nous avons tout d'abord entrepris une étude phylogénétique d'*E. coli* P4. En effet, bien que citée et qualifiée dans la littérature comme environnementale et opportuniste, aucune souche MPEC n'a fait, à notre connaissance, l'objet de ce type d'étude. Pourtant, cette démarche est importante pour mieux comprendre l'évolution d'une espèce et l'acquisition de la pathogénicité par une souche bactérienne. L'étude phylogénétique d'*E. coli* P4 a consisté, d'une part, à déterminer, par PCR, à quel groupe phylogénétique de l'espèce *E. coli* elle appartient et, d'autre part, à affiner et confirmer les résultats phylogénétiques obtenus en analysant quelques parties de son génome core. Les résultats obtenus ont montré que cette souche mammopathogène appartient au groupe phylogénétique A, groupe rassemblant majoritairement des souches commensales non pathogènes. De plus, comparé à tous les génomes *E. coli* entièrement séquencés à l'heure actuelle, le génome core d'*E. coli* P4 se rapproche le plus du génome core de la souche *E. coli* K12 MG1655, commensale non pathogène, appartenant également au groupe phylogénétique A. *Cette proximité phylogénétique d'*E. coli *P4 avec les souches commensales non pathogènes du groupe phylogénétique A de l'espèce* E. coli *va constituer le premier point de discussion de ce manuscrit. Quelques hypothèses sur la situation phylogénétique des autres souches MPEC et sur la création du pathotype MPEC au sein du groupe phylogénétique A vont être proposées.*

Afin de répondre à une problématique soulevée lors de travaux antérieurs effectués au laboratoire, concernant l'identification des acteurs de la caséinolyse élevée du lait observée lors d'une mammite bovine induite par *E. coli* P4, une recherche d'une sécrétion d'enzymes extracellulaires à activité caséinolytique par *E. coli* P4 a été, dans un deuxième temps, entreprise. Une sécrétion effective et, vraisemblablement, constitutive de quatre protéases a pu être mise en évidence par la technique de zymographie caséines. Cependant, les différents procédés expérimentaux utilisés pour mesurer l'activité de ces enzymes ont révélé un potentiel caséinolytique très faible, voire nul, soutenant l'absence d'un rôle direct de la bactérie dans l'hydrolyse élevée des protéines du lait observée lors d'une mammite bovine induite par cette bactérie. La découverte de protéases extracellulaires sécrétées par *E. coli* P4 constitue, tout de même, un fait intéressant, étant donné l'importance que peut avoir ce type d'enzyme dans la virulence chez des souches pathogènes de l'espèce *E. coli*. *La mise en évidence d'une sécrétion probablement constitutive de quatre protéases*

extracelulaires par E. coli *P4 va constituer le deuxième point de discussion de ce manuscrit. La réalité du potentiel caséinolytique de ces quatre protéases va tout d'abord être analysée, puis le rôle particulier des protéases sécrétées par d'autres souches* E. coli *pathogènes va être examiné.*

Ayant constaté que, contrairement à *E. coli* K12 MG1655, *E. coli* P4 est capable de se développer dans un lait issu de mammite bovine, c'est-à-dire dans un lait riche en substances antibactériennes, et une étude ayant rapporté dans la littérature sa capacité également à adhérer et à envahir des cellules épithéliales mammaires (Gonen *et al*., 2007), nous nous sommes intéressés à rechercher des déterminants génétiques expliquant ces différentes propriétés de virulence.

La première démarche expérimentale a été de rechercher, par PCR, la présence dans son génome de différents gènes codant des facteurs de virulence connus chez les autres pathotypes de l'espèce *E. coli*. Sur 26 gènes de virulence examinés, seul un, *traT*, a pu être retrouvé. *Ce type de recherche va constituer le troisième point de discussion de ce manuscrit. La pertinence des facteurs recherchés, en relation avec les propriétés de virulence phénotypiques connues d'*E. coli *P4, va tout d'abord être analysée, puis quelques hypothèses expliquant les résultats obtenus vont être proposées.*

La seconde démarche expérimentale entreprise, pour tenter d'expliquer la pathogénicité d'*E. coli* P4, a été de rechercher la présence d'îlots génomiques potentiels en aval de loci d'ARNt dans son génome. Cette démarche expérimentale permet souvent en effet, d'identifier des parties variables au sein d'un génome bactérien et donc de découvrir, éventuellement, de nouvelles propriétés spécifiques d'une souche. Sept loci d'ARNt, au niveau desquels une insertion d'ADN étranger a pu être mise en évidence, ont ainsi été détectés sur le génome de la souche *E. coli* P4. *La présence d'ADNs étrangers à proximité de ces loci va constituer le quatrième et dernier point de discussion de ce manuscrit. La nature des ADNs insérés en aval des loci* asnT*,* leuX*,* pheV*,* serU *et* thrW *va tout d'abord être analysée, puis le cas du locus* argW *sera traité avec une attention particulière au rôle que pourrait jouer la protéine codée par l'ORF 4-*argW *dans la pathogénicité d'*E. coli *P4*.

1. *E. coli* P4 appartient au groupe phylogénétique A de l'espèce *E. coli*

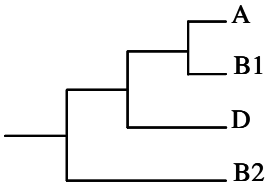
Nos travaux ont montré qu'*E. coli* P4 appartient au groupe phylogénétique A de l'espèce *E. coli*. Cette détermination est la première jamais effectuée à notre connaissance sur une souche MPEC. D'une part, les études rapportant la caractérisation de souches *E. coli* à l'origine de mammites bovines portent, la plupart du temps, sur une recherche de facteurs de virulence ou sur une analyse de génotype et ne déterminent jamais le groupe phylogénétique de ces souches. D'autre part, les études évaluant la distribution phylogénétique de souches de différents pathotypes de l'espèce *E. coli*, ne prennent jamais en considération le pathotype MPEC.

La connaissance du groupe phylogénétique d'une seule souche mammopathogène, celui d'*E. coli* P4, permet de réfléchir à la question suivante : est-ce que toutes les souches MPEC appartiennent uniquement au groupe A, ou, au contraire, sont-elles réparties dans deux, trois, ou dans les quatre groupes phylogénétiques de l'espèce *E. coli* ? En regardant la répartition des autres pathotypes de l'espèce *E. coli*, aucun d'entre eux n'est exclusivement associé à un seul groupe phylogénétique (tableau 14). Les souches ExPEC et les souches ETEC, EHEC et EIEC sont retrouvées au minimum dans deux groupes (B2/D et A/B1, respectivement) ; les souches commensales non pathogènes, EAEC, DAEC et EPEC sont retrouvées, quant à elles, dans les quatre groupes. Ainsi, il est fort probable que des souches MPEC autres qu'*E. coli* P4 appartiennent à un autre groupe que le groupe A. Etant donné que les résultats obtenus lors de l'analyse MLST d'*E. coli* P4 l'ont rapproché d'une souche commensale non pathogène et que les génotypes des souches MPEC semblent aussi variés que ceux des souches *E. coli* environnementales isolées de fourrage ou de fécès de bovins en bonne santé (Yang *et al*., 2004), une distribution des souches MPEC dans les quatre groupes phylogénétiques, tout comme les souches commensales, peut être attendue. Il serait intéressant de vérifier cette hypothèse en déterminant le groupe phylogénétique d'un plus grand nombre de souches MPEC. Par la suite, il serait également intéressant de regarder si des souches MPEC n'appartenant pas au même groupe phylogénétique présentent des propriétés de virulence différentes. Il a été déjà dit en effet dans ce manuscrit que le bagage génétique de virulence d'une souche *E. coli* pouvait être lié à la nature de son génome core.

Le groupe phylogénétique A de l'espèce *E. coli*, auquel appartient *E. coli* P4, est le groupe où se concentrent majoritairement les souches commensales non pathogènes de cette espèce et où le plus faible taux de gènes codant des facteurs de virulence est recensé (Picard *et al*., 1999). Par conséquent, la présence de souches pathogènes au sein de ce groupe n'est pas expliquée par l'existence d'un ancêtre pathogène dont les descendants auraient, pour la plupart, perdu le caractère virulent, mais par

l'existence d'un ancêtre non pathogène dont quelques descendants seulement auraient acquis un caractère virulent, ceci à la suite d'évènements de transferts génétiques horizontaux. D'une manière générale d'ailleurs, l'ancêtre commun aux quatre principaux groupes phylogénétiques de l'espèce *E. coli* est considéré comme non pathogène (Lecointre *et al*., 1998). Le groupe phylogénétique B2 serait le premier à avoir été créé, puis le groupe D et enfin les groupes A et B1 (figure 38).

Figure 38 : Arbre représentant l'apparition des quatre groupes phylogénétiques de l'espèce *E. coli*, d'après Lecointre *et al*., 1998.



La plus ou moins grande virulence de chaque groupe proviendrait uniquement de l'acquisition de facteurs de virulence par transfert génétique horizontal et de l'expression de ces derniers, actions toutes deux étroitement liées au génome core de chaque groupe phylogénétique. L'apparition de la souche mammopathogène *E. coli* P4 au sein du groupe phylogénétique A est donc probablement due, également, à des transferts génétiques horizontaux. Le groupe A, tout comme le groupe B1, est, d'ailleurs, caractérisé par le génome core capable d'accueillir la plus large gamme de gènes codant des facteurs de virulence. Des gènes spécifiques de virulences intestinale et extra-intestinale sont identifiés dans le génome de souches de ce groupe (tableau 14). Ceci n'est pas le cas des génomes cores des groupes B2 et D, dans lesquels aucun gène spécifique de la virulence intestinale des souches ETEC, EHEC et EIEC, souches intestinales considérées comme les plus sévères (les souches EAEC, DAEC et EPEC sont considérées quant à elles de sévérité moyenne), n'a pu être détecté jusqu'à présent (tableau 14). *E. coli* P4 pourrait avoir acquis certains gènes de virulence de souches pathogènes appartenant également au groupe phylogénétique A, ou à l'un des trois autres principaux groupes phylogénétiques de l'espèce *E. coli* (figure 39). Cette hypothèse peut notamment être confortée par le fait que, tout comme les souches des autres pathotypes de l'espèce *E. coli*, intestinaux ou extra-intestinaux, *E. coli* P4 présente des propriétés d'adhérence et d'invasion cellulaires (Gonen *et al*., 2007). Des transferts génétiques pourraient donc être à l'origine de ce partage de propriétés. Les résultats d'analyse des loci d'ARNt d'*E. coli* P4 présentant des structures altérées ont d'ailleurs montré la présence de séquences nucléotidiques qui pouvaient être retrouvées dans le génome de souches *E. coli* pathogènes aussi bien intestinales qu'extra-intestinales.

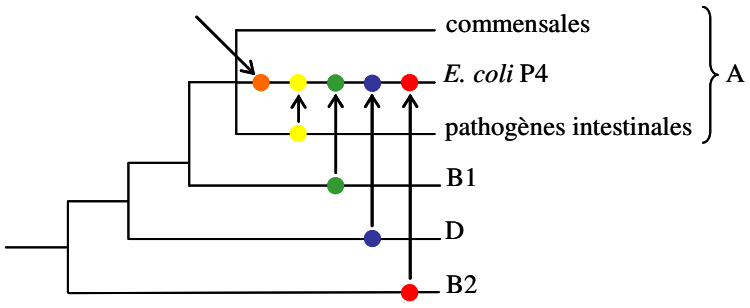


Figure 39 : Situation et hypothèses sur l'apparition d'*E. coli* P4 au sein de l'espèce *E. coli*. Les points correspondent à un ou des gènes de virulence et les flèches indiquent des évènements de transfert génétique horizontal.

Ce transfert de gènes entre souches *E. coli* de différents groupes phylogénétiques a notamment été proposé par Lecointre *et al*. pour les gènes de virulence *kps* et *pap*, responsables de la synthèse d'une capsule de type II et d'adhésines fimbriaires de type P, respectivement et communs aux souches uropathogènes des groupes B2 et D (Lecointre *et al*., 1998). Ces auteurs suggèrent que ces deux types de gènes auraient été acquis par transfert génétique horizontal, dans un premier temps, par les souches du groupe phylogénétique B2, puis, que celles-ci les auraient transmis aux souches du groupe phylogénétique D, toujours par transfert génétique horizontal. Ces mêmes auteurs n'envisagent pas la présence de ces deux types de gènes chez un ancêtre commun aux deux groupes phylogénétiques, car cet ancêtre commun est partagé par les groupes phylogénétiques A et B1 (figure 38). Or, les gènes *kps* et *pap* sont très rarement identifiés dans le génome de souches appartenant aux groupes A et B1 ; leur présence chez un ancêtre commun signifierait donc que la majorité des souches de ces deux groupes les auraient perdus au cours du temps, ce qui est peu probable. Par ailleurs, l'origine horizontale des gènes *kps* et *pap* est soutenue par leur présence sur des éléments génétiques mobiles et par leur contenu en paires de bases guanine et cytosine différent de celui du génome core des souches *E. coli*. *E. coli* P4 pourrait également avoir acquis certains gènes de virulence de souches pathogènes appartenant à des espèces autres qu'*E. coli* et partageant la même niche écologique qu'elle (figure 39). La notion de niche écologique est, en effet, importante. Par exemple, les gènes *kps* et *pap*, présentés précédemment, semblent typiques des souches uropathogènes et n'ont pas été identifiés dans le génome de souches pathogènes intestinales (Boyd et Hartl, 1998). Ceci montre donc bien que les souches uropathogènes du groupe B2 ont acquis pour la première fois ces deux types de gènes à la suite d'un transfert génétique horizontal réalisé avec des souches bactériennes qui partagent leur niche écologique et pas celle des souches intestinales.

2. Implication d'*E. coli* P4 dans la caséinolyse observée lors d'une mammite bovine

2.1. Mise en doute du rôle caséinolytique des quatre protéases extracellulaires mises en évidence

Nous avons montré qu'*E. coli* P4 est capable de sécréter au moins quatre protéases extracellulaires dans différents milieux de culture et à différentes phases de croissance. Une expression constitutive peut donc être suggérée pour ces quatre enzymes protéolytiques. Ces dernières ont été mises en évidence par la technique de zymographie, en utilisant comme substrat les caséines et ont ainsi été qualifiées de caséinolytiques. Cependant, les différentes approches expérimentales menées par la suite, ont révélé que l'intensité de leur activité caséinolytique semblait très faible, voire insignifiante. Ces résultats permettent de soulever deux interrogations. La caséinolyse ou plus généralement l'hydrolyse de protéines dans le but de fournir des éléments nutritifs à la bactérie, est-elle le rôle premier de ces enzymes ? Les "quatre" enzymes détectées contribuent-elles à un rôle direct d'*E. coli* P4 dans la caséinolyse du lait observée lors d'une mammite ?

En considérant uniquement les résultats obtenus *in vitro*, la réponse à la première question posée serait un non. En d'autres termes, les caséines ne seraient pas les substrats spécifiques et préférentiels des quatre enzymes détectées et ces dernières n'auraient ainsi aucun rôle nutritif pour la bactérie. Deux observations appuient cette hypothèse. Premièrement, les quatre enzymes ont pu être détectées dans les milieux extracellulaires des cultures d'*E. coli* P4 réalisées en milieu MS, alors que ce milieu ne contient aucun substrat nutritif à hydrolyser. Deuxièmement, aucune croissance d'*E. coli* P4 dans un milieu ne contenant que des caséines comme sources de carbone et d'azote n'a pu être observée, alors que les quatre enzymes extracellulaires étaient détectées. Le fait d'affirmer que les caséines ne sont pas les substrats préférentiels des quatre protéases mises en évidence, bien que le rôle protéolytique de ces dernières ait été démontré par zymographie caséines, ne semble pas aussi invraisemblable qu'il paraît. Les caséines sont, en effet, comme l'insuline, la gélatine et la globine, des substrats couramment utilisés dans les techniques de zymographie pour mettre en évidence l'activité protéolytique de protéines, sans être pour autant les substrats préférentiels des protéines en question. Ainsi, le rôle protéolytique de nombreuses protéines intracellulaires des bactéries *E. coli* a été découvert grâce à l'utilisation des caséines ou de la gélatine, sans que ces dernières ne soient les véritables substrats des protéines étudiées (ces types de substrats ne sont d'ailleurs pas naturellement présents dans les cellules bactériennes) (Sreedhara Swamy et Goldberg, 1982). L'activité

caséinolytique non spécifique des quatre protéases détectées est peut-être tout simplement dévoilée avec les conditions expérimentales de la zymographie. Ces dernières incluent, notamment, la présence de SDS dans les gels de concentration et de séparation et dans le tampon d'échantillon. Or, le SDS est un agent dénaturant les protéines. Ainsi, si la renaturation des protéines (étape effectuée après la migration électrophorétique et avant l'incubation du gel d'électrophorèse dans la solution permettant de dévoiler l'activité protéolytique) n'est pas totale, celles-ci peuvent présenter un domaine de fixation du substrat quelque peu modifié, d'où l'accueil non spécifique possible des caséines.

La réponse à la deuxième question posée serait également un non, mais un non conditionnel. Un rôle direct d'*E. coli* P4, via les quatre protéases mises en évidence, dans la caséinolyse du lait observée lors d'une mammite ne peut être en effet soutenu, si l'intensité de l'activité de ces protéases n'augmente pas dans les conditions créées lors de l'inflammation. Cette augmentation pourrait avoir pour origine, soit une expression en plus grande quantité des protéases, soit une activation des protéases. Ces deux hypothèses sont envisageables *in vivo*.

En ce qui concerne la première hypothèse, il est vrai que l'absence de détection d'activité caséinolytique significative, lors des différentes démarches expérimentales menées, était peut-être due à la présence, dans les conditions *in vitro* créées, d'une trop faible quantité des quatre protéases mises en évidence. L'essai de purification conduit en milieu synthétique soutient d'ailleurs ce fait. *In vivo*, lors de l'inflammation, les quatre protéases détectées pourraient être exprimées en plus grande quantité suite, par exemple, à une augmentation du taux de transcription de leurs gènes codants. Il est bien connu, en effet, que le taux de transcription de certains gènes bactériens est étroitement lié à certaines conditions environnementales données. Une de ces conditions peut être le stress oxydant, particulièrement présent, notamment, dans la mamelle bovine lors d'un évènement de mammite (la description de ce stress oxydant sera faite dans le paragraphe 3.1.1 de ce chapitre) (Burvenich *et al*., 2003). De nombreuses protéines dont l'expression est particulièrement induite en condition de stress oxydant ont été, effectivement, identifiées chez *E. coli* (Farr et Kogoma, 1991). Un approche par électrophorèse bidimentionnelle a montré, par exemple, que l'expression de plus de 30 protéines était augmentée lorsque des cellules *E. coli* étaient traitées avec de l'eau oxygénée (VanBogelen *et al*., 1987). Ainsi, il serait intéressant, en perspectives, de comparer les taux de transcription *in vitro* et *in vivo* des gènes codant les quatre enzymes extracellulaires mises en évidence, lors d'une incubation d'*E. coli* P4 dans un milieu de culture classique et lors d'une inflammation après l'infusion de la bactérie dans une mamelle bovine, respectivement. Cependant, cette étude nécessiterait, au préalable, l'identification des quatre enzymes et présenterait probablement quelques difficultés de mise en oeuvre, notamment pour l'obtention d'ARNm de bonne qualité. La comparaison, par électrophorèse bidimentionnelle, des protéines extracellulaires bactériennes exprimées dans les deux situations précédentes, *in vitro* et *in vivo*, serait plus difficile à envisager. L'électrophorégramme obtenu *in vivo*

serait en effet probablement trop complexe à analyser, du fait de la présence d'un trop grand nombre de protéines endogènes du lait issu de mammite bovine.
En ce qui concerne la deuxième hypothèse, l'activation des quatre protéases, lors de l'inflammation, pourrait correspondre à l'accueil d'un cofacteur favorable, ou à une coupure protéolytique (très souvent observée dans le cas des protéases, comme par exemple dans le cas de la plasmine) catalysée par un acteur propre à la bactérie ou à la glande mammaire. Il est également possible que les quatre protéases mises en évidence soient naturellement plus actives sur les caséines lorsque celles-ci se trouvent dans un état oxydé, état probablement induit par le stress oxydant important qui se développe dans la mamelle bovine lors d'un évènement de mammite. Une étude a d'ailleurs montré qu'en milieu oxydant les caséines étaient plus susceptibles à la protéolyse (Wiking et Nielsen, 2004). Cette plus grande vulnérabilité à la protéolyse est expliquée par un changement conformationnel des caséines, suite à la fixation de groupements aldéhydes libérés lors de l'oxydation des constituants lipidiques du lait. Cette hypothèse pourrait être vérifiée en purifiant les caséines d'un lait issu de mammite bovine et en les incubant avec les protéines extracellulaires d'*E. coli* P4 cultivée en milieux LB et MS.

Si, toutefois, l'absence d'un rôle direct d'*E. coli* P4 dans la caséinolyse du lait observée lors d'une mammite se confirme, l'étude de son rôle indirect devra être envisagée. Celui-ci pourrait être lié à l'induction d'une maturation différente des neutrophiles à leur arrivée dans le sang, maturation différente forcément couplée à un potentiel protéolytique différent de ces cellules.

2.2. Rôle des protéases extracellulaires sécrétées par les bactéries *E. coli* pathogènes

La sécrétion de protéases extracellulaires, ou, plus généralement, de protéines extracellulaires, par des souches *E. coli* commensales non pathogènes, souches proches d'un point de vue phylogénétique d'*E. coli* P4, est très rarement rapportée dans la littérature. Une étude de ce genre a été publiée en 2006 (Nandakumar *et al*., 2006) et concerne la souche *E. coli* K12 W3110. Au sein des souches *E.* coli pathogènes, de nombreuses protéines extracellulaires ont, au contraire, été identifiées, notamment des protéases.

La majorité de ces protéases, sécrétées par des souches *E. coli* pathogènes, appartiennent à la famille des SPATEs. Les SPATEs, ou Serine Protease Autotransporters of *Enterobacteriaceae*, sont des protéases à sérine dirigeant leur propre transport du cytoplasme vers le milieu extracellulaire et spécifiquement retrouvées chez les entérobactéries (Henderson *et al*., 1998). Ces enzymes partagent plusieurs caractéristiques communes : une séquence signale N-terminale, un domaine central sécrété correspondant à leur forme mature et un domaine C-terminal portant la fonction d'auto-transport. Les

principales protéines SPATEs découvertes à ce jour chez les bactéries *E. coli* pathogènes sont les protéines Sat, Pet, EspC, EspP, Pic, EatA, Hbp et Vat. Quelques propriétés de ces protéases sont présentées dans le tableau 26.

Tableau 26 : Caractéristiques principales de quelques protéines SPATEs identifiées chez des souches *E. coli* pathogènes. PAI : îlot de pathogénicité.

Nom	Pathotype prédominant détenteur	Localisation du gène codant	Substrat	Référence
Sat	UPEC	Chromosomique (PAI)	Fodrine, autres protéines de la membrane plasmique, du cytosquelette et du noyau	Maroncle *et al*., 2006
Pet	EAEC	Plasmidique	Fodrine	Villaseca *et al*., 2000
EspC	EPEC	Chromosomique (PAI)	Fodrine	Navarro-Garcia *et al*., 2004
EspP	EHEC	Plasmidique	Pepsine, facteur V de coagulation humain	Brunder *et al*., 1997
Pic	EAEC	Chromosomique	Mucine, protéines du système du complément, hémagglutinine	Henderson *et al*., 1999
EatA	ETEC	Plasmidique	Encore inconnu	Patel *et al*., 2004
Hbp	*E. coli* causant des infections intra-abdominales	Chromosomique (épisome)	Hémoglobine	Otto *et al*., 1998
Vat	UPEC	Chromosomique (PAI)	Encore inconnu	Parreira et Gyles, 2003

L'action protéolytique des protéines Sat, Pet et EspC sur la fodrine conduit à la désorganisation du cytosquelette des cellules hôtes (Maroncle *et al*., 2006). La fodrine est en effet une protéine eucaryote impliquée dans la stabilisation des structures membraires, dans le maintien de la forme des cellules et dans la liaison des filaments d'actine à la membrane plasmique (Viel et Branton, 1996). Le clivage du facteur V de coagulation humain par la protéine EspP entraîne notamment une exacerbation des colites hémorragiques (Brunder *et al*., 1997). L'action de la protéine Pic sur les protéines du système du complément stoppe la cascade d'activation de ce dernier (Henderson *et al*., 1999). Enfin, l'hydrolyse de l'hémoglobine par la protéine Hbp s'accompagne d'une libération de fer qui profite aux bactéries pathogènes (Otto *et al*., 1998).

Un exemple de protéase extracellulaire n'appartenant pas à la famille des SPATEs et sécrétée par une bactérie *E. coli* pathogène est la protéine StcE, métalloprotéase à zinc codée par un gène situé sur le plasmide de virulence pO157 de la souche EHEC O157:H7 Sakaï (Grys *et al*., 2005). La sécrétion de cette protéase se fait via un système de sécrétion de type II codé également sur le plasmide pO157. Trois substrats ont pu être identifiés pour cette protéase : la molécule C1-INH (C1 esterase inhibitor), la glycoprotéine 340 (gp340) et la mucine 7 (MUC7). Les deux derniers substrats cités sont des protéines glycosylées protégeant la surface des cellules hôtes. La molécule C1-INH a

pour fonction quant à elle, la régulation du système du complément et d'autres mécanismes typiques d'une inflammation.

L'analyse fonctionnelle des protéases sécrétées par les souches *E. coli* pathogènes montre donc qu'elles correspondent à d'importants facteurs de virulence. Ainsi, si l'absence d'une activité caséinolytique spécifique devait se confirmer pour les quatre protéases mises en évidence chez *E. coli* P4, il serait intéressant de déterminer si ces quatre enzymes protéolytiques ne présenteraient pas plutôt un rôle dans la virulence de la souche. Ses capacités à adhérer et à envahir les cellules épithéliales mammaires (Gonen *et al*., 2007) pourraient soutenir cette hypothèse. L'entrée dans les cellules hôtes requiert souvent, en effet, la désorganisation du cytosquelette, désorganisation catalysée par les protéases Sat, Pet et EspC chez des souches UPEC, EAEC et EPEC, respectivement. L'identification des quatre enzymes mises en évidence permettrait également de savoir si l'une d'entre elles ne correspondrait pas à une des protéases connues des souches *E. coli* pathogènes décrites précédemment.

3. Recherche, au sein du génome d'*E. coli* P4, de gènes codant des facteurs de virulence connus chez les autres pathotypes de l'espèce *E. coli*

3.1. Pertinence des gènes de virulence recherchés

3.1.1. Gènes bactériens de résistance aux défenses immunitaires hôtes

Nous avons montré qu'*E. coli* P4 était capable de se développer aussi bien dans un lait issu de mammite bovine, que dans un lait sain. Or, dans un lait issu de mammite bovine, plusieurs constituants à vocation antibactérienne, majoritairement impliqués dans la réponse immunitaire innée spécifique des vertébrés, voient leur concentration augmenter (Sordillo *et al*., 1997). Ces constituants sont, d'une part, les neutrophiles et, d'autre part, la lactoferrine, la transferrrine, la molécule du complément, la xanthine oxydase et l'oxyde nitrique synthase.

Les neutrophiles, après être entrés en contact avec les bactéries infectieuses, les internalisent et les tuent selon deux voies possibles : une appelée flambée oxydative, dépendante d'entités oxygénées réactives et une indépendante de ces dernières (Burvenich *et al*., 2003). La flambée oxydative est déclenchée dès l'établissement du contact bactérie / neutrophile, par l'activation d'une enzyme, la

NADPH oxydase, présente sur la membrane des neutrophiles. Cette enzyme catalyse l'oxydation du NADPH tout en formant des anions superoxyde (O_2^-). A partir des ces anions superoxyde, une cascade de réactions catalysées par des enzymes du neutrophile peut se dérouler, aboutissant à la formation de diverses entités oxygénées réactives néfastes pour les bactéries, tels que l'eau oxygénée (H_2O_2), le radical hydroxyl ($OH^\cdot$) et l'ion hypochloreux (OCl^-) (figure 40). Ces entités générées dans le cytoplasme des neutrophiles sont également libérées à l'extérieur des cellules immunitaires.

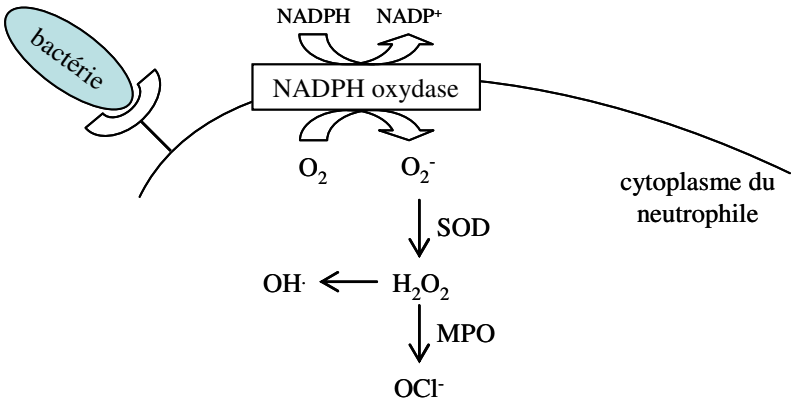


Figure 40 : Schématisation de la flambée oxydative catalysée au sein d'un neutrophile suite à un contact avec une bactérie infectieuse. SOD : superoxyde dismutase, MPO : myélopéroxydase.

La voie de destruction du pathogène, indépendante des entités oxygénées réactives, implique diverses enzymes et peptides contenus dans les neutrophiles et capables d'altérer la perméabilité membranaire des bactéries. Un exemple d'enzymes mises en jeu est l'élastase, attaquant la protéine de la membrane externe OmpA d'*E. coli* (Belaaouaj *et al*., 2000). Parmi les exemples de peptides, peuvent être cités des peptides antimicrobiens, appelés β–défensines, possédant une activité bactéricide *in vitro*, aussi bien vis-à-vis de *St. aureus* que d'*E. coli* (Selsted *et al*., 1993).
La lactoferrine et la transferrrine sont toutes deux des agents bactériostatiques de par leur capacité à lier le fer pour le rendre indisponible aux bactéries, la molécule du complément peut jouer le rôle d'opsonine et / ou d'agent bactéricide et enfin, la xanthine oxydase et l'oxyde nitrique synthase produisent toutes deux de l'oxyde nitrique, pouvant être ensuite transformé en peroxynitrite à pouvoir bactéricide (Sordillo *et al*., 1997 ; Rainard et Riollet, 2006).

D'après les différentes propriétés antibactériennes d'un lait issu de mammite, *E. coli* P4 doit, pour pouvoir se développer dans ce milieu, éviter de se faire reconnaître comme entité étrangère et contrer toutes les attaques dont elle est la cible, telles que la lyse, la privation en fer et l'oxydation. Afin de savoir si *E. coli* P4 est capable de mener à bien une ou plusieurs de ces actions de résistance, nous avons recherché, par PCR dans son génome, la présence des gènes *neuA-neuC*, *traT* et *iucD*,

responsables, chez les autres souches *E. coli* pathogènes, de la synthèse d'une capsule, d'une protéine de la membrane externe et d'un système d'approvisionnement en fer, respectivement.

3.1.1.1. Gènes *neuA-neuC*

Les gènes *neuA-neuC*, de localisation chromosomique, codent deux enzymes impliquées dans la synthèse d'une capsule chez les souches *E. coli* pathogènes (Russo, 2002). Les capsules, appelées également K-antigènes (K-Ag), sont des structures présentes à la surface des cellules bactériennes qui présentent plusieurs fonctions de résistance. Du fait qu'elles recouvrent les cellules bactériennes, elles peuvent empêcher tout simplement le dépôt de la molécule du complément sur les membranes cellulaires. Ce phénomène semble toutefois proportionnel à la quantité de capsule présente et varier avec les différentes natures de capsules existantes (Johnson, 1991).Quelques acides spécifiques fixés aux composants glucidiques de certains K-Ag sont capables de lier des inhibiteurs de la molécule du complément et de bloquer ainsi la cascade d'activation de cette dernière (Leying *et al*., 1990). Enfin, les capsules peuvent prévenir la phagocytose des bactéries de par leur charge négative élevée et leur hydrophobicité et certaines seraient même faiblement immunogènes, du fait de leur ressemblance avec d'autres constituants cellulaires hôtes (Johnson, 1991). Les gènes *neuA-neuC* n'ont pas pu être mis en évidence au sein du génome d'*E. coli* P4. Ce résultat suggère que, soit *E. coli* P4 n'exprime pas de capsule à sa surface, soit un système de synthèse, de ce type de structure, différent de celui connu au sein de l'espèce *E. coli* est présent sur son génome. Une caractéristique connue d'*E. coli* P4 peut toutefois apporter une propriété de résistance similaire à celle des capsules : l'expression d'un O-antigène. Certains O-Ag, particulièrement ramifiés, peuvent prévenir, en effet, tout comme les capsules, le dépôt de la molécule du complément, du fait de la création d'un encombrement stérique à la surface des cellules bactériennes (Hull, 1997). Les O-Ag seraient également dotés d'une autre propriété de résistance : leur présence rendrait la membrane externe des bactéries moins fluide, ce qui empêcherait l'insertion du complexe d'attaque membranaire de la molécule du complément (Russo, 2002). L'O-antigène exprimé par *E. coli* P4 est de type 32 (Vangroenweghe *et al*., 2004). Ce dernier est un hétéropolymère composé, en quantité équimolaire, d'unités de glucose, de galactose, d'acide glucuronique, de N-acétylglucosamine et de N-acétylfucosamine. Le nombre de répétitions de ces unités nous est pour l'instant inconnu pour pouvoir évaluer le degré de résistance qu'elles apportent à la bactérie, mais l'expression d'un O-Ag par *E. coli* P4 pourrait donc, en partie, expliquer sa résistance en milieu antibactérien.

3.1.1.2. Gène *traT*

Le gène *traT*, de localisation plasmidique, code la protéine TraT de la membrane externe des bactéries à Gram négatif, capable d'inhiber l'action du complexe d'attaque membranaire de la

molécule du complément (Hancock et Piers, 1994). Initialement identifié sur les plasmides conjugatifs de l'espèce *E. coli*, ce gène a également été retrouvé sur des plasmides de souches bactériennes d'autres espèces, telles que *Salmonella typhimurium* et *Yersinia enterolitica* (Sukupolvi et O'Connor, 1990). Ces plasmides ne semblent pas pouvoir se transmettre par le biais de la conjugaison, ne portent aucun autre gène *tra* à proximité du gène *traT*, mais renferment au contraire différents gènes impliqués dans la virulence de ces souches. L'origine du gène *traT* sur ces plasmides n'est pas encore expliquée. Une première hypothèse consiste à dire que ces plasmides seraient d'anciens plasmides conjugatifs, ayant perdu progressivement tous les autres gènes impliqués dans la conjugaison, pour ne garder que le gène *traT* conférant aux bactéries qui les possèdent un avantage sélectif pour la pathogénicité. Une deuxième hypothèse suggère que ces plasmides auraient acquis le gène *traT* par un évenement de recombinaison ou de transposition. Le couple d'oligonucléotides spécifique du gène *traT* est le seul couple avec lequel un produit d'amplification de taille attendue a pu être obtenu par PCR à partir du génome d'*E. coli* P4. Si le séquençage de ce produit confirme son identité au gène *traT*, ce fait atteste donc l'existence chez *E. coli* P4 d'un autre système de résistance face aux défenses immunitaires hôtes. De plus, il suggère que son génome renferme peut-être au moins un plasmide. Cependant, des travaux à ce sujet n'ayant pas été davatange engagés, la nature du plasmide impliqué, conjugatif ou non, ne peut être à l'heure actuelle établie.

3.1.1.3. Gène *iucD*

Le gène *iucD* code une enzyme impliquée dans la synthèse de l'aérobactine. Cette dernière est un sidérophore, molécule très répandue dans le monde bactérien pour s'approvisionner en fer. De par leur très haute affinité pour cet élément, les sidérophores sont en effet capables de le détacher des chélateurs hôtes au profit des bactéries (Griffiths *et al*., 1988). Au sein de l'espèce *E. coli*, trois entités de ce type ont été identifiées, l'entérobactine, l'aérobactine et la yersiniabactine et il a également été montré que la production de citrate pouvait être utilisée par certaines souches pour croître en milieu appauvri en fer (Griffiths, 1994, 1997). Seul le système aérobactine a été recherché au sein du génome d'*E. coli* P4, des données épidémiologiques ayant montré que sa production était restreinte aux souches pathogènes à potentiel invasif de l'espèce *E. coli*.

L'aérobactine est une molécule de petite taille (616 Da), formée par l'association de deux résidus de lysine et d'un citrate (figure 41) (Johnson, 1991). Une molécule d'aérobactine sécrétée extrait le fer piégé dans les chélateurs hôtes et le délivre à un récepteur de la membrane externe des bactéries sans être altérée, pouvant ainsi être recyclée et réutilisée continuellement. Le système aérobactine est sous la dépendance de cinq gènes regroupés dans un opéron : quatre gènes *iuc* (iron uptake:chelate) codent les enzymes catalysant la synthèse de la molécule et un gène *iut* (iron uptake:tranport) code le récepteur de la membrane externe de la molécule (figure 41) (Johnson, 1991).

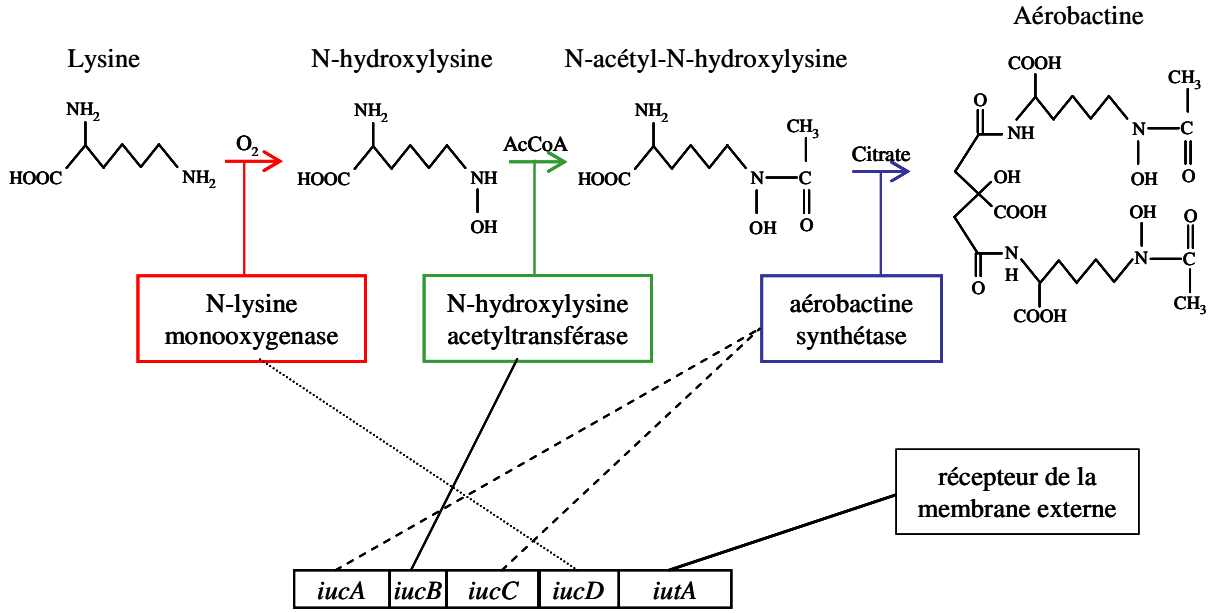


Figure 41 : Voie de biosynthèse de l'aérobactine et gènes et protéines impliqués, d'après Johnson, 1991.

Aucun résultat positif d'amplification n'a pu être obtenu lors de la recherche, par PCR, du gène *iucD* au sein du génome d'*E. coli* P4. Etant donné, cependant, le caractère essentiel du fer dans la croissance des bactéries, *E. coli* P4 doit, très probablement, pour pouvoir se développer dans un lait issu de mammite, disposer d'un système pour le fixer ou l'assimiler. Plusieurs hypothèses peuvent donc être envisagées. La première serait l'existence d'un système aérobactine présentant au niveau de ses gènes codants, notamment au niveau du gène *iucD*, quelques variations alléliques comparé aux gènes actuellement connus, qui sont pour l'instant assez bien conservés. La deuxième hypothèse serait l'existence d'un système caractéristique habituellement de souches bactériennes appartenant à une espèce autre qu'*E. coli*. Divers autres systèmes de sidérophores ont en effet été identifiés dans le monde bactérien. Des gènes dirigeant par exemple la synthèse d'une vibriobactine, d'une acinetobactine, d'une mycobactine, d'une pyocheline ou d'une anguibactine, principaux autres systèmes de sidérophores bactériens recensés chez *Vibrio cholerae*, *Acinetobacter calcoaceticus*, *Mycobacterium tuberculosis*, *P. aeruginosa* ou *Vibrio anguillarum*, respectivement (Crosa et Walsh, 2002), pourraient être ainsi présents dans le génome d'*E. coli* P4. Ces gènes pourraient avoir été acquis lors d'un transfert génétique horizontal. Ce transfert pourrait avoir eu lieu notamment avec une bactérie *P. aeruginosa*, espèce présente, entre autre, dans l'environnement, tout comme les MPEC et responsable de pathologies au niveau du tractus urinaire chez l'homme, donc de tropisme extra-intesinal, tout comme les MPEC. La troisième et dernière hypothèse formulée ici serait bien sûr l'existence d'un nouveau système d'approvisionnement en fer, encore jamais décrit.

3.1.2. Gènes bactériens d'adhérence et d'invasion aux cellules hôtes

Des travaux, réalisés *in vivo* chez la souris, ont montré qu'*E. coli* P4 était capable d'adhérer à et d'envahir des cellules épithéliales (Gonen *et al*., 2007). Ces deux propriétés ont également été observées chez d'autres souches MPEC (Döpfer *et al*., 2000, 2001 ; Lammers *et al*., 2001 ; Dogan *et al*., 2006 ; Passey *et al*., 2008). Les mécanismes d'adhérence et d'invasion cellulaires impliqués ne sont, à l'heure actuelle, pas précisément renseignés. En ce qui concerne l'étape d'adhérence, des appendices de nature encore inconnue et ressemblant à de fins fils ont été observés à deux reprises entre des cellules *E. coli* et des cellules mammaires (Döpfer *et al*., 2000 ; Lammers *et al*., 2001). Les processus d'adhérence et d'invasion ne semblent pas entraîner la formation de lésions à la surface des cellules mammaires (Döpfer *et al*., 2000). Il a également été constaté que l'étape d'invasion est inhibée en présence d'inhibiteurs de phosphokinases (Döpfer *et al*., 2001 ; Dogan *et al*., 2006), mais qu'elle ne s'accompagne d'aucun réarrangement d'actine au sein des cellules hôtes (Passey *et al*., 2008). Enfin, il a été montré que les souches MPEC pouvaient survivre dans les cellules épithéliales mammaires pendant près de cinq jours, 48 heures et 24 heures (Döpfer *et al*., 2001, Dogan *et al*., 2006 et Passey *et al*., 2008, respectivement). Cette survie intracellulaire semble être expliquée par l'internalisation des bactéries dans une structure de type vacuole, dont les caractéristiques ont commencé à être examinées (Passey *et al*., 2008). Cette vacuole peut être marquée avec des molécules typiques du mécanisme d'endocytose et ne semble jamais fusionner avec des compartiments lysosomiaux acides ou phagocytaires.

La mise en évidence de ces propriétés d'adhérence et d'invasion cellulaire d'*E. coli* P4 et d'autres souches MPEC nous a donc conduit à rechercher une deuxième catégorie de gènes de virulence au sein du génome d'*E. coli* P4, à savoir des gènes codant divers facteurs d'adhérence et d'invasion cellulaires. Parmi les différents facteurs de virulence assurant ces fonctions au sein de l'espèce *E. coli*, nous nous sommes focalisés sur la recherche de ceux spécifiques, notamment, aux souches UPEC, pour deux raisons. La première est que les souches UPEC présentent, tout comme les souches MPEC, un tropisme extra-intestinal. La deuxième est qu'il a été rapporté des cas d'infections urinaires récurrentes causées par des souches UPEC (Mulvey *et al*., 2001 ; Schilling *et al*., 2002), suggérant la persistance de ces dernières dans les tissus hôtes, tout comme il a été mis en évidence des cas récurrents de mammites et des souches MPEC persistantes. Parmi les souches UPEC, une grande variété de composants cellulaires de surface, appelés adhésines et responsables d'une adhérence aux cellules hôtes, ont été identifiés. La classification de ces composants, pas toujours évidente, repose sur leurs propriétés structurales et leurs spécificités de récepteurs sur les cellules hôtes (Mainil, 2003).

Trois principaux groupes d'adhésines peuvent ainsi être définis, sachant qu'au sein de chaque groupe de nombreux sous-groupes sont ensuite répertoriés (tableau 27).

Tableau 27 : Présentation des principaux groupes d'adhésines recensés chez les souches UPEC, d'après Mainil, 2003. Gal : galactose ; NeuAc : N-acétyl-neuraminique ; GalNAc : N-acétyl-D-galactosamine ; GlcNAc : N-acétyl-D-glucosamine ; DAF : Decay Accelerating Factor.

Groupe	Sous-groupe	Structure	Spécificité de récepteurs hôtes
I	-	Fimbriaire	Molécule glycosylée contenant au moins une entité mannose
II	P	Fimbriaire	Molécule glycosylée contenant au moins l'entité Gal-(α1-4)-Gal
	S		Molécule glycosylée contenant au moins l'entité NeuAc-(α2-3)-Gal-(β1-3)-GalNAc
	F1C		Molécule glycosylée contenant au moins l'entité GalNAc-(β1-4)-Gal
	F17		Molécule glycosylée contenant au moins l'entité GlcNAc en position terminale
III	Afa	Afimbriaire	DAF (excepté pour les types Afa7 et Afa8) : glycoprotéine membranaire régulant la cascade du complément et protégeant les cellules eucaryotes contre une action mal dirigée de ce dernier
	M		Glycophorine A
	F1845	Fimbriaire	DAF : glycoprotéine membranaire régulant la cascade du complément et protégeant les cellules eucaryotes contre une action mal dirigée de ce dernier
	Dr		
	Nfa	Capsule-like	Molécule glycosylée contenant au moins l'entité GlcNAc en position terminale

Au sein du génome d'*E. coli* P4, la présence de gènes impliqués dans la synthèse d'adhésines des sous-groupes P, S et F17 du groupe II, a été tout d'abord recherchée. Les adhésines du groupe II, de nature fimbriaire, sont composées essentiellement d'une protéine d'ancrage à la membrane externe, puis d'une sous-unité majeure répétée formant un cylindre et enfin, en partie terminale, de diverses sous-unités mineures dont l'une porte la fonction d'adhésine proprement dite (Mainil, 2003). Ces adhésines du groupe II ont été réparties dans différents sous-groupes selon leur spécificité de récepteur. De plus, au sein de chaque sous-groupe, plusieurs variants génétiques ont été identifiés : *pap* et *prs* pour les adhésines de type P, *sfaI* et *sfaII* pour les adhésines de type S et *f17a*, *f17b, f17c* et *f17d* pour les adhésines de type F17, par exemple. Ces variants diffèrent principalement au niveau de la sous-unité majeure (dont le nom est dans la plupart des cas associé à la lettre A) et de la sous-unité mineure porteuse de la fonction d'adhésine (dont le nom est dans la plupart des cas associé à la lettre G). Les gènes qui ont été recherchés chez *E. coli* P4 sont les gènes codant les sous-unités mineures PapE et PapF des adhésines fimbriaires de type P, les gènes codant la sous-unité mineure SfaD et la protéine chaperone SfaE des adhésines fimbriaires de type S et les gènes codant la sous-unité majeure des adhésines fimbriaires de type F17a, F17b, F17c et F17d. Tous les résultats obtenus ont été

négatifs. Jusqu'à présent, les sous-unités mineures ne portant pas la fonction d'adhésine et les protéines chaperones des adhésines fimbriaires de type P et S sont assez conservées, même entre les différents variants génétiques de ces deux sous-groupes (Mainil, 2003). Ainsi, soit les adhésines de type P et S ne sont effectivement pas présentes chez *E. coli* P4, soit un nouveau variant génétique pour l'un de ces deux types d'adhésine, voire les deux, existe chez *E. coli* P4 et peut-être chez d'autres souches MPEC, avec des variations génétiques non habituelles dans les gènes codant les sous-unités mineures et les protéines chaperones. De même, la non-detection des quatre variants génétiques de la sous-unité majeure des adhésines de type F17, actuellement répertoriés, implique que, soit *E. coli* P4 n'exprime pas ce type d'adhésine, soit un cinquième variant encore jamais décrit et peut-être typique des souches MPEC existerait. L'hypothèse d'autres variants génétiques des adhésines de type F17 a déjà été soulevée par d'autres auteurs (Bertin *et al*., 1996 ; Mainil *et al*., 2000 ; Stordeur *et al*., 2002).
La présence de gènes impliqués dans la synthèse d'adhésines du sous-groupe Afa du groupe III, a été, d'autre part, recherchée. Les adhésines du groupe III ont des structures non fimbriaires (Mainil, 2003). Deux d'entre elles, les adhésines de type Dr et F1845, sont décrites comme fimbriaires, mais les appendices extracellulaires qui les caractérisent présentent toutefois une taille beaucoup plus petite que celles des appendices des adhésines du groupe II. Ainsi, les adhésines du groupe III sont composées essentiellement d'une protéine d'ancrage à la membrane externe, d'une sous-unité porteuse de la fonction d'adhésine proprement dite et d'une sous-unité porteuse d'une fonction d'invasine. Ces adhésines du groupe III ont également été réparties dans différents sous-groupes selon leur structure et leur spécificité de récepteur, puis au sein de chaque sous-groupe plusieurs variants génétiques ont été identifiés : *afa1* à *afa8* pour les adhésines de type Afa et *nfa1* à *nfa6* pour les adhésines de type Nfa, par exemple. Ces variants génétiques diffèrent principalement au niveau de la sous-unité porteuse de la fonction d'adhésine (dont le nom est souvent associé à la lettre E). Les gènes qui ont été recherchés chez *E. coli* P4 sont les gènes codant les sous-unités porteuses des fonctions d'adhésine et d'invasine des adhésines afimbriaires de type Afa7 et Afa8. Ce choix s'expliquait par le fait que les adhésines Afa apportent, aux bactéries qui les expriment, les fonctions d'adhérence et d'invasion de manière combinée, deux propriétés observées chez *E. coli* P4. Par ailleurs, la recherche plus particulière des variants génétiques Afa7 et Afa8 reposait sur le fait qu'à l'heure actuelle, seuls ces deux variants sont retrouvés chez les souches *E. coli* pathogènes bovines. Les résultats obtenus ont de nouveau été négatifs. Ainsi, les mêmes hypothèses que précédemment peuvent être formulées : soit *E. coli* P4 n'exprime pas d'adhésine afimbriaire de type Afa, soit *E. coli* P4 exprime un nouveau variant génétique de ce type d'adhésine.

3.2. Hypothèses expliquant les résultats obtenus

Plusieurs hypothèses ont été évoquées dans le paragraphe précédent pour expliquer l'absence, presque systématique, de détection de gènes de virulence, connus par ailleurs, au sein du génome d'*E. coli* P4 et des souches MPEC en général (tableau 15), bien que les propriétés phénotypiques de virulence de ces dernières sont désormais évidentes (Döpfer *et al*., 2000, 2001 ; Lammers *et al*., 2001 ; Dogan *et al*., 2006 ; Passey *et al*., 2008). Ces hypothèses sont : la présence de nouveaux gènes encore non identifiés au sein de l'espèce *E. coli* ; la présence de gènes identifiés au sein de l'espèce *E. coli*, mais présentant quelques variations alléliques chez *E. coli* P4 et peut-être chez d'autres MPEC ; et enfin, la présence de gènes habituellement associés à des bactéries appartenant à des espèces autres qu'*E. coli*. Les deux premières hypothèses citées ont été émises pour tous les gènes de virulence recherchés et la troisième dans le cas de la recherche du système aérobactine. Une étude, réalisée à la fin de l'année 2007, permet de soutenir ces différentes hypothèses et semble très intéressante pour commencer à résoudre ce problème des bactéries mammopathogènes.

Cette étude rapporte le séquençage complet du génome de la souche *St. aureus* ET3-1 (Herron-Olson *et al*., 2007), isolat représentatif des clones de l'espèce *St. aureus* les plus fréquemment isolés de mammites bovines à travers le monde (Fitzgerald *et al*., 1997). Le génome de cette souche présente une taille de 2,74 Mb et renferme 2589 ORFs. L'analyse de ces ORFs a montré que plusieurs d'entre elles codaient des protéines aux fonctions de virulence, telles que des protéines impliquées dans la synthèse de capsules et de toxines, dans des systèmes hydrolytiques, d'adhérence, de détoxification, de résistance à des antibiotiques, ou encore d'adaptation à des conditions atypiques (tableau 28). Certaines des ORFs de virulence sont retrouvées sur le génome de souches *St. aureus* pathogènes pour l'homme. Cependant, un pourcentage élevé de différences nucléotidiques ponctuelles entre les séquences des ORFs de la souche bovine ET3-1 et des souches humaines est observé, entraînant au niveau de la structure primaire des protéines codées par ces ORFs de grandes différences en composition d'acides aminés. Ainsi, ce premier constat soutient la deuxième hypothèse proposée pour *E. coli* P4. L'analyse des 2589 ORFs identifiées sur le génome de la souche *St. aureus* ET3-1 a également révélé l'existence de 88 ORFs codant des protéines n'ayant aucune séquence homologue significative dans les banques de données et de 33 ORFs codant des protéines homologues à des protéines enregistrées dans les banques de données mais appartenant uniquement à des souches bactériennes d'espèces autres que *St. aureus*. Ce deuxième constat vient soutenir quant à lui les première et troisième hypothèses proposées pour *E. coli* P4. Ces 88 et 33 ORFs inconnues et homologues à des ORFs de souches bactériennes d'espèces autres que *St. aureus*, respectivement, sont pour la plupart retrouvées au sein d'éléments génétiques mobiles, accompagnées de quelques unes des ORFs de virulence présentées précédemment (figure 42 et tableau 29).

Tableau 28 : Recensement des ORFs aux fonctions de virulence identifiées sur le génome de la souche *St. aureus* ET3-1, d'après Herron-Olson *et al.*, 2007.

Fonction	Nombre d'ORFs impliquées	Exemples de protéines codées
Synthèse de capsules	22	Enzymes catalysant la synthèse de capsules …
Toxicité	35	Leucocidines Hémolysines Protéines homologues aux streptolysines …
Hydrolyse	27	Protéases à sérine Coagulase …
Adhérence	19	Protéines de liaison aux immunoglobulines, à la fibronectine, au fibrinogène …
Détoxification	11	Superoxyde dismutase Hydroperoxyde réductase Catalase Protéines de résistance à l'arsenic et aux anions toxiques …
Résistance aux antibiotiques	38	Protéines de résistance aux quinolones, à la méthicilline, à la pénicilline, à la tétracycline …
Adaptation aux conditions atypiques	15	Protéines de réponse aux stress alcalin, froid, osmotique …

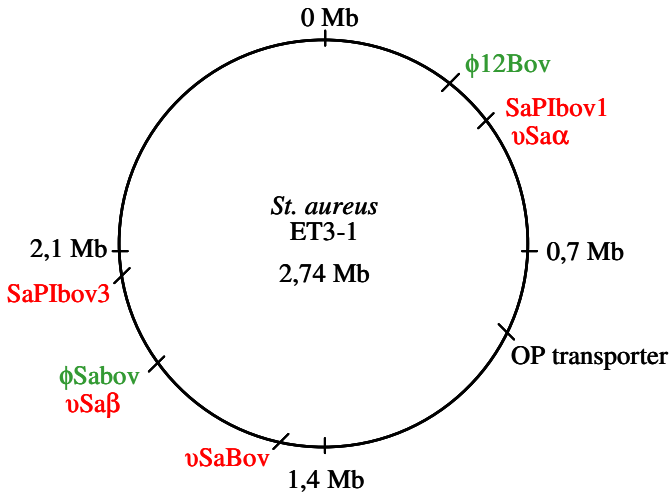


Figure 42 : Localisation des huit principaux éléments génétiques mobiles identifiés sur le chromosome de la souche *St. aureus* ET3-1, d'après Herron-Olson *et al.*, 2007. Les éléments inscrits en couleur verte et rouge sont des prophages et des îlots de pathogénicité, respectivement.

Tableau 29 : Nombre et fonction des gènes contenus dans les huit principaux éléments génétiques mobiles identifiés sur le génome de la souche *St. aureus* ET3-1, d'après Herron-Olson *et al*., 2007.

Elément	Nombre total de gènes	Nombre de gènes codant								
		des protéines de fonction inconnue	des toxines	des protéines aux fonction de mobilité	des protéines d'origine phagique	des protéines sans homologie dans les banques de données	des protéines homologues à des protéines de bactéries *St. aureus*	des protéines homologues à des protéines de bactéries autres que *St. aureus*	des protéases	des lipoprotéines
SaPIbov1	23	19	3	1	0	0	0	0	0	0
SaPIbov3	28	13	0	3	1	5	4	2	0	0
ϕSaBov	59	0	0	1	47	7	0	4	0	0
ϕ12Bov	13	0	0	1	7	3	1	1	0	0
υSaBov	14	0	0	0	0	5	1	8	0	0
υSaα	35	0	9	0	0	3	17	0	0	6
υSaβ	38	0	8	0	0	5	20	0	5	0

Le contenu d'un de ces éléments génétiques mobiles, l'élément υSaBov, est particulièrement intéressant. Sur les quatorze gènes qui composent cet élément, cinq codent des protéines sans homologie dans les banques de données et huit codent des protéines homologues à des protéines de bactéries autres que *St. aureus*. Deux de ces dernières sont homologues à des gènes codant des streptolysines identifiés pour la première fois sur le génome de souches *S. pyogenes* (Nizet *et al*., 2000). Les streptolysines sont des protéines capables de se lier aux molécules de cholestérol constituant la membrane des neutrophiles, provoquant par la suite la désintégration des granules lysosomiaux et la libération du contenu toxique de ces derniers dans le cytoplasme des cellules.

L'analyse du génome de la souche *St. aureus* ET3-1, le premier et le seul génome d'une souche bactérienne inductrice de mammites bovines entièrement séquencé à ce jour, semble donc apporter des données concrètes pour commencer à décrire plus sûrement la pathogénicité des souches *St. aureus* bovines. Le fait d'identifier dans ce génome des gènes nouveaux, des gènes habituellement associés à des espèces autres que *St. aureus* et des gènes retrouvés chez d'autres pathotypes de l'espèce *St. aureus* présentant toutefois de grandes variations alléliques, constitue une découverte très importante pour la caractérisation des souches bactériennes inductrices de mammites bovines en général. En effet, ces nouveautés montrent que tenter d'expliquer la pathogénicité de ces souches en se basant sur les propriétés bien connues d'autres agents pathogènes, n'est peut-être pas toujours adéquat, du moins suffisant. Ceci semble particulièrement vrai pour les souches inductrices de mammites bovines de l'espèce *E. coli*. Il est donc raisonnable de penser que le séquençage complet d'un génome d'une souche *E. coli* inductrice de mammites bovines révèlerait des informations toutes aussi nouvelles, intéressantes et utiles pour tenter d'expliquer la pathogénicité des MPEC.

4. *E. coli* P4 renferme dans son génome plusieurs îlots génomiques potentiels à proximité de différents loci d'ARNt

Le criblage de loci d'ARNt chez des souches *E. coli* mammopathogènes n'avait encore jamais été décrit à notre connaissance. Sur 17 loci d'ARNt criblés au sein du génome d'*E. coli* P4, six, au niveau desquels une insertion d'ADN étranger a pu être mise en évidence, ont été identifiés : *argW*, *asnT*, *leuX*, *pheV*, *serU* et *thrW*. Alors que l'analyse phylogénétique effectuée a rapproché le génome core d'*E. coli* P4 de celui de la souche *E. coli* K12 MG1655, des séquences nucléotidiques très différentes de celles présentes dans le génome d'*E. coli* K12 MG1655 ont au contraire été déterminées au niveau de ces six loci d'ARNt.

4.1. Analyse des loci d'ARNt altérés partiellement caractérisés

4.1.1. Cas du locus *asnT*

Aucune ORF complète n'a pu être détectée dans la région amplifiée et séquencée en 3' du locus *asnT* sur le génome d'*E. coli* P4. Cependant, la comparaison de la séquence obtenue avec les séquences des banques de données a montré l'existence d'une séquence similaire (96 % d'identité) au sein du génome de la souche *E. coli* C ATCC 8739, souche commensale non pathogène appartenant au groupe phylogénétique A de l'espèce *E. coli*. Chez cette dernière, la séquence en question renferme le début d'un pseudogène présentant 82 % d'identité avec la séquence d'une ORF codant une intégrase putative chez la souche *Sh. boydii* Sb 227. Chez *E. coli* C ATCC 8739, le pseudogène est retrouvé en 3' de deux loci d'ARNt-Asn et est suivi par d'autres pseudogènes, ainsi que par des ORFs codant des protéines hypothétiques de fonction inconnue. De même, chez *Sh. boydii* Sb 227, l'ORF codant l'intégrase putative, également située en 3' d'un locus d'ARNt-Asn, est suivie par des gènes codant des protéines de fonction inconnue, des transposases et des intégrases. A ce stade des travaux, aucune hypothèse ne peut donc être avancée concernant le rôle de la région *asnT* dans la pathogénicicté ou tout simplement dans la nature d'*E. coli* P4. Il est intéressant de noter toutefois que le locus *asnT* chez *E. coli* P4 n'est pas associé à la présence de l'îlot génomique HPI (High Pathogenicity Island). Ce dernier, identifié initialement chez des bactéries du genre *Yersinia* et porteur de gènes permettant la synthèse et l'expression d'un sidérophore appelé yersiniabactine, a été mis en évidence en 3' du locus *asnT* du génome de nombreuses souches *E. coli*, aussi bien pathogènes, appartenant notamment aux pathotypes UPEC (Schubert *et al*., 1999), NMEC (Clermont *et al*., 2001),

ETEC et EPEC (Wang *et al*., 2002) et EAEC (Hu *et al*., 2005), que non pathogènes, comme des souches probiotiques (Schubert *et al*., 1999). Ces mises en évidence ont reposé sur le séquençage complet de l'îlot ou sur la recherche par PCR de certains gènes spécifiques de l'îlot et ont abouti à la découverte de structures dans certains cas intactes et dans d'autres cas altérées de l'îlot. Quoiqu'il en soit, l'élément HPI semble très répandu au sein de l'espèce *E. coli* et il y est très souvent associé au locus *asnT* (Schmidt et Hensel, 2004). Certains auteurs pensent que cette intégration caractéristique est dûe à la présence d'une intégrase spécifique du locus *asnT* au sein de l'îlot (Schubert *et al*., 1999). Le fait de retrouver l'élément HPI au sein du génome de souches *E. coli* présentant des tropismes très différents suggère que, contrairement à ce que son nom indique, il ne serait pas ou peu impliqué dans la virulence des souches qui le possèdent. De plus, la plupart de ces dernières disposent dans leur génome d'un système codant un autre type de sidérophore dont l'affinité pour le fer serait supérieure à celle de la yersiniabactine. Ainsi, d'autres fonctions du HPI sont parfois proposées, comme celle de moduler les défenses cellulaires hôtes (Ewald *et al*., 1994 ; Rakin *et al*., 1999). Dans quelques cas, l'élément HPI a été identifié dans des régions autres que *asnT*, comme le locus *asnUVW* par exemple (Dezfulian *et al*., 2003). Cette localisation atypique ne peut être envisagée pour *E. coli* P4, car le criblage de ce locus au sein de son génome a montré qu'il était intact. Ainsi, excepté une autre localisation atypique encore non décrite dans la littérature, la présence de l'élément HPI au sein du génome d'*E. coli* P4 est peu probable. Ce fait sous-entend qu'*E. coli* P4 (ou les souches qui l'ont précédée d'un point de vue phylogénétique) ait, soit, jamais possédé cet élément du fait d'une occupation de niches écologiques différentes de celles des souches l'ayant acquis, soit, perdu cet élément au cours du temps.

4.1.2. Cas du locus *leuX*

En ce qui concerne la nature de l'îlot génomique présent à proximité du locus *leuX*, au sein du génome d'*E. coli* P4, une des hypothèses avancées, d'après les résultats de séquences partielles obtenus, est que ce locus pourrait être associé à une forme altérée de l'élément SpLE5 présent au niveau du même locus chez la souche *E. coli* O157:H7 Sakaï. Cette forme altérée se traduirait, d'une part, par la présence d'une intégrase (première protéine codée par l'élément) dont l'ORF renferme plusieurs mutations ayant conduit, notamment, à la création d'un codon stop prématuré (figure 31) et, d'autre part, par la délétion très probable de plusieurs fragments de l'élément, comme le suggèrent les résultats de son criblage par PCR grâce à différents couples d'oligonucléotides spécifiques de l'intégralité de sa séquence. L'élément SpLE5 d'*E. coli* O157:H7 Sakaï contient, notamment, en plus de l'intégrase évoquée précédemment, quatre autres ORFs codant des protéines putatives à fonction de mobilité : deux transposases d'IS, une résolvase et une deuxième intégrase. L'hypothèse de la

présence d'une forme altérée de l'élément SpLE5 chez *E. coli* P4 semble tout à fait probable, étant donné que cette situation est retrouvée dans le génome d'une autre souche *E. coli*, la souche *E. coli* EDL933. Chez cette dernière, l'élément SpLE5 a effectivement pu être identifié, mais sous une forme partiellement identique à celle présente chez *E. coli* O157:H7 Sakaï. Comparée à la forme présente chez *E. coli* O157:H7 Sakaï, la forme identifiée chez *E. coli* EDL933 est dépourvue d'une ORF codant une protéine hypothétique de fonction inconnue et comporte deux ORFs de plus petite taille, codant également une protéine hypothétique de fonction inconnue pour l'une et une protéine putative à fonction de transposase pour l'autre. Les cinq protéines putatives à fonction de mobilité (résolvase, intégrases, transposases), codées par des ORFs de l'élément SpLE5 présent chez *E. coli* O157:H7 Sakaï, pourraient avoir gouverné différents évènements de recombinaison à l'origine des structures variables présentes chez *E. coli* P4 et EDL933. Il serait ainsi intéressant, par la suite, de séquencer les différents produits d'amplification obtenus lors du criblage par PCR de l'élément SpLE5 au sein du génome d' *E. coli* P4, pour identifier les éventuels types de réarrangements de cet élément qui se sont peut-être produits.

4.1.3. Cas du locus *pheV*

Les résultats obtenus dans la caractérisation du locus *pheV* d'*E. coli* P4 soutiennent très bien la présence d'un îlot génomique. En effet, ont été identifés, d'une part, un gène codant une intégrase probablement fonctionnelle localisé immédiatement en 3' du gène *pheV* et, d'autre part, une région localisée immédiatement en 5' du gène *yghD* (premier gène constant retrouvé en 3' du locus *pheV* au sein des génomes de l'espèce *E. coli*) contenant une répétition de 23 pb du gène *pheV* et trois ORFs potentielles codant des protéines putatives, dont deux sont similaires à des protéines déjà identifées au sein d'autres îlots génomiques de l'espèce *E. coli*. Les quatre ORFs détectées au total présentent également des pourcentages en paires de bases guanine et cytosine différents de celui des génomes complets de l'espèce *E. coli*. L'intégrase putative mise en évidence est similaire à une intégrase putative de type P4 codée par une ORF située également immédiatement en 3' d'un locus d'ARNt-Phe au sein du génome de la souche *Sh. flexneri* 2a 301. Chez cette dernière, l'ORF codant l'intégrase putative est suivie, notamment, par plusieurs ORFs codant des protéines hypothétiques de fonction inconnue, une ORF codant une ADN hélicase et deux ORFs codant des protéases à sérine. Le contenu de cette région génomique de *Sh. flexneri* 2a 301 est très différent du contenu des îlots génomiques renfermant les protéines similaires aux protéines codées par les trois ORFs identifiées en 5' du gène *yghD* d'*E. coli* P4. La première de ces trois ORF code une protéine putative similaire à une protéine hypothétique de fonction inconnue mise en évidence dans l'îlot GEI_{II} de la souche commensale probiotique *E. coli* Nissle 1917, également localisé en 3' du locus *pheV* (Grozdanov *et al*., 2004). La

deuxième ORF code une protéine putative similaire à une protéine hypothétique de fonction inconnue mise en évidence au sein du génome de la souche entéropathogène *E. coli* B7A. Le génome de cette dernière étant encore en cours d'annotation, l'analyse de la position de cette protéine hypothétique n'a pas pu être faite. Enfin, la troisième ORF code une protéine putative similaire à une protéine hypothétique de fonction inconnue mise en évidence au sein de l'îlot LEE-PAI de la souche bovine entéropathogène *E. coli* RW1374 (Jores *et al*., 2005). Tout comme l'îlot GEI_{II}, l'îlot LEE-PAI, d'une taille de 111 kb, est situé en 3' du locus *pheV* du génome d'*E. coli* RW1374. Il renferme notamment, l'îlot LEE (Locus of Enterocyte Effacement) spécifique des souches EHEC et EPEC, le gène *efa1* codant un facteur d'adhérence et les gènes *lifA* et *ent* codant une lymphostatine (toxine bloquant l'activation des lymphocytes) et une toxine homologue aux toxines Shiga, respectivement. L'îlot génomique vraisemblablement présent en 3' du locus *pheV* du génome d'*E. coli* P4 semble donc correspondre à un assemblage complexe de fragments identifiés au sein de génomes de souches bactériennes présentant des tropismes très différents. Ce fait correspond à une des caractéristiques des îlots génomiques. L'îlot GEI_{II} de la souche *E. coli* Nissle 1917 renferme par exemple plusieurs fragments partiellement identiques à des portions de l'îlot PAI-*pheV* de la souche uropathogène *E. coli* CFT073. Sur l'îlot GEI_{II}, ces fragments ne sont pas forcément positionnés dans le même ordre que sur l'îlot PAI-*pheV* et sont séparés par des régions nucléotidiques non identifiées dans l'îlot PAI-*pheV*. Etant donné la fonction inconnue des trois protéines similaires aux protéines codées par les trois ORFs situées en 5' du gène *yghD* d'*E. coli* P4, le rôle de la région génomique *pheV* dans la pathogénicité ou la nature de cette souche mammopathogène ne peut de nouveau pas encore être élucidé, de même que les différents avènements ayant conduit à sa formation. Il est intéressant d'ajouter également que, sur les six îlots mis en évidence chez *E. coli* P4, les données obtenues sur l'îlot vraisemblablement présent en 3' du locus *pheV* permettent d'affirmer que ce dernier, en particulier, est peut-être toujours mobile. En effet, au sein de cet îlot ont été identifiés, d'une part, un gène codant une intégrase, probablement fonctionnelle, vers l'extrémité 5' et, d'autre part une répétition directe des 23 derniers nucléotides du gène *pheV* vers l'extrémité 3'. Ces deux éléments peuvent dicter à eux seuls l'excision de l'îlot par un mécanisme de recombinaison site-spécifique. Ce phénomène a déjà été observé en effet pour les îlots de pathogénicité I_{536} et II_{536} de la souche *E. coli* uropathogène 536 (Blum *et al*., 1994). Il a été montré que la fréquence d'excision de ces deux îlots était de l'ordre de 10^{-6} pour l'îlot I_{536} et 10^{-4} pour l'îlot II_{536} lors du passage de la phase exponentielle à la phase stationnaire des bactéries cultivées à 20°C (Middendorf *et al*., 2001).

4.1.4. Cas du locus *thrW*

Comme pour le cas du locus *asnT*, aucune ORF complète n'a pu être détectée dans la région amplifiée et séquencée en 3' du locus *thrW* sur le génome d'*E. coli* P4. Cependant, les deux séquences des banques de données, appartenant aux souches *Sh. flexneri* 2a 301 et *E. coli* O157:H7 Sakaï et similaires à la séquence obtenue, renferment toutes deux la fin d'un gène codant une intégrase putative. Chez *Sh. flexneri* 2a 301, le gène codant l'intégrase putative est situé au sein d'une région génomique variable localisée en 3' d'un locus d'ARNt-Thr. Cette région contient des ORFs codant d'autres intégrases putatives, des transposases d'IS, des protéines hypothétiques de fonction inconnue et des protéines phagiques. Chez *E. coli* O157:H7 Sakaï, le gène codant l'intégrase est le premier gène du prophage Sp1 situé immédiatement en 3' du locus *thrW*. Ce prophage est retrouvé sur le génome de la souche *E. coli* EDL933 sous une forme quelque peu modifiée : trois ORFs sont absentes et une ORF n'est identique qu'à moitié à l'ORF correspondante de l'élément Sp1. La région génomique variable de *Sh. flexneri* 2a 301 et le prophage Sp1 présentent des contenus tout à fait différents, seules les ORFs codant les intégrases sont similaires, partageant 82,5 % d'identité. Sur toutes les autres ORFs de l'élément Sp1, seule une, codant une ADN-invertase, est retrouvée sur le génome de *Sh. flexneri* 2a 301 (à une position éloignée de 2,3 Mb environ de la région génomique variable citée précédémment), avec une conservation du taux d'identité nucléotidique de 91 %. Ainsi, aucune hypothèse ne peut être faite quant au contenu de la région génomique variable située en 3' du locus *thrW* sur le génome d'*E. coli* P4. L'analyse comparative réalisée avec les souches *E. coli* O157:H7 Sakaï et *Sh. flexneri* 2a 301 montre effectivement que des gènes codant des intégrases et identifiés sur les génomes de souches bactériennes différentes peuvent être similaires sans que les régions génomiques attenantes ne soient identiques, même partiellement.

4.1.5. Cas du locus *serU*

La séquence déterminée immédiatement en 3' du locus *serU* sur le génome d'*E. coli* P4 renferme cinq premières ORFs codant des protéines à fonction de mobilité, puis deux ORFs codant une protéine putative inconnue des banques de données et une protéine putative impliquée dans des réactions de glycosylation (figure 33). Les cinq premières ORFs à fonction de mobilité occupent près des trois quarts de la séquence établie et semblent présenter des origines très différentes. Ainsi, l'ORF 7-*serU* serait similaire à une ORF codant une intégrase retrouvée sur l'ADN d'un bactériophage ; l'ORF 6-*serU* est identique à une ORF codant une transposase d'IS retrouvée sur le chromosome de la souche entérotoxinogène *E. coli* E24377A ; l'ORF 5-*serU* est identique à une ORF codant une transposase retrouvée sur le plasmide F ; l'ORF 4-*serU* est identique à une ORF codant une protéine

hypothétique retrouvée sur le plasmide F ; et l'ORF 3-*serU* est similaire à une ORF codant une transposase d'IS retrouvée sur le chromosome de la souche commensale *E. coli* HS. Ce fait suggère que l'insertion de ces cinq ORFs sur le chromosome d'*E. coli* P4 ait eu lieu après différents évènements de recombinaison indépendants. Comme il l'a été dit lors de l'analyse du locus *leuX* d'*E. coli* P4, avec la suggestion d'un élément SpLE5 altéré, les réarrangements des îlots génomiques sont, en effet, des phénomènes non négligeables, de par leur contenu en enzymes de mobilité. De telles situations ont été rapportées pour l'îlot de pathogénicité *cag* des bactéries *Helicobacter pylori* (Akopyants *et al*., 1998), l'îlot de pathogénicité HPI des bactéries *Yersinia* (Buchrieser *et al*., 1998), l'îlot de pathogénicité LEE des bactéries *E. coli* entéropathogènes (Tauschek *et al*., 2002) et l'îlot de pathogénicité *she* des bactéries *Shigella* (Rajakumar *et al*., 1997 ; Sakellaris *et al*., 2004). L'ORF 2-*serU* d'*E. coli* P4 présente une séquence inconnue des banques de données. Enfin, l'ORF 1-*serU* code une protéine putative présentant une similarité avec une protéine impliquée dans la conversion de la partie O-antigène des lipopolysaccharides. Etant donné, cependant, le faible taux d'identité que partagent ces deux protéines (24 %) et l'absence d'un domaine conservé typique de la fonction de glycosylation, cette dernière ne peut être avancée avec certitude pour la protéine codée par l'ORF 1-*serU* d'*E. coli* P4. Bien que caractérisée que partiellement, la région génomique variable située en 3' du locus *serU* sur le génome d'*E. coli* P4 semble présenter une structure très chimérique et complexe. Son rôle dans le caractère de la souche ne peut être encore envisagé.

4.2. Analyse de l'îlot génomique *argW* entièrement caractérisé

4.2.1. Caractéristiques générales

Avant de discuter des paramètres propres à l'îlot *argW* d'*E. coli* P4, trois caractéristiques concernant l'îlot *argW* des souches *E. coli* en général peuvent être commentées. La première porte sur les limites de cet îlot. Si sont considérés comme bordures de l'îlot uniquement les gènes partagés par tous les génomes *E. coli* séquencés jusqu'à présent, ces limites sont les gènes *argW* et *dsdA*. Au contraire, si sont considérés comme bordures de l'îlot le gène d'ARNt et la répétition de l'extrémité 3' de celui-ci (créée lors de l'insertion génomique de l'îlot par un mécanisme de recombinaison site-spécifique), ces limites semblent être les gènes *argW* et *dsdC* dans le cas d'*E. coli* P4, tout comme dans les cas d'*E. coli* K12 MG1655, K12 W3110, APEC01, UTI89 et CFT073, mais pas comme dans les cas d'*E. coli* O157:H7 Sakaï, EDL933 et E24377A chez qui le gène *dsdC* ainsi que la première moitié du gène *dsdX* sont absents (*dsdC* et *dsdX* étant les deux gènes précédant le gène *dsdA* sur les génomes des autres souches). La deuxième caractéristique qui peut être commentée au sujet de l'îlot *argW* des souches *E. coli* est sa taille, qui varie de 6 à 100 kb (tableau 30).

Tableau 30 : Taille des îlots génomiques insérés en 3' du locus *argW* au sein du génome de différentes souches *E. coli*.

Souche	K12 MG1655	W3110	O157:H7 Sakaï	EDL933	CFT073	UTI89	APEC01	E24377	P4
Taille (kb)	12,8	12,8	14,3	14,3	14,8	43,1	100	5,8	11,2

La taille de l'îlot *argW* d'*E. coli* P4 (11,2 kb) s'inscrit dans cette fourchette et se rapproche de celle des îlots *argW* des souches *E. coli* K12 MG1655, K12 W3110, O157:H7 Sakaï, EDL933 et CFT073. Deux cas particuliers se présentent également : le cas du génome de la souche *E. coli* HS pour lequel aucune insertion d'ADN étranger n'est détectée entre les gènes *argW* et *dsdC* et le cas du génome de la souche *E. coli* 536 pour lequel les gènes *dsdC*, *dsdX* et *dsdA* sont retrouvés à une position très éloignée du gène *argW*. Le fait de retrouver le gène *dsdC* au sein du génome d'*E. coli* HS alors qu'aucun îlot génomique n'est présent en 3' du locus *argW* de cette souche soutient la deuxième hypothèse proposée précédemment au sujet des limites de l'îlot *argW* chez l'espèce *E. coli*, à savoir les gènes *argW* et *dsdC*. L'absence du gène *dsdC* et de la première moitié du gène *dsdX* au sein du génome des souches *E. coli* O157:H7 Sakaï, EDL933 et E24377A résulte probablement de la perte de ces régions lors de l'insertion génomique de l'îlot. La troisième et dernière caractéristique de l'îlot *argW* de l'espèce *E. coli* qui va être commentée ici, est son contenu. Des contenus très différents peuvent être observés entre les souches, comme l'illustre la figure 43.

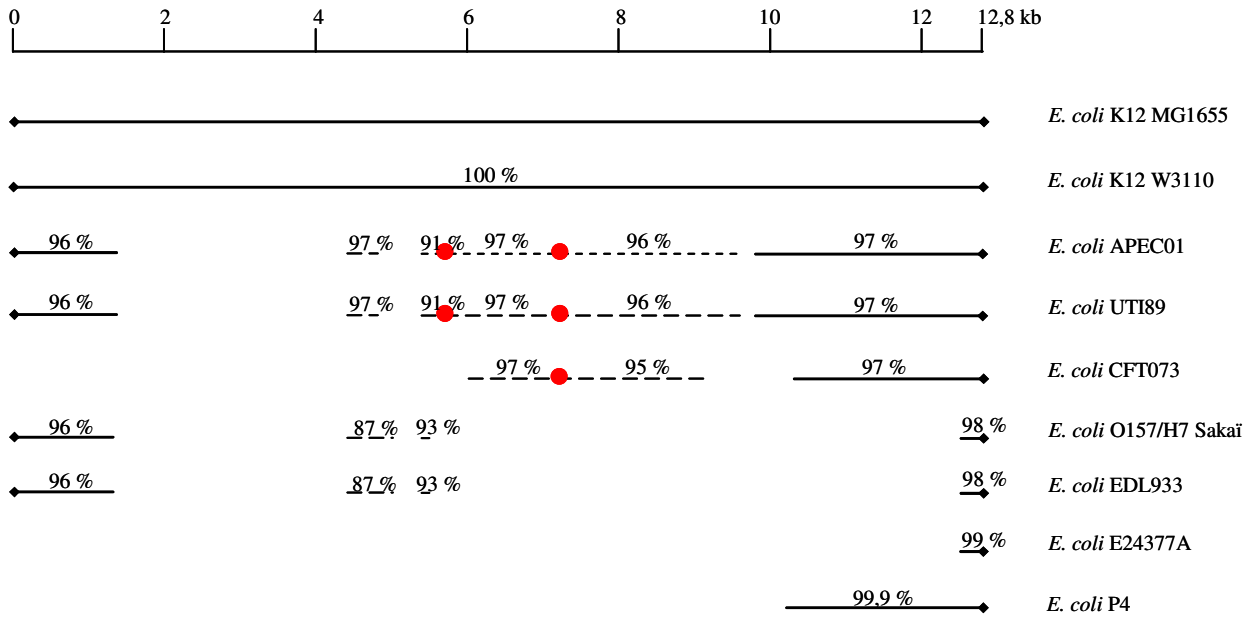


Figure 43 : Recherche de séquences nucléotidiques similaires à la séquence de l'îlot *argW* d'*E. coli* K12 MG1655 grâce au serveur blastn du site NCBI. Les traits en ligne continue indiquent la présence de séquences similaires, chez d'autres souches *E. coli,* localisées également au niveau de leur îlot *argW*. Les traits en ligne pointillée indiquent la présence de séquences similaires, chez d'autres souches *E. coli,* à des positions éloignées de leur locus *argW*, séquences qui sont interrompues par d'autres indiquées par des points rouges. Enfin, les pourcentages indiquent les taux d'identité nucléotidique entre la séquence de l'îlot *argW* d'*E. coli* K12 MG1655 et les séquences similaires chez les autres souches *E. coli*.

Cette dernière représente, de façon schématisée, le résultat de la recherche de séquences nucléotidiques similaires à la séquence de l'îlot *argW* de la souche *E. coli* K12 MG1655 grâce au serveur blastn du site NCBI. Elle montre clairement qu'au sein des génomes des souches pathogènes intestinales, telles que *E. coli* O157:H7 Sakaï, EDL933 et E24377A, très peu de séquences similaires à la séquence de l'îlot *argW* d'*E. coli* K12 MG1655 sont retrouvées. Dans les cas des souches *E. coli* O157:H7 Sakaï et EDL933, il s'agit de séquences homologues aux extrémités 5' et 3' de l'îlot, ainsi qu'à la fin de la première moitié de l'îlot. Dans le cas de la souche *E. coli* E24377A, il s'agit uniquement de séquences homologues à l'extrémité 3' de l'îlot. Dans les trois cas, les séquences homologues aux extrémités de l'îlot *argW* d'*E. coli* K12 MG1655 correspondent également aux extrémités des îlots *argW* des trois souches pathogènes intestinales. Par contre, les séquences d'*E. coli* O157:H7 Sakaï et EDL933 homologues à la fin de la première moitié de l'îlot *argW* d'*E. coli* K12 MG1655 correspondent à des séquences situées à une position très différente du locus *argW* dans leurs génomes. Au sein des génomes des souches pathogènes extra-intestinales, telles qu'*E. coli* CFT073, UTI89 et APEC01, une plus grande proportion de séquences similaires à la séquence de l'îlot *argW* d'*E. coli* K12 MG1655 est mise en évidence. Pour les trois souches, des séquences homologues à l'extrémité 5' et/ou 3' de l'îlot, ainsi qu'à de grandes régions internes de l'îlot, sont retrouvées. Comme précédemment, les séquences homologues aux extrémités 5' et 3' de l'îlot d'*E. coli* K12 MG1655 correspondent aux extrémités des îlots *argW* de ces trois souches pathogènes extra-intestinales et les séquences homologues aux régions internes de l'îlot d'*E. coli* K12 MG1655 correspondent par contre, à des séquences tout à fait indépendantes et éloignées du locus *argW* dans leurs génomes.

Bien qu'il comporte des ORFs pour lesquelles des ORFs similaires aient été trouvées dans les banques de données, l'îlot *argW* d'*E. coli* P4 peut être considéré comme un nouvel îlot au sein de l'espèce *E. coli*. En effet, comme le montre la figure 44, très peu de séquences similaires à la séquence nucléotidique de cet îlot sont identifiées au sein des génomes entièrement séquencés de l'espèce *E. coli* cités dans le tableau 10.

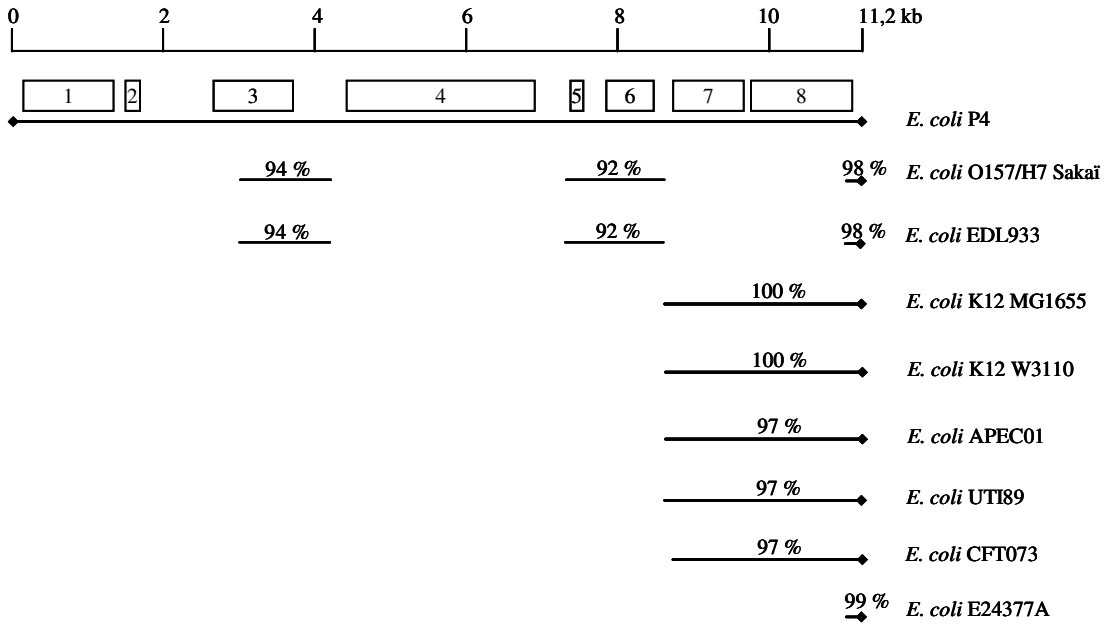


Figure 44 : Recherche de séquences nucléotidiques similaires à la séquence de l'îlot *argW* d'*E. coli* P4 grâce au serveur blastn du site NCBI. Les traits en ligne continue indiquent la présence de séquences similaires, chez d'autres souches *E. coli,* localisées également au niveau de leur îlot *argW*. Les pourcentages indiquent les taux d'identité nucléotidique entre la séquence de l'îlot *argW* d'*E. coli* P4 et les séquences similaires chez les autres souches *E. coli*. Les rectangles dessinés au dessus du trait représentant la séquence de l'îlot *argW* d'*E. coli* P4 correspondent aux différentes ORFs détectées au sein de cet îlot, décrites dans le tableau 25.

Au sein des génomes des souches *E. coli* commensales et pathogènes extra-intestinales (K12 MG1655, K12 W3110, APEC01, UTI89 et CFT073), des séquences similaires à l'extrémité 3' de l'îlot seulement sont retrouvées. Ces séquences incluent les ORFs 7-*argW* et 8-*argW* d'*E. coli* P4, ORFs également présentes à l'extrémité 3' des îlots *argW* de ces souches et correspondant aux gènes *dsdC* et *dsdX*, respectivement. Au sein du génome de la souche *E. coli* E24377A, seule une séquence de très petite taille (0,2 kb) est similaire à l'extrémité 3' de l'îlot d'*E. coli* P4, séquence correspondant à la fin du gène *dsdX*. Enfin, au sein des génomes des souches pathogènes intestinales *E. coli* O157:H7 Sakaï et EDL933, en plus de la séquence de petite taille présentée précédemment pour *E. coli* E24377A, deux séquences similaires à des fragments internes de l'îlot d'*E. coli* P4 peuvent être identifiées. Ces séquences se situent également au sein des îlots *argW* des deux souches.

Les îlots *argW* des souches *E. coli* O157:H7 Sakaï et EDL933 sont identiques. D'une taille de 14,3 kb, ils renferment quatorze ORFs codant successivement : une intégrase putative (similaire à l'intégrase du prophage Sf6-like), trois protéines faiblement similaires à des protéines de bactériophages (dont deux similaires à des protéines du bactériophage P22 nécessaires au transfert et à l'injection de l'ADN), une protéine hypothétique inconnue des banques de données, une protéine putative de réplication (faiblement similaire à une protéine de réplication du plasmide pVT736-1 d'une souche *Haemophilus actinomycetemcomitans*), trois protéines hypothétiques inconnues des banques de

données, une résolvase putative (similaire à une résolvase du plasmide pMT1 d'une souche *Yersinia pestis*) et quatre protéines impliquées dans le métabolisme de sucres (similaires à des protéines d'autres souches *E. coli*) (figure 45).

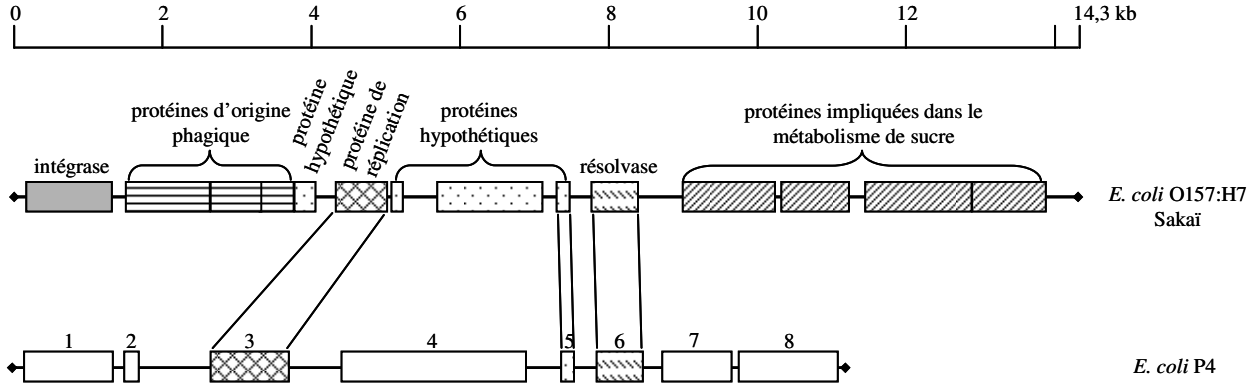


Figure 45 : Représentation schématique de l'îlot *argW* d'*E. coli* O157:H7 Sakaï et alignement avec l'îlot *argW* d'*E. coli* P4. Les chiffres inscrits au dessus des rectangles sur l'îlot d'*E. coli* P4 indiquent les numéros des ORFs de cet îlot, décrites dans le tableau 25. Les ORFs similaires entre les deux îlots sont rejointes par des traits verticaux.

Les fragments des îlots *argW* des souches *E. coli* O157:H7 Sakaï et EDL933 similaires à des portions de l'îlot *argW* d'*E. coli* P4 incluent l'ORF codant la protéine putative de réplication, l'ORF codant la résolvase putative et une des trois ORFs codant des protéines hypothétiques inconnues des banques de données séparant les deux précédentes (figure 45). Il semble donc que l'îlot d'*E. coli* P4 ait acquis une région de l'îlot d'*E. coli* O157:H7 Sakaï / EDL933 tout en remplaçant une partie de son contenu (ou inversement). En effet, la perte de deux ORFs codant des protéines hypothétiques inconnues des banques de données est compensée chez *E. coli* P4 par la présence de l'ORF 4-*argW* codant la protéine putative au domaine conservé Sir2 (figure 45).

Une des particularités de l'îlot *argW* d'*E. coli* P4 est son pourcentage en paires de bases guanine et cytosine dont la valeur, de l'ordre de 40 %, s'éloigne de près de 10 % de la valeur moyenne des génomes *E. coli* (50 %). Un tel écart a déjà été enregistré pour l'îlot LEE (Locus of Enterocyte Effacement) de certaines souches *E. coli* entérohemolytiques et entéropathogènes (Schmidt et Hensel, 2004). Ce plus faible pourcentage sous-entend pour les six ORFs propres à l'îlot *argW* d'*E. coli* P4 (ORF 1-*argW*, ORF 2-*argW*, ORF 3-*argW,* ORF 4-*argW,* ORF 5-*argW* et ORF 6-*argW*) une origine autre que l'espèce *E. coli*. L'ORF 1-*argW* est similaire à une ORF codant une intégrase au sein du génome de la souche *Salmonella enterica* ATCC9150. L'espèce *Salmonella enterica* appartient, comme *E. coli*, à la classe *Gamma Proteobacteria* et à l'ordre *Enterobacteriales* et présente des génomes aux pourcentages en paires de bases guanine et cytosine de l'ordre de 51-52 %. L'intégrase

de la souche *Salmonella enterica* ATCC9150 est elle-même similaire à l'intégrase du bactériophage P4. L'ORF 2-*argW* est similaire à une ORF codant un régulateur transcriptionnel au sein du génome d'une souche *Yersinia intermedia*, espèce appartenant également à la classe *Gamma Proteobacteria* et à l'ordre *Enterobacteriales* et présentant des génomes aux pourcentages en paires de bases guanine et cytosine de l'ordre de 47 %. L'ORF 3-*argW* est similaire à une ORF présente au sein de l'îlot *argW* d'*E. coli* O157:H7 Sakaï, elle-même similaire à une ORF codant une protéine de réplication localisée sur le plasmide pVT736-1 d'une souche *Haemophilus actinomycetemcomitans*, espèce appartenant à la classe *Gamma Proteobacteria* et à l'ordre *Pasteurellales* et présentant des génomes aux pourcentages en paires de bases guanine et cytosine de l'ordre de 37-38 %. L'ORF 4-*argW* est similaire à une ORF codant une protéine hypothétique au sein du génome de la souche *Rhodopseudomonas palustris*, espèce appartenant à la classe *Alpha Proteobacteria* et à l'ordre *Rhizobiales* et présentant des génomes aux pourcentages en paires de bases guanine et cytosine de l'ordre de 64-66 %. L'ORF 5-*argW* est similaire à une ORF présente au sein de l'îlot *argW* d'*E. coli* O157:H7 Sakaï pour qui aucune séquence similaire dans les banques de données n'est pour l'instant disponible. Enfin, l'ORF 6-*argW* est similaire à une ORF également présente au sein de l'îlot *argW* d'*E. coli* O157:H7 Sakaï, elle-même similaire à une ORF codant une résolvase localisée sur le plasmide pMT1 d'une souche *Yersinia pestis*, espèce appartenant à la classe *Gamma Proteobacteria* et à l'ordre *Enterobacteriales* et présentant des génomes aux pourcentages en paires de bases guanine et cytosine de l'ordre de 47 %. Ainsi, des similarités avec des ORFs de différentes origines sont effectivement mises en évidence pour les ORFs de l'îlot *argW* d'*E. coli* P4. Les différents évènements ayant conduit à l'assemblage de ces ORFs et à la création de l'îlot semblent donc difficiles à retracer. Etant donné que l'intégrase codée par l'ORF 1-*argW* ne semble plus fonctionnelle, cet îlot n'est probablement plus mobile et la structure découverte ici peut donc être considérée comme stable.

Pour conclure ce paragraphe, une hypothèse concernant l'apport de l'îlot *argW* dans le caractère ou la nature d'*E. coli* P4 peut être émise. Trois ORFs identifiées (les ORFs 2-*argW*, 3-*argW* et 4-*argW*) codent des protéines dont les domaines conservés ou les protéines similaires témoignent d'un rôle dans les processus de réplication et de transcription de l'ADN, ainsi que dans les processus de déacétylation des protéines. Ceci suggère donc, pour l'îlot *argW* d'*E. coli* P4, une fonction dans les processus de régulation ou de contrôle de l'expression et de l'activité de certaines protéines.

4.2.2. Hypothèses sur le rôle de la protéine codée par l'ORF 4-*argW*

L'ORF 4-*argW* détectée au sein de l'îlot *argW* d'*E. coli* P4 code une protéine putative présentant un domaine conservé de protéine Sir2. Les protéines Sir2, également appelées sirtuins, sont

des enzymes catalysant des réactions de déacétylation NAD^+ dépendantes (North et Verdin, 2004). Les processus d'acétylation / déacétylation des protéines régulent une grande variété de fonctions cellulaires, telles que les intéractions protéine / protéine, ADN / protéine et la stabilisation des protéines (Kouzarides, 2000). L'acétylation post-traductionnelle réversible des protéines a lieu sur les résidus lysine et est catalysée par les activités opposées des histones acétyltransférases (HATs) et des histones déacétylases (HDACs). Contrairement à ce que leurs noms indiquent, ces enzymes agissent aussi bien sur les protéines histones que sur d'autres substrats. Les HDACs ont été réparties dans trois classes différentes, I, II et III, selon leur degré de similarité avec des répresseurs transcriptionnels identifiés initialement chez la levure. Les protéines Sir2 appartiennent à la classe III des HDACs.

Le membre fondateur de la famille des protéines Sir2 a été découvert chez la levure *Saccharomyces cerevisiae*, au début des années 1990. Il s'agit d'un répresseur transcriptionnel appelé Sir2p, fixé au niveau de séquences répétées sur le chromosome, comme les télomères et les loci d'ADNr (Gottschling *et al*., 1990 ; Smith et Boeke, 1997). Depuis, plusieurs protéines similaires ont été identifiées chez les bactéries, les archées, les plantes, les invertébrés et les vertébrés (Frye, 2000). Il semble que quelques bactéries ne possèdent pas de sirtuins (dont certaines appartiennent notamment aux genres *Rickettsia*, *Borrelia*, *Chlamydia*, *Mycoplasma*, *Synechocystis*, *Neisseria*, *Treponema* et *Ureaplasma*), mais tous les organismes eucaryotes étudiés à ce jour en possèdent. En général, une seule protéine Sir2 est identifiée par organisme bactérien, à quelques exceptions près, comme le cas de la souche *Streptomyces coelicolor* A3 qui présente deux protéines Sir2. En revanche, plus d'une protéine Sir2 peuvent être identifiées par organisme eucaryote : la levure, la drosophile et l'homme, par exemple, présentent ainsi cinq, cinq et sept protéines Sir2, respectivement.

Une analyse phylogénétique de toutes les protéines Sir2 recensées a permis de les ordonner dans cinq classes différentes (I, II, III, IV et U) (Frye, 2000). Les sirtuins identifiées chez les bactéries appartiennent aux classes III et U, la classe III regroupant des sirtuins d'origine aussi bien bactérienne qu'eucaryote et la classe U regroupant uniquement des sirtuins de bactéries à Gram positif. Quelque soit leur classe, les sirtuins partagent quelques motifs conservés, regroupés dans un domaine core d'une taille de 206 acides aminés environ ; la longueur et le contenu de leurs extrémités N et C-terminales peuvent, au contraire, varier. Parmi les motifs conservés, sont relevés un cluster de résidus cystéine et un domaine riche en résidus leucine, suggérant tous deux la possibilité, pour les protéines Sir2, d'intéragir avec d'autres protéines et de se lier à l'ADN.

Deux activités enzymatiques propres aux protéines Sir2 ont pu être mises en évidence : une mineure d'ADP-ribosyltransférase et une prédominante de déacétylase, toutes deux requérant la présence de NAD^+ et d'un substrat acétylé (North et Verdin, 2004). Des études effectuées principalement chez la levure et l'homme ont montré que, de par leur activité déacétylase, les protéines Sir2 pouvaient exercer plusieurs rôles, notamment dans la répression de la transcription, la réparation de l'ADN

endommagé, la stabilité du chromosome, le retard du vieillissement cellulaire, la bonne progression du cycle cellulaire et le bon développement embryonnaire.

Les protéines d'origine bactérienne présentant un domaine conservé de protéine Sir2 sont encore peu étudiées. Au sein de l'espèce *E. coli*, une protéine Sir2 mise en évidence est la protéine CobB, qui catalyse la déacétylation et ainsi l'activation de l'acétyl-CoA synthétase (Zhao *et al*., 2004). La protéine codée par l'ORF 4-*argW* d'*E. coli* P4 ne présente aucune similarité avec la protéine CobB ni avec aucune autre protéine de l'espèce *E. coli*. Plusieurs rôles joués par cette protéine peuvent être envisagés, en relation avec les propriétés de virulence d'*E. coli* P4. Un premier rôle serait la répression de la transcription de certains gènes de virulence une fois l'infection ou la colonisation de la glande mammaire établie. Ce phénomène est en effet connu et très important chez les souches bactériennes pathogènes (Dagberg et Uhlin, 1992). Il permet aux bactéries ayant atteint leur site d'infection d'arrêter la synthèse et l'expression de facteurs de virulence, ceci dans plusieurs buts comme, par exemple : moins ou ne plus stimuler les défenses immunitaires hôtes pour se protéger de leurs actions bactéricides, arrêter les dommages cellulaires ou tissulaires causés à l'hôte pour pouvoir profiter des métabolismes de celui-ci et s'installer durablement en son sein. Un deuxième rôle que pourrait jouer la protéine codée par l'ORF 4-*argW* d'*E. coli* P4 serait en relation avec l'environnement très oxydant qui s'installe dans la glande mammaire lors d'une infection de type mammite. Il a été évoqué dans le paragraphe 3.1.1 de ce chapitre, le processus de flambée oxydative généré par les neutrophiles, aboutissant à la production de diverses entités oxygénées réactives (O_2^-, H_2O_2, OH^-, OCl^-). Ces entités, synthétisées dans le cytoplasme des neutrophiles, se retrouvent également dans l'espace extracellulaire et sont connues pour induire d'importantes altérations au niveau de constituants cellulaires essentiels tels que les protéines, l'ADN et les lipides membranaires. Il a été montré que l'activité des protéines Sir2 nécessitait et était stimulée par la présence de NAD^+ (Buck *et al*., 2004). Or, en milieu oxydant la concentration de ce cofacteur oxydé est susceptible d'augmenter. Ainsi, lors d'un stress oxydant, comme celui qui se produit pour les souches mammopathogènes lors de l'inflammation mammaire, les protéines Sir2 pourraient être activées dans le but de ralentir l'activité cellulaire par répression de la transcription de certains gènes et/ou de réparer les dommages que les entités oxygénées réactives pourraient causer sur le chromosome. Un troisième rôle que pourrait jouer la protéine codée par l'ORF 4-*argW* d'*E. coli* P4 serait lié à ses capacités d'invasion cellulaire. Il a été montré que certaines protéines Sir2 régulaient le passage d'une phase à une autre du cycle cellulaire, notamment l'entrée et la sortie de la mitose (Dryden *et al*., 2003). Cette régulation passe par la déacétylation des tubulines α, rendant la structure des microtubules plus lâches, plus flexibles (North *et al*., 2003). Ce phénomène serait intéressant et profitable à *E. coli* P4 pour son entrée dans les cellules épithéliales hôtes. Il nécessiterait cependant l'exposition de la protéine Sir2 en surface de la cellule bactérienne ou même sa sécrétion puis son internalisation par les cellules hôtes afin d'y exercer son rôle pour faciliter par la suite l'internalisation de la bactérie.

4.3. Conclusion

L'analyse de certains loci d'ARNt au sein du génome d'*E. coli* P4 a permis de révéler quelques caractéristiques distinguant cette souche des souches *E. coli* commensales non pathogènes. La caractérisation partielle de cinq d'entre eux, *asnT*, *leuX*, *pheV*, *serU* et *thrW*, a été initiée et devra être poursuivie pour détermminer leur apport fonctionnel. La caractérisation complète du locus *argW* a permis d'identifier une ORF codant une protéine porteuse d'un domaine conservé Sir2, dont il serait intéressant par la suite d'étudier l'expression et le rôle précis, en conditions de mammite. Tout comme chez les eucaryotes, des fonctions bénéfiques à la bactérie lui seraient en effet peut-être attribuées. Pour compléter ce criblage des loci d'ARNt, les contextes génomiques de 19 autres loci d'ARNt, figurant parmi les 36 loci connus pour accueillir des ADNs étrangers, mais non criblés au cours de ces travaux, pourrait être également analysés. Ce criblage global des loci d'ARNt, facile à mettre en œuvre, pourrait être mené sur le génome d'autres souches mammopathogènes. Ceci permettrait de vérifier si la présence d'îlots génomiques à ces endroits du génome est commune au sein de ce pathotype et si les mêmes îlots ou au contraire des îlots différents sont retrouvés chez différentes souches MPEC. Un contenu différent en îlots génomiques entre souches MPEC serait peut-être lié également à des degrés de virulence différents. Sachant enfin que les parties variables d'un génome bactérien ne sont pas uniquement concentrées en 3' des loci d'ARNt, mais peuvent se situer également ailleurs dans le génome, des études d'hybridations soustractives entre le génome d'*E. coli* P4 et les génomes de souches *E. coli* présentant des tropismes différents pourraient être également envisagées. Ces dernières permettraient, éventuellement, de découvrir des régions génomiques propres à la souche mammopathogène, peut-être à l'origine donc de la spécificité de sa pathogénicité. Bien sûr, le séquençage complet du génome d'*E. coli* P4 permettrait également de mettre en évidence ces régions propres.

Références bibliographiques

Abbott S.L., O'Connor J., Robin T., Zimmer B.L., and Janda J.M. (2003). Biochemical properties of a newly described *Escherichia* species, *Escherichia albertii. J. Clin. Microbiol.* **41**, 4852-4854.

Adiri R.S., Gophna U., and Ron E.Z. (2003). Multilocus sequence typing (MLST) of *Escherichia coli* O78 strains. *FEMS Microbiol. Lett*. **222**, 199-203.

Aguero M.E., Aron L., DeLuca A.G., Timmis K.N., and Cabello F.C. (1984). A plasmid-encoded outer membrane protein, TraT, enhances resistance of *Escherichia coli* to phagocytosis. *Infect. Immun.* **46**, 740-746.

Akopyants N.S., Clifton S.W., Kersulyte D., Crabtree J.E., Youree B.E., Reece C.A., Bukanov N.O., Drazek E.S., Roe B.A., and Berg D.E. (1998). Analyses of the *cag* pathogenicity island of *Helicobacter pylori. Mol. Microbiol.* 28, 37-53.

Allison G.E., Angeles D., Tran-Dinh N., and Verma N.K. (2002). Complete genomic sequence of SfV, a serotype-converting temperate bacteriophage of *Shigella flexneri*. *J. Bacteriol.* **184**, 1974-1987.

Amano M., Ohkusu K., Kusaba K., Ikeda H., Nagasawa Z., Aoki Y., Kawamura Y., Kobatake S., Tanaka T., Matsuura S., and Ezaki T. (2005). Quantitative microarray-based DNA-DNA hybridization assay for measuring genetic distances among bacterial species and its application to the identification of family *Enterobacteriaceae. Microbiol. Immunol.* **49**, 255-263.

Andrup L., Damgaard J., and Wassermann K. (1993). Mobilization of small plasmids in *Bacillus thuringiensis* subsp. *Israelensis* is accompanied by specific aggregation. *J. Bacteriol.* **175**, 6530-6536.

Anton A.I., Martinez-Murcia A.J., and Rodriguez-Valera F. (1998). Sequence diversity in the 16S-23S intergenic spacer region (ISR) of the rRNA operons in representatives of the *Escherichia coli* ECOR collection. *J. Mol. Evol*. **47**, 62-72.

Arakawa Y., Murakami M., Suzuki K., Ito H., Wacharotayankun R., Ohsuka S., Kato N., and Ohta M. (1995). A novel integron-like element carrying the metallo-β-lactamase gene bla_{IMP}. *Antimicrob. Agents Chemother.* **39**, 1612-1615.

Austin J.J., Smith A.B., and Thomas R.H. (1997). Paleontology in a molecular world: the search for authentic ancient DNA. *Trends Ecol. Evol.* **12**, 303-306.

Bach S., de Almeida A., and Carniel E. (2000). The *Yersinia* high-pathogenicity island is present in different members of the family Enterobacteriaceae. *FEMS Microbiol. Lett.* **183**, 289-294.

Banks D.J., Beres S.B., and Musser J.M. (2002). The fundamental contribution of phages to GAS evolution, genome diversification and strain emergence. *Trends Microbiol.* **10**, 515-521.

Barksdale L., and Arden S.B. (1974). Persisting bacteriophage infections, lysogeny, and phage conversions. *Annu. Rev. Microbiol.* **28**, 265-299.

Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P., Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J., Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L., Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H., Yeh K.C., Davis R.W., Federspiel N.A., and Long S.R. (2001). Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. *Proc. Natl. Acad. Sci. USA* **98**, 9883-9888.

Barrow P.A., and Hill A.W. (1989). The virulence characteristics of strains of *Escherichia coli* isolated from cases of bovine mastitis in England and Wales. *Vet. Microbiol.* **20**, 35-48.

Bauer F., Hertel C., and Hammes W.P. (1999). Transformation of *Escherichia coli* in foodstuffs. *Syst. Appl. Microbiol.* **22**, 161-168.

Bayles K.W., Wesson C.A., Liou L.E., Fox L.K., Bohach G.A., and Trumble W.R. (1998). Intracellular *Staphylococcus aureus* escapesthe endosome and induces apoptosis in epithelial cells. *Infect. Immun*. **66**, 336-342.

Beaber J.W., Hochhut B., and Waldor M.K. (2002). Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element drived from *Vibrio cholerae*. *J. Bacteriol.* **184**, 4259-4269.

Bean A., Williamson J., and Cursons R.T. (2004). Virulence genes of *Escherichia coli* strains isolated from mastitic milk. *J. Vet. Med.* **51**, 285-287.

Belaaouaj A., Kim K.S., and Shapiro S.D. (2000). Degradation of outer membrane protein A in *Escherichia coli* killing by neutrophil elastase. *Science* **289**, 1185-1188.

Bender C., Palmer D., Peñaloza-Vàzquez A., Rangaswamy V., and Ulrich M. (1996). Biosynthesis of cotonatine, a thermoregulated phytotoxin produced by the phytopathogen *Pseudomonas syringae. Arch. Microbiol.* **166**, 71-75.

Benz I., and Schmidt M.A. (1993). Diffuse adherence of enteropathogenic *Escherichia coli* strains, processing of AIDA-I. *Int. J. Med. Microbiol. Virol. Parasitol. Infect. Dis.* **278**, 197-208.

Berg T., Firth N., Apisiridej S., Hettiaratchi A., Leelaporn A., and Skurray R.A. (1998). Complete nucleotide sequence of pSK41: evolution of staphylococcal conjugative multiresistance plasmids. *J. Bacteriol*. **180**, 4350-4359.

Bertani L.E., and Six E.W. (1988). The P2-like phages and their parasite, P4. In R. Calendar (ed.), The bacteriophages, vol. 2, Plenum Publishing Corp., New York, pp.73-143.

Bertin Y., Martin C., Oswald E., and Girardeau J.P. (1996). Rapid and specific detection of F17-related pilin and adhesion genes in diarrheic and septicemic *Escherichia coli* strains by multiplex PCR. *J. Clin. Microbio*. **34**, 2921-2928.

Bertin Y., Martin C., Girardeau J.P., Pohl P., and Contrepois M. (1998). Association of genes encoding P fimbriae, CS31A antigen and EAST 1 toxin among CNF1-producing *Escherichia coli* strains from cattle with septicaemia and diarrhea. *FEMS Microbiol Lett*. **162**, 235-239.

Bettelheim K.A. (2002). The genus *Escherichia*. In : M. Dworkin *et al*., eds., The Prokaryotes: An evolving electronic resource for the microbiological community, 3rd edition, release3.9, 4/1/2002, Springer-Verlag, New York.

Bishop A.L., Baker S., Jenks S., Fookes M., Gaora P.O., Pickard D., Anjum M., Farrar J., Hien T.T., Ivens A., and Dougan G. (2005). Analysis of the hypervariable region of the *Salmonella enterica* genome associated with tRNA (leuX). *J. Bacteriol.* **187**, 2469-2482.

Bissonnette L., and Roy P.H. (1992). Characterization of In0 of *Pseudomonas aeruginosa* plasmid pVS1, an ancestor of integrons of multiresistance plasmids and transposons of Gram-Negative bacteria. *J. Bacteriol.* **174**, 1248-1257.

Blattner F. R., Plunkett III G., Bloch C. A., Perna N. T., Burland V., Riley M., Collado-Vides J., Glasner J. D., Rode C. K., Mayhew G. F., Gregor J., Davis N. W., Kirkpatrick H. A., Goeden M. A., Rose D. J., Mau B., and Shao Y. (1997). The complete genome sequence of *Escherichia coli* K12. *Science* **277**, 1453-1462.

Blowey R.W., and Edmondson P.W. (1995). Mastitis control in dairy herds. Ipswich, Farming Press, pp. 29.

Blum G., Ott M., Lischewski A., Ritter A., Imrich H., Tschäpe H., and Hacker J. (1994). Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infect. Immun.* **62**, 606-614.

Boutet P., Bureau F., and Lekeux P. (2006). La mammite bovine : de l'initiation à la résolution. *Ann. Méd. Vet.* **150**, 1-26.

Boyd D.A., Peters G.A., Ng L.K., and Mulvey M.R. (2000). Partial characterization of a genomic island associated with the multidrug resistance region of *Salmonella enterica* Typhimurium DT104. *FEMS Microbiol. Lett.* **189**, 285-291.

Boyd D.A., Peters G.A., Cloeckaert A., Boumedine K.S., Chaslus-Dancla E., Imberechts H., and Mulvey M.R. (2001). Complete nucleotide sequence of a 43-kilobase genomic island associated with the multidrug resistance region of *Salmonella enterica* serovar Typhimurium DT104 and its identification in phage type DT120 and serovar Agona. *J. Bacteriol.* **183**, 5725-5732.

Boyd E.F., and Hartl D. (1998). Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution. *J. Bacteriol.* **180**, 1159-1165.

Bradley A.J., and Green M.J. (2000). A study of the incidence and significance of intramammary enterobacterial infections acquired during the dry period. *J. Dairy Sci.* **83**, 1957-1965.

Bradley A.J., and Green M.J. (2001a). Aetiology of clinical mastitis in six Somerset dairy herds. *Vet. Record* **148**, 683-686.

Bradley A.J., and Green M.J. (2001b). Adaptation of *Escherichia coli* to the bovine mammary gland. *J. Clin. Microbiol.* **39**, 1845-1849.

Bradley A.J. (2002). Bovine mastitis: an evolving disease. *Vet. J.* **164**, 116-128.

Bramley A.J. (1976). Variations in the susceptibility of lactating and non-lactating bovine udders to infection when infused with *Escherichia coli*. *J. Dairy Res*. **43**, 205-211.

Bray J. (**1945**). Isolation of antigenically homogeneous strains of *Bact. coli neopolitanum* from summer diarrhoea of infants. *J. Pathol. Bacteriol*. 57, 239-247.

Brenner D.J., Davis B.R., Steigerwalt A.G., Riddle C., McWhorter A.C., Allen S.D., Farmer III J.J., Saitoh Y., and Fanning G.R. (1982a). Atypical biogroups of *Escherichia coli* found in clinical specimens and description of *Escherichia hermannii* sp. nov. *J. Clin. Microbiol.* **15**, 703-713.

Brenner D.J., McWhorter A.C., Leete Knutson J.K., and Steigerwalt A.G. (1982b). *Escherichia vulneris*: a new species of *Enterobactericeae* associated with human wounds. *J. Clin. Microbiol.* **15**, 1133-1140.

Brenner D.J. (1984). Facultatively anaerobic Gram-negative rods. In N.R. Krieg and J.G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1, Williams & Wilkins, Baltimore, Md., pp. 408-516.

Brunder W., Schmidt H., and Karch H. (1997). EspP, a novel extracellular serine protease of enterohemorrhagic *Escherichia coli* O157:H7 cleaves human coagulation factor V. *Mol. Microbiol.* **24**, 767-778.

Brüssow H., and Hendrix R.W. (2002). Phage genomics: small is beautiful. *Cell* **108**, 13-16.

Brüssow H., Canchaya C., and Hardt W.D. (2004). Phage and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev.* **68**, 560-602.

Buchrieser C., Prentice M., and Carniel E. (1998). The 102-kilobase unstable region of *Yersinia pestis* comprises a high-pathogenicity island linked to a pigmentation segment which undergoes internal rearrangement. *J. Bacteriol.* **180**, 2321-2329.

Buck S.W., Gallo C.M., and Smith J.S. (2004). Diversity in the Sir2 family of protein deacetylases. *J. Leukoc. Biol.* **75**, 939-950.

Burgess N.R.H., McDermott S.N., and Whiting J. (1973). Aerobis bacteria in the hind-gut of the cockroach, *Blatta orientalis. J. Hyg.* **71**, 1-7.

Burns A.L., Ball H.J., and Finlay D.A. (1996). CNF producing *Escherichia coli* isolated from cattle in Northern Ireland. *Vet. Microbiol.* **49**, 235-241.

Burvenich C., Van Merris V., Mehrzad J., Diez-Fraile A., and Duchateau L. (2003). Severity of *E. coli* mastitis is mainly determined by cow factors. *Vet. Res.* **34**, 521-564.

Burvenich C., Monfardini E., Mehrzad J., Capuco A.V., and Paape M.J. (2004). Role of neutrophil polymorphonuclear leukocytes during bovine coliform mastitis: physiology or pathology ? *Verh. K. Acad. Geneeskd. Belg.* **66**, 97-150.

Campbell A.M. (1992). Chromosomal insertion sites for phages and plasmids. *J. Bacteriol.* **174**, 7495-7499.

Cao T.B., and Saier M.H. (2001). Conjugal type IV macromolecular transfer systems of Gram-negative bacteria: organismal distribution, structural constraints and evolutionary conclusions. *Microbiology* **147**, 3201-3214.

Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J., Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S., Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D., Pohl T., Portetelle D., Puhler A., Purnelle B., Ramsperger U., Renard C., Thebault P., Vandenbol M., Weidner S., and Galibert F. (2001). Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9877-9882.

Carattoli A. (2001). Importance of integrons in the diffusion of resistance. *Vet. Res*. **32**, 243-259.

Carniel E., Guilvout I., and Prentice M. (1996). Characterization of a large chromosomal **'**high-pathogenicity island**'** in biotype 1B *Yersinia enterocolitica*. *J. Bacteriol.* **178**, 6743-6751.

Carroll E.J., and Jasper D.E. (1977). Bactericidal activity of standard bovine serum against coliform bacteria isolated from udders and the environment of dairy cows. *Am. J. Vet. Res.* **38**, 2019-2022.

Chamier B., Lorenz M.G., and Wackernagel W. (1993). Natural transformation of *Acinetobacter calcoaceticus* by plasmid DNA adsorbed on sand and groundwater aquifer material. *Appl. Environ. Microbiol.* **59**, 1662-1667.

China B., Pirson V., and Mainil J. (1996). Typing of bovine attaching and effacing *Escherichia coli* by multiplex in vitro amplification of virulence-associated genes. *Appl. Environ. Microbiol.* **62**, 3462-3465.

Cheng Q., Campbell E.A., Naughton A.M., Johnson S., and Masure H.R. (1997). The com locus controls genetic transformation in *Streptococcus pneumoniae. Mol. Microbiol.* **23**, 683-692.

Chopin A., Bolotin A., Sorokin A., Ehrlich S.D., and Chopin M. (2001). Analysis of six prophages in *Lactococcus lactis* IL1403: different genetic structure of temperate and virulent phage populations. *Nucleic Acids Res.* **29**, 644-651.

Claverys J.P., and Lacks S.A. (1986). Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. *Microbiol. Rev.* **50**, 133-165.

Clermont O., Boacorsi S., and Bingen E. (2000). Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl. Environ. Microbiol.* **66**, 4555-4558.

Clermont O., Bonacorsi S., and Bingen E. (2001). The Yersinia high-pathogenicity island is highly predominant in virulence-associated phylogenetic groups of *Escherichia coli*. *FEMS Microbiol. Lett.* **196**, 153-157.

Clewell D.B., Flannagan S.E., and Jaworski D.D. (1995). Unconstrained bacterial promiscuity: the Tn*916*-Tn*1545* family of conjugative transposons. *Trends Microbiol.* **3**, 229-236.

Clewell D.B., Francia M.V., Flannagan S.E., and An F.Y. (2002). Enterococcal plasmid transfer: sex pheromones, transfer origins, relaxases, and the *Staphylococcus aureus* issue. *Plasmid* **48**, 193-201.

Cloeckaert A., Praud K., Doublet B., Demartin M., and Weill F.X. (2006). Variant *Salmonella* genomic island 1-L antibiotic resistance gene cluster in *Salmonella enterica* serovar Newport. *Antimicrob. Agents Chemother.* **50**, 3944-3946.

Cohen E., Kessler D., and Levine H. (2005). Recombinaison dramatically speeds up evolution of finite populations. *Phys. Rev. Lett.* **94**, 098102.

Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R., Wheeler P.R., Honoré N., Garnier T., Churcher C., Harris D., Mungall K., Basham D., Brown D., Chillingworth T., Connor R., Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N., Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S., Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M., Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R., and Barrell B.G. (2001). Massive gene decay in the leprosy bacillus. *Nature* **409**, 1007-1011.

Copley S.D. (2000). Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach. *Trends Biochem. Sci.* **25**, 261-265.

Correa M.G.P., and Marin J.M. (2002). O-serogroups, *eae* gene and EAF plasmid in *Escherichia coli* isolates from cases of bovine mastitis in Brazil. *Vet. Microbiol.* **85**, 125-132.

Crosa J.H., and Walsh C.T. (2002). Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiol. Mol. Biol. Rev.* **66,** 223-249.

Cvitkovitch D.G. (2001). Genetic competence and transformation in oral streptococci. *Crit. Rev. Oral Bio. Med.* **12**, 217-243.

Czeczulin J.R., Whittam T.S., Henderson I.R., Navarro-Garcia F., and Nataro J.P. (1999). Phylogenetic analysis of enteroaggregative and diffusely adherent *Escherichia coli*. *Infect. Immun.* **67**, 2692-2699.

Dagberg B., and Uhlin B.E. (1992). Regulation of virulence-associated plasmid genes in enteroinvasive *Escherichia coli*. *J. Bacteriol.* **174**, 7606-7612.

Dalrymphe B.P. (1982). Plasmid encoded DNA primases. Ph.D. thesis, University of Leicester, Leicester, England.

Datta N. (1979). Plasmid classification: incompatibility grouping. In K.N. Timmis and A. Pühler (ed.), Plasmids of medical, environmental and commercial importance. Elseviers/North Holland Publishing Co., Amsterdam, pp. 3-12.

Demaneche S., Kay E., Gourbiere F., and Simonet P. (2001a). Natural transformation of *Pseudomonas fluorescens* and *Agrobacterium tumefaciens* in soil. *Appl. Environ. Microbiol.* **67**, 2617-2621.

Demaneche S., Bertolla F., Buret F., Nalin R., Sailland A., Auriol P., Vogel T.M., and Simonet P. (2001b). Laboratory-scale evidence for lightning-mediated gene transfer in soil. *Appl. Environ. Microbiol.* **67**, 3440-3444.

Desjardins P., Picard B., Kaltenböck B., Elion J., and Denamur E. (1995). Sex in *Escherichia coli* does not disrupt the clonal structure of the population: evidence from random amplified polymorphic DNA and restriction-fragment lenght polymorphism. *J. Mol. Evol.* **41**, 440-448.

Deurenberg R.H., Vink C., Kalenic S., Friedrich A.W., Bruggeman C.A., and Stobberingh E.E. (2006). The molecular evolution of methicillin-resistant *Staphylococcus aureus. Clin. Microbiol. Infect.* **13**, 222-235.

de Vries J., and Wackernagel W. (2004). Microbial horizontal gene transfer and the DNA release from transgenic crop plants. *Plant Soil* **266**, 91-104.

Dezfulian H., Batisson I., Fairbrother J.M., Lau P.C.K., Nassar A., Szatmari G., and Harel J. (2003). Presence and characterization of extraintestinal pathogenic *Escherichia coli* virulence genes in F165-positive *E. coli* strains isolated from diseased calves and pigs. *J. Clin. Microbiol*. **41**, 1375-1385.

Dogan B., Klaessig S., Rishniw M., Almeida R.A., Olivier S.P., Simpson K., and Schukken Y.H. (2006). Adherent and invasive *Escherichia coli* are associated with persistent bovine mastitis. *Vet. Microbiol.* **116**, 270-282.

Doolittle R.F., Feng D.F., Tsang S., Cho G., and Little E. (1996). Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* **271**, 470-477.

Döpfer D., Barkema H.W., Lam T.J.G.M., Schukken Y.H., and Gaastra W. (1999). Recurrent clinical mastitis caused by *Escherichia coli* in dairy cows. *J. Dairy Sci.* **82**, 80-85.

Döpfer D., Almeida R.A., Lam T.J.G.M., Nederbragt H., Olivier S.P., and Gaastra W. (2000). Adhesion and invasion of *Escherichia coli* from single and recurrent cases of bovine mastitis *in vitro*. *Vet. Microbiol.* **74**, 331-343.

Döpfer D., Nederbragt H., Almeida R.A., and Gaastra W. (2001). Studies about the mechanism of internalization by mammary epithelial cells of *Escherichia coli* isolated from persistent bovine mastitis. *Vet. Microbiol.* **80**, 285-296.

Doublet B., Boyd D.A., Mulvey M.R., and Cloeckaert A. (2005). The *Salmonella* genomic island 1 is an integrative mobilizable element. *Mol. Microbiol.* **55**, 1911-1924.

Dougan G., Haque A., Pickard D., Frankel G., O'Goara P., and Wain J. (2001). The *Eshcerichia coli* gene pool. *Curr. Opin. Microbiol.* **4**, 90-94.

Dougherty B.A., Hill C., Weidman J.F., Richardson D.R., Venter J.C., and Ross R.P. (1998). Sequence and analysis of the 60 kb conjugative, bacteriocin-producing plasmid pMRC01 from Lactococcus lactis DPC3147. *Mol. Microbiol*. **29**, 1029-1038.

Dreiseikelmann B. (1994). Translocation of DNA across bacterial membranes. *Microbiol*. *Rev*. **58**, 293-316.

Dryden S.C., Nahhas F.A., Nowak J.E., Goustin A.S., and Tainsky M.A. (2003). Role for human SIRT2 NAD-dependant deacetylase activity in control of mitotic exit in the cell cycle. *Mol. Cell. Biol.* **23**, 3173-3185.

Dutta, P. R., Capello, R., Navarro-Garcia, F., and Nataro, J. P. (2002). Functional comparison of serine protease autotransporters of enterobacteriaceae. *Infect. Immun.*, **70**, 7105-7113.

Escherich T. (1885). Die Darmbakterien des Neugeborenen und Säuglings. *Fortschr. Med.* **3**, 515-522.

Escobar-Paramo P., Clermont O., Blanc-Potard A.B., Bui H., Le Bouguenec C., and Denamur E. (2004). A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli. Mol. Biol. Evol.* **21**, 1085-1094.

Evans E., and Alani E. (2000). Roles for mismatch repair factors in regulating genetic recombination. *Mol. Cell. Biol*. **20**, 7839-7844.

Ewald J.H., Heesemann J., Rudiger H., and Autenrieth I.B. (1994). Interaction of polymorphonuclear leukocytes with *Yersinia enterocolitica*: role of the *Yersinia* virulence plasmid and modulation by the iron-chelator desferrioxamine B. *J. Infect. Dis.* **170**, 140-150.

Fang W., and Pyörälä S. (1996). Mastitis-causing *Escherichia coli*: serum sensitivity and susceptibility to selected antibacterials in milk. *J. Dairy Sci.* **79**, 76-82.

Farmer III J.J., Fanning G.R., Davis B.R., O'Hara C.M., Riddle C., Hickman-Brenner F.W., Asbury M.A., Lowery III V.A., and Brenner D.J. (1985). *Escherichia fergusonii* and *Enterobacter taylorae*, two new species of *Enterobacteriaceae* isolated from clinical specimens. *J. Clin. Microbiol.* **21**, 77-81.

Farr S.B., and Kogoma T. (1991). Oxidative stress responses in *Escherichia coli* and *Salmonella typhimurium*. *Microbiol. Rev.* **55**, 561-585.

Feng G., Tsui H.C.T., and Winkler M.E. (1996). Depletion of the cellular amounts of the MutS and MutH methyl-directed mismatch repair proteins in stationary-phase *Escherichia coli* K12 cells. *J. Bacteriol.* **178**, 2388-2396.

Ffrench-Constant R., Waterfield N., Daborn P., Joyce S., Bennett H., Au C., Dowling A., Boundy S., Reynolds S., and Clarke D. (2003). *Photorhabdus:* towards a functional genomic analysis of a symbiont and pathogen. *FEMS Microbiol. Rev.* **26**, 433-456.

Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P., Vorhölter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J., Golding B., and Pühler A. (2001).Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P., Vorhölter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J., Golding B., and Pühler A. (2001). The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endosymbiont *Sinorhizobium meliloti*. *Proc. Natl. Acad. Sci. USA* **98,** 9889-9894.

Finkel S.E., and Kolter R. (2001). DNA as a nutrient: a novel role for bacterial competence gene homologs. *J. Bacteriol.* **183**, 6288-6293.

Fitzgerald J.R., Meaney W.J., Hartigan P.J., Smyth C.J., and Kapur V. (1997). Fine structure molecular epidemiological analysis of *Staphylococcus aureus* recovered from cows. *Epidemiol. Infect.* **119**, 261-269.

Fong K.P.Y., Goh C.B.H., and Tan H.M. (2000). The genes for benzene catabolism in *Pseudomonas putida* ML2 are flanked by two copies of the insertion element IS*1489*, forming a class-I-type catabolic transposon, Tn*5542*. *Plasmid* **43**, 103-110.

Forst S., and Nealson K. (1996). Molecular biology of the symbiotic-pathogenic bacteria *Xenorhabdus* spp. and *Photorhabdus* spp. *Microbiol. Rev.* **60**, 21-43.

Forterre P., Gribaldo S. et Brochier C. (2005). Luca : à la recherche du plus proche ancêtre commun universel. *Med. Sci.* **21**, 860-865.

Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult,C.J., Kerlavage A.R., Sutton G., Kelley J.M., Fritchman R.D., Weidman J.F., SmallK.V., Sandusky M., Fuhrmann J., Nguyen D., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M., Tomb J.F., Dougherty B.A., Bott K.F., Hu P.C., Lucier T.S., Peterson S.N., Smith H.O., Hutchison C.A. III, and Venter J.C. (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397-403.

Freeman V.J. (1951). Studies on the virulence of bacteriophage infected strains of *Corynebacterium diphtheriae. J. Bacteriol.* **61**, 675-688.

Freiberg C., Fellay R., Bairoch A., Broughton W.J., Rosenthal A., and Perret X. (1997). Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature* **387**, 394-401.

Frost L.S., Ippen-Ihler K., and Skurray R.A. (1994). Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol. Rev.* **58**, 162-210.

Frye RA. (2000). Phylogenetic classification of prokaryotic and eukaryotic Sir2-like proteins. *Biochem Biophys Res Commun*. **273**, 793-798.

Fukushima M., Kakinuma K., and Kawaguchi R. (2002). Phylogenetic anlysis of *Salmonella*, *Shigella* and *Escherichia coli* strains on the basis of the *gyrB* gene sequence. *J. Clin. Microbiol.* **40**, 2779-2785.

Gaillard M., Vallaeys T., Vorhölter F.J., Minoia M., Werlen C., Sentchilo V., Pühler A., and van der Meer J.R. (2006). The clc element of *Pseudomonas* sp. strain B13, a genomic island with various catabolic properties. *J. Bacteriol*. **188**, 1999-2013.

Garcia-Martinez J., Martinez-Murcia A.J., Rodriguez-Valera F., and Zorraquino A. (1996). Molecular evidence supporting the existence of two major groups in uropathogenic *Escherichia coli*. *FEMS Immunol. Med. Microbiol.* **14**, 231-244.

Germon P., Roche D., Melo S., Mignon-Grasteau S., Dobrindt U., Hacker J., Schouler C., and Moulin-Schouleur M. (2007). tDNA locus polymorphism and ecto-chromosomal DNA insertion hot-spots are related to the phylogenetic group of *Escherichia coli* strains. *Microbiology* **153**, 826-837.

Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L., Msadek T., Zouine M., Couvé E., Lalioui L., Poyart C., Trieu-Cuot P., and Kunst F. (2002). Genome sequence of *Streptococcus agalactiae*, a pathogen causive invasive neonatal disease. *Mol. Microbiol.* **45**, 1499-1513.

Gioia D.D., Peel M., Fava F., and Wyndham R.C. (1998). Structures of homologous composite transposons carrying *cbaABC* genes from Europe and North America. *Appl. Environ. Microbiol.* **64**, 1940-1946.

Gonen E., Vallon-Eberhard A., Elazar S., Harmelin A., Brenner O., Rosenshine I., Jung S., and Shpigel N.Y. (2007). Toll-like receptor 4 is needed to restrict the invasion of *Escherichia coli* P4 into mammary gland epithelial cells in a murine model of acute mastitis. *Cell. Microbiol*. **9**, 2826-2838.

González V., Santamaría R.I., Bustos P., Hernández-González I., Medrano-Soto A., Moreno-Hagelsieb G., Janga S.C., Ramírez M.A., Jiménez-Jacinto V., Collado-Vides J., and Dávila G. (2006). The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. *Proc. Natl. Acad. Sci. U S A*. **103**, 3834-383.

Gottesman M.E., and Weisberg R.A. (2004). Little lambda, who made thee? *Microbiol. Mol. Biol. Rev.* **68**, 796-813.

Gottschling D.E., Aparicio O.M., Billington B.L., and Zakian V.A. (1990). Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell* **63**, 751-762.

Greated A., Lambersten L., Williams P.A., and Thomas C.M. (2002). Complete sequence of the IncP-9 TOL plasmid pWWO from *Pseudomonas putida. Environ. Microbiol.* **4**, 856-871.

Griffiths E., Chart H., and Stevenson P. (1988). High-affinity iron uptake systems and bacterial virulence. In : Roth J.A. (Ed.), Virulence mechanisms of bacterial pathogens. ASM Press : Washington, D.C., 121-137.

Griffiths E. (1994). Iron acquisition systems in *Escherichia coli*. In : Gyles C.L. (Ed.), *Escherichia coli* in domestic animals and humans. CAB International : Wallingford, 533-566.

Griffiths E. (1997). Iron and the virulence of *Escherichia coli*. In : Sussman M. (Ed.), *Escherichia coli* : mechanisms of virulence. Cambridge University Press : Cambridge, 331-371.

Griffith F. (1928). The significance of pneumococcal types. *J. Hyg.* **27**, 113-159.

Grindley N.D.F., Whiteson K.L., and Rice P.A. (2006). Mechanisms of site-specific recombination. *Annu. Rev. Biochem.* **75**, 567-605.

Grohmann E., Muth G., and Espinosa M. (2003). Conjugative plasmid transfer in Gram positive bacteria. *Microbiol. Mol. Biol. Rev.* **67**, 277-301.

Grozdanov L., Raasch C., Schulze J., Sonnenborn U., Gottschalk G., Hacker J., and Dobrindt U. (2004). Analysis of the genome structure of the nonpathogenic probiotic *Escherichia coli* strain Nissle 1917. *J. Bacteriol.* **186**, 5432-5441.

Grys T.E., Siegel M.B., Lathem W.W., and Welch R.A. (2005). The StcE protease contributes to intimate adherence of enterohemorrhagic *Escherichia coli* O157:H7 to host cells. *Infect. Immun.* **73**, 1295-1303.

Guyer D.M., Kao J.S., and Mobley H.L. (1998). Genomic analysis of a pathogenicity island in uropathogenic *Escherichia coli* CFT073: distribution of homologous sequences among isolates from patients with pyelonephritis, cystitis, and catheter-associated bacteriuria and from fecal samples. *Infect. Immun.* **66**, 4411-4417.

Haase J., Lurz R., Grahn A.M., Bamford D.H., and Lanka E. (1995). Bacterial conjugation mediated by plasmid RP4: RSF1010 mobilization, donor-specific phage propagation, and pilus production require the same Tra2 core components of a proposed DNA transport complex. *J. Bacteriol.* **177**, 4779-4791.

Hacker J., Bender L., Ott M., Wingender J., Lund B., Marre R., and Goebel W. (1990). Deletions of chromosomal regions coding for fimbriae and hemolysins occur *in vitro* and *in vivo* in various extraintextinal *Escherichia coli* isolates. *Microb. Pathog.* **8**, 213-225.

Hacker J., Blum-Oehler G., Mühldorfer I., and Tschäpe H. (1997). Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol. Microbiol.* **23**, 1089-1097.

Hacker J., and Kaper J.B. (2000). Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* **54**, 641-679.

Haddadi K., Prin-Mathieu C., Moussaoui F., Faure G. C., Vangroenweghe F., Burvenich C., and Le Roux Y. (2006). Polymorphonuclear neutrophils and *Escherichia coli* proteases involved in proteolysis of casein during experimental *E. coli* mastitis. *Int. Dairy J.* **16**, 639-647.

Hamann J., and Krömker V. (1997). Potential of specific milk composition variables for cow health management. *Livestock Prod. Sci.* **48**, 201-208.

Hancock R.E.W., and Piers K. (1994). Outer membrane proteins. In Gyles C.L. (Ed.), *Escherichia coli* in domestic animals and humans. CAB International : Wallingford, 495-532.

Hansson K., Sundström L., Pelletier A., and Roy P.H. (2002). IntI2 integron integrase in Tn*7*. *J. Bacteriol.* **184**, 1712-1721.

Harmon R.J. (1994). Symposium: Mastitis and genetic evaluation for somatic cell count. Physiology of mastitis and factors affecting somatic cell counts. *J. Dairy Sci.* **77**, 2103-2112.

Hayashi K., Morooka N., Yamamoto Y., Fujita K., Isono K., Choi S., Ohtsubo E., Baba T., Wanner B.L., Mori H., and Horiuchi T. (2006). Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol. Syst. Biol*. **2**, 1-5.

Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C. G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T., Kuhara S., Shiba T., Hattori M., and Shinagawa H. (2001). Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res*. **8**, 11-22.

Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A., Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D., Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P., McDonald L., Utterback T., Fleishmann R.D., Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C., and Fraser C.M. (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**, 477-483.

Heinemann J.A., and Sprague G.F. (1989). Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature* **340**, 205-209.

Hejnova J., Pages D., Rusniok C., Glaser P., Sebo P., and Buchrieser C. (2006). Specific regions of genome plasticity and genetic diversity of the commensal *Escherichia coli* A0 34/86. *Int. J. Med. Microbiol.* **296**, 541-546.

Henderson I.R., Navarro-Garcia F., and Nataro J.P. (1998). The great escape: structure and function of the autotransporter proteins. *Trends Microbiol*. **6,** 370-378.

Henderson I.R., Czeczulin J., Eslava C., Noriega F., and Nataro J.P. (1999). Characterization of pic, a secreted protease of *Shigella flexneri* and enteroaggregative *Escherichia coli*. *Infect. Immun.* **67**, 5587-5596.

Herron-Olson L., Fitzgerald J.R., Musser J.M., and Kapur V. (2007). Molecular correlates of host specialization in *Staphylococcus aureus*. *PloS ONE* **10**, 1-13.

Hinnebusch J., and Tilly K. (1993). Linear plasmids and chromosomes in bacteria. *Mol. Microbiol.* **10**, 917-922.

Hiramatsu K., Chui L., Kuroda M., and Ito T. (2001). The emergence and evolution of methicillin resistant *Staphylococcus aureus*. *Trends Microbiol.* **9**, 486-493.

Hochhut B., Jahreis K., Lengeler J.W., and Schmid K. (1997). CTn*scr94*, a conjugative transposon found in enterobacteria. *J. Bacteriol.* **179**, 2097-2102.

Hochhut B., and Waldor M.K. (1999). Site-specific integration of the conjugal *Vibrio cholerae* SXT element into *prfC*. *Mol. Microbiol.* **32**, 99-110.

Hochhut B., Lotfi Y., Mazel D., Faruque S.M., Woodgate R., and Waldor M.K. (2001). Molecular analysis of the antibiotic resistance gene clusters in the *Vibrio cholerae* O139 and O1 SXT constins. *Antimicrob. Agents Chemother* **45**, 2991-3000.

Hou Y.M. (1999). Transfer RNAs and pathogenicity islands. *Trends Biochem. Sci.* **24**, 295-298.

Hu J., Kan B., Liu Z.H., and Yu S.Y. (2005). Enteroaggragative *Escherichia coli* isolated from Chinese diarrhea patients with high-pathogenicity island of Yersinia is involved in synthesis of siderophore yersiniabactin. *World J. Gastroenterol.* **11**, 5816-5820.

Huang D.B., Mohanty A., DuPont H.L., Okhuysen P.C., and Chiang T. (2006). A review of an emerging enteric pathogen: enteroaggragative *Escherichia coli*. *J. Med. Microbiol.* **55**, 1303-1311.

Hull S. (1997). *Escherichia coli* lipopolysaccharide in pathogenesis and virulence. In : Sussman M. (Ed.), *Escherichia coli* : mechanisms of virulence. Cambridge University Press : Cambridge, 145-167.

Huys G., Cnockaert M., Janda J.M., and Swings J. (2003). *Escherichia albertii* sp. Nov., a diarrhoeagenic species isolated from stool specimens of Bangladeshi children. *Int. J. Syst. Evol. Microbiol.* **53**, 807-810.

Israel D.A. (2001). Genetic exchange. In : Mobley H., Mendz G., and Hazell S. (Ed.), *Helicobacter pylori*: Physiology and genetics. ASM Press, Washington, D.C. chapter 28.

Jiang S.C., and Paul J.H. (1998). Gene transfer by transduction in the marine environment. *Appl. Environ. Microbiol.* **64**, 2780-2787.

Johnson J.T., Kariyawasam S., Wannemuehler Y., Mangiamele P., Johnson S.J., Doetkott C., Skyberg J.A., Lynne A.M., Johnson J.R., and Nolan L.K. (2007). The genome sequence of avian pathogenic *Escherichia coli* strain O1:K1:H7 shares strong similarities with human extraintestinal pathogenic *E. coli* genomes. *J. Bacteriol.* **189**, 3228-3236.

Johnson J.R. (1991). Virulence factors in *Escherichia coli* urinary tract infection. *Clin. Microbiol. Rev.* **4**, 80-128.

Johnson J.R. (2000). *Shigella* and *Escherichia coli* at the crossroads: Machiavellian masqueraders or taxonomic treachery? *J. Med. Microbiol.* **49**, 583-585.

Jores J., Wagner S., Rumer L., Eichberg J., Laturnus C., Kirsch P., Schierack P., Tschäpe H., and Wieler L.H. (2005). Description of a 111-kb pathogenicity island (PAI) encoding various virulence features in the enterohemorrhagic *E. coli* (EHEC) strain RW1374 (O103:H2) and detection of a similar PAI in other EHEC strains of serotype 0103:H2. *Int. J. Med. Microbiol*. **294**, 417-425.

Kaartinen L., Salonen E.M., Vaheri A., and Sandholm M. (1988). Activation of plasmin in mastitic milk. *Acta. Vet. Scand*. **29**, 485-491.

Kahn M.E., Barany F., and. Smith H.O. (1983). Transformasomes: specialized membranous structures that protect DNA during *Haemophilus* transformation. *Proc. Natl. Acad. Sci. USA*. **80**, 6927–6931.

Kaipainen T., Pohjanvirta T., Shpigel N.Y., Shwimmer A., Pyörälä S., and Pelkonen S. (2002). Virulence factors of *Escherichia coli* isolated from bovine clinical mastitis. *Vet. Microbiol.* **85**, 37-46.

Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S., Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T., Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M., Takeuchi C., Yamada M., and Tabata S. (2000). Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res*. **7**, 331-338.

Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T., Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K., Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M., and Tabata S. (2002). Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res.* **9**, 189-197.

Kaper J.B., Nataro J.P., and Mobley H.L.T. (2004). Pathogenic *Escherichia coli. Nat. Rev. Microbiol.* **2**, 123-140.

Keil H., Keil S., Pickup R.W., and Williams P.A. (1985). Evolutionary conservation of genes coding for *meta* pathway enzymes within TOL plasmids pWW0 and pWW53. *J. Bacteriol.* **164**, 887-895.

Konstantinidis K.T., and Tiedje J.M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. USA* **102**, 2567-2572.

Konstantinidis K.T., Ramette A., and Tiedje J.M. (2006). The bacterial species definition in the genomic era. *Phil. Trans. R. Soc. B* **361**, 1929-1940.

Kouzarides T. (2000). Acetylation: a regulatory modification to rival phosphorylation? *EMBO J.* **19**, 1176-1179.

Kuhnert P., Hacker J., Mühldorfer I., Burnens A.P., Nicolet J., and Frey J. (1997). Detection system for *Escherichia coli*-specific virulence genes: absence of virulence determinants in B and C strains. *Appl. Environ. Microbiol.* **63**, 703-709.

Lalioui L., Jouve M., Gounon P., and Le Bouguenec C. (1999). Molecular cloning and characterization of the afa-7 and afa-8 gene clusters encoding afimbrial adhesins in *Escherichia coli* strains associayed with diarrhea or septicaemia in calves. *Infect. Immun*. **67**, 5048-5059.

Lammers A., Vorstenbosch C.J., Erkens J.H.F., and Smith H.E. (2001). The major bovine mastitis pathogens have different cell tropisms in cultures of bovine mammary gland cells. *Vet. Microbiol.* **80**, 255-265.

Landy A. (1989). Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu. Rev. Biochem*. **58**, 913-949.

Lanka E., and Barth P.T. (1981). Plasmid RP4 specifies a deoxyribonucleic acid primase involved in its conjugal transfer and maintenance. *J. Bacteriol.* **148**, 769-781.

Lawrence J.G., and Ochman H. (1997). Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* **44**, 383-397.

Lawrence J.G., and Roth J.R. (1999). Genomic flux: genome evolution by gene loss and acquisition. In: R.L. Charlebois (Ed.), Organization of the prokaryotic genome, American Society for Microbiology, Washington, pp. 263-288.

Lecointre G., Rachdi L., Darlu P., and Denamur E. (1998). *Escherichia coli* molecular phylogeny using the incongruence lenght difference test. *Mol. Biol. Evol.* **15**, 1685-1695.

Lee C.Y., and Iandolo J.J. (1986). Integration of staphylococcal phage L54a occurs by site-specific recombination: structural analysis of the attachment sites. *Proc. Natl. Acad. Sci. USA* **83**, 5475-5478.

Lee H.Y., and Côté J.C. (2006). Phylogenetic analysis of γ-proteobacteria inferred from nucleotide sequence comparaison of the house-keeping genes *adk*, *aroE* and *gdh*: comparaisons with phylogeny inferred from 16S rRNA gene sequences. *J. Gen. Appl. Microbiol.* **52**, 147-158.

Le Roux Y., Colin O., and Laurent F. (1995). Proteolysis in samples of quarter milk with varying somatic cell counts. 1. Comparaison of some indicators of endogeneous proteolysis in milk. *J. Dairy Sci*. **78**, 1289-1297.

Leying H., Suerbaum S., Kroll H.P., Stahl D., and Opferkuch W. (1990). The capsular polysaccharide is a major determinant of serum resistance in K1-positive blood culture isolates of *Escherichia coli*. *Infect. Immun.* **58**, 222-227.

Liebert C.A., Hall R.M., and Summers A.O. (1999). Transposon Tn*21*, flagship of the floating genome. *Microbiol. Mol. Biol. Rev.* **63**, 507-522.

Lillehaug D., and Birkeland N.K. (1993). Characterization of genetic elements required for site-specific integration of the temperate lactococcal bacteriophage ϕLC3 and construction of integration-negative ϕLC3 mutants. *J. Bacteriol.* **175**, 1745-1755.

Lipman L.J.A., de Nijs A., Lam T.J.G.M., and Gaastra W. (1995). Identification of *Escherichia coli* strains from cows with clinical mastitis by serotyping and DNA polymorphism patterns with REP and ERIC primers. *Vet. Microbiol.* **43**, 13-19.

Lira W.M., Macedo C., and Marin J.M. (2004). The incidence of Shiga toxin-producing *Escherichia coli* in cattle with mastitis in Brazil. *J. Appl. Microbiol.* **97**, 861-866.

Lloyd A.L., Rasko D.A., and Mobley H.L.T. (2007). Defining genomic islands and uropathogen-specific genes in uropathogenic *Escherichia coli. J. Bacteriol.* **189**, 3532-3546.

Łobocka M.B., Rose D.J., Plunkett G. 3rd, Rusin M., Samojedny A., Lehnherr H., Yarmolinsky M.B., and Blattner F.R. (2004). Genome of bacteriophage P1. *J. Bacteriol*. **186**, 7032-68.

Lodinova R., Jouja V., and Lanc A. (1967). Influence of the intestinal flora on the development of immune reactions in infants. *J. Bacteriol.* **93**, 797-800.

Lodinova R., Jouja V., Vinsova N., Vocel J., and Melkova J. (1980). New attempts and possibilities in prevention and treatment of intestinal coli-infections in infants. *Czech Med.* **3**, 47-58.

Lodinova-Zadnikova R., Tlaskalova-Hogenova H., and Sonnenborn U. (1992). Local and serum antibody response in full-term and premature infants after artificial colonization of the intestine with the *E. coli* strain Nissle 1917 (Mutaflor). *Pediatr. Allergy Immunol.* **3**, 43-48.

Lodinova-Zadnikova R., and Sonnenborn U. (1997). Effect of preventive administration of a non-pathogenic *Escherichia coli* strain on the colonization of the intestine with microbial pathogens in newborn infants. *Biol. Neonate* **71**, 224-232.

Lohuis J.A., Van Leeuwen W., Verheijden J.H., Smit J.A., Brand A., and Van Miert A.S. (1988). Growth of *Escherichia coli* in whole and skim milk from endotoxin-induced mastitic quarters: in vitro effects of deferoxamine, zinc, and iron supplementation. *J. Dairy Sci.* **71**, 2772-2781.

Long E., Capuco A.V., Wood D.L., Sonstegard T., Tomita G., Paape M.J., and Zhao X. (2001). *Escherichia coli* induces apoptosis and proliferation of mammary cells. *Cell Death Diffe*r. **8**, 808-816.

Lorenz M.G., Aardema B.W., and Wackernagel W. (1988). Highly efficient genetic transformation of *Bacillus subtilis* attached to sand grains. *J. Gen. Microbiol.* **134**, 107-112.

Lorenz M.G., and Wackernagel W. (1990). Natural genetic transformation of *Pseudomonas stutzeri* by sand-adsorbed DNA. *Arch. Microbiol.* **154**, 380-385.

Lorenz M.G., and Wackernagel W. (1994). Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol. Rev.* **58**, 563-602.

MacLean A.M., Finan T.M., and Sadowsky M.J. (2007). Genomes of the symbiotic nitrogen-fixing bacteria of legumes. *Plant Physiol.* **144**, 615-622.

Maeda K., Nojiri H., Shinatani M., Yoshida T., Habe H., and Omori T. (2003). Complete nucleotide sequence of carbazole/dioxin-degrading plasmid pCAR1 in *Pseudomonas resinovorans* strain CA10 indicates its mosaicity and the presence of large catabolic transposon Tn*4676. J. Mol. Biol.* **326**, 21-33.

Mahillon J., and Chandler M. (1998). Insertion sequences. *Microbiol. Mol. Biol. Rev.* **62**, 725-774.

Mainil J., Gerardin J., and Jacquemin E. (2000). Identification of the F17 fimbrial subunit- and adhesin-encoding (*f17A* and *f17G*) gene variants in necrotoxigenic *Escherichia coli* isolates from cattles, pigs and humans. *Vet. Microbiol.* **73**, 327-335.

Mainil J. (2003). Facteurs de virulence et propriétés spécifiques des souches invasives d'*Escherichia coli* : I) Les adhésines et facteurs de colonisation. *Ann. Med. Vet.* **147**, 105-126.

Majewski J., and Cohan F.M. (1998). The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus. Genetics* **148**, 13-18.

Majewski J., Zawadzki P., Pickerill P., Cohan F.M., and Dowson C.G. (2000). Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation. *J. Bacteriol.* **182**, 1016-1023.

Manoil C., and Rosenbusch J.P. (1982). Conjugation-deficient mutants of *Escherichia coli* distinguish classes of functions of the outer membrane OmpA protein. *Mol. Gen. Genet.* **187**, 148-156.

Maroncle N.M., Sivick K.E., Brady R., Stokes F.E., and Mobley H.L. (2006). Protease activity, secretion, cell entry, cytotoxicity, and cellular targets of secreted autotransporter toxin of uropathogenic *Escherichia coli*. *Infect. Immun.* **74**, 6124-6134.

Martins A., and Cunha M.L.R.S. (2007). Methicillin resistance in *Staphylococcus aureus* and coagulase-negative staphylococci: epidemiological and molecular aspects. *Microbiol. Immunol.* **51**, 787-795.

May E.W., and Craig N.L. (1996). Switching from cut-and-paste to replicative Tn7 transposition. *Science* **272**, 401-404.

Mazodier P., and Davies J. (1991). Gene transfer between distantly related bacteria. *Annu. Rev. Genet.* **25**, 147-171.

McDaniel T.K., Jarvis K.G., Donnenberg M.S., and Kaper J.B. (1995). A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc. Natl. Acad. Sci. USA* **92**, 1664-1668.

Mellies J.L., Barron A.M.S., and Carmona A.M. (2007). Enteropathogenic and enterohemorrhagic *Escherichia coli* virulence gene regulation. *Infect. Immun.* **75**, 4199-4210.

Michelutti I., Le Roux Y., Rainard P., Poutrel B., and Laurent F. (1999). Cinétiques des variations de la composition protéique du lait au cours d'une mammite expérimentale à *Escherichia coli*. *Le Lait* **79**, 535-549.

Middendorf B., Blum-Oehler G., Dobrindt U., Mühldorfer I., Salge S., and Hacker J. (2001). The pathogenicity islands (PAIs) of the uropathogenic *Escherichia coli* strain 536: island probing of PAI II_{536}. *J. Infect. Dis.* **183**, 17-20.

Mills S., McAuliffe O.E., Coffey A., Fitzgerald G.F., and Ross R.P. (2006). Plasmids of lactococci - genetic accessories or genetic necessities? *FEMS Microbiol. Rev*. **30**, 243-273.

Mochizuki S., Hiratsu K., Suwa M., Ishii T., Sugino F., Yamada K., and Kinashi H. (2003). The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. *Mol Microbiol*. **48**, 1501-1510.

Moe D., Garbarsch C., and Kirkeby S. (1994). The protein effect on determination of DNA with Hoechst 33258. *J. Biochem. Biophys. Methods* **28**, 263-276.

Moll A., Manning P.A., and Timmis K.N. (1980). Plasmid-determined resistance to serum bactericidal activity: a major outer membrane protein, the *traT* gene product, is responsible for plasmid-specified serum resistance in *Escherichia coli. Infect. Immun.* **28**, 359-367.

Montenegro M.A., Bitter-Suemann D., Timmins J.K., Aguero M.E., Cabello F.C., Sanyal S.C., and Timmins K.N. (1985). *traT* gene sequences, serum resistance and pathogenicity-related factors in clinical isolates of *Escherichia coli* and other Gram-negative bacteria. *J. Gen. Microbiol.* **131**, 1511-1521.

Mortier-Barrière I., de Saizieu A., Claverys J.P., and Martin B. (1998). Competence-specific induction of *rec*A is required for full recombination proficiency during transformation in *Streptococcus pneumoniae*. *Mol. Microbiol.* **27**, 159-170.

Morton T.M., Eaton D.M., Johnston J.L., and Archer G.L. (1993). DNA sequence and units of transcription of the conjugative transfer gene complex (trs) of *Staphylococcus aureus* plasmid pGO1. *J. Bacteriol*. **175**, 4436-4447.

Moussaoui F., Michelutti F., Le Roux Y., and Laurent F. (2002). Mechanisms involved in milk endogenous proteolysis induced by a lipopolysaccharide experimental mastitis. *J. Dairy Sci.*, **85**, 2562-2570.

Moussaoui F., Vangroenweghe F., Haddadi K., Le Roux Y., Laurent F., Duchateau L., and Burvenich C. (2004). Proteolysis in milk during experimental *Escherichia coli* mastitis. *J. Dairy Sci*. **87**, 2923-2931.

Mulvey M.A., Schilling J.D., and Hultgren S.J. (2001). Establishment of a persistent *Escherichia coli* reservoir during the acute phase of a bladder infection. *Infect. Immun.* **69**, 4572-4579.

Mulvey M.R., Boyd D.A., Olson A.B., Doublet B., and Cloeckaert A. (2006). The genetics of *Salmonella* genomic island 1. *Microbes Infect.* **8**, 1915-1922.

Munro G.L., Grieve P.A., and Kitchen B.J. (1984). Effect of mastitis of milk yield, milk composition, processing properties and yield and quality of milk product. *Austr. J. Dairy Techn*. **39**, 7-16.

Murinda S.E., Nguyen L.T., Landers T.L., Draughon F.A., Mathew A.G., Hogan J.S., Smith K.L., Hancock D.D., and Oliver S.P. (2004). Comparaison of *Escherichia coli* isolates from humans, food, and farm and companion animals for presence of Shiga toxin-producing *E. coli* virulence markers. *Foodborne Pathog. Dis.* **1**, 178-184.

Nakabachi A., Yamashita A., Toh H., Ishikawa H., Dunbar H., Moran N., and Hattori M. (2006). The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* **314**, 259-260.

Nakaya R., Nakamura A., and Murata Y. (1960). Resistance transfer agents in *Shigella*. *Biochem. Biophys. Res. Commun.* **3**, 654-659.

Nandakumar M.P., Cheung A. and Marten M.R. (2006). Proteomic analysis of extracellular proteins from *Escherichia coli* W3110. *J. Proteome Res.* **5**, 1155-1161.

Nascimento A.M.A., and Chartone-Souza E. (2003). Operon *mer*: bacterial resistance to mercury and potential for bioremediation of contaminated environments. *Genet. Mol. Res.* **2**, 92-101.

Nataro J.P., Scaletsky I.C., Kaper J.B., Levine M.M., and Trabulsi L.R. (1985). Plasmids-mediated factors conferring diffuse and localized adherence of enteropathogenic *Escherichia coli*. *Infect. Immun.* **48**, 378-383.

Nataro J.P., Kaper J.B., Robins-Browne R., Prado V., Vial P., and Levine M.M. (1987). Patterns of adherence of diarrheagenic *Escherichia coli* to Hep-2 cells. *Pediatr. Infect. Dis. J.* **6**, 829-831.

Nataro J.P., and Kaper J.B. (1998). Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.* **11**, 142-201.

Navarro-Garcia F., Canizalez-Roman A., Sui B.Q., Nataro J.P., and Azamar Y. (2004). The serine protease motif of EspC from enteropathogenic *Escherichia coli* produces epithelial damage by a mechanism different from that of Pet toxin from enteroaggregative *E. coli. Infect. Immun.* **72**, 3609-3621.

Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A., McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M., Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D., Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O., Salzberg S.L., Smith H.O., Venter J.C., and Fraser C.M. (1999). Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323-329.

Nemeth J., Muckle C.A., and Lo R.Y. (1991). Serum resistance and the *trat* gene in bovine mastitis-causing *Escherichia coli* isolates. *Vet. Microbiol.* **28**, 343-351.

Nemeth J., Muckle C.A., and Gyles C.L. (1994). In vitro comparaison of bovine mastitis and fecal *Escherichia coli* isolates. *Vet. Microbiol.* **40**, 231-238.

Nesbo C.L., Dlutek M., and Doolittle W.F. (2006). Recombination in thermotoga: implications for species concepts and biogeography. *Genetics* **172**,759-769.

Nickerson S.C., and Pankey J.W. (1984). Neutrophils migration through teat and tissues of bovine mammary quarters experimentally challanged with *Staphylococcus aureus. J. Dairy Sci*. **67**, 826-834.

Nikaido H. (1994). Outer membrane. In: Neidhardt F.C., Ingraham J.L., Low K.B., Magasanik B., Umbarger H.E. (Eds.), *Escherichia coli* and *Salmonella* Typhimurium : cellular and molecular biology. American Society for Microbiology : Washington, 29-47.

Nizet V., Beall B., Bast D.J., Datta V., Kilburn L., Low D.E., and De Azavedo J.C. (2000). Genetic locus for streptolysin S production by group A streptococcus. *Infect. Immun.* **68**, 4245-4254.

Nojiri H., Shintani M., and Omori T. (2004). Divergence of mobile genetic elements involved in the distribution of xenobiotic-catabolic capacity. *Appl. Microbiol. Biotechnol.* **64**, 154-174.

Normark B.H., and Normark S. (2002). Evolution and spread of antibiotic resistance. *J. Intern. Med*. **252**, 91-106.

North B.J., Marshall B.L., Borra M.T., Denu J.M., and Verdin E. (2003). The human Sir2 ortholog, SIRT2, is an NAD^+-dependent tubulin deacetylase. *Mol. Cell.* **11**, 437-444.

North B.J., and Verdin E. (2004). Sirtuins: Sir2-related NAD-dependent protein deacetylases. *Genome Biol.* **5**, article 224.

Novick R.P. (1987). Plasmid incompatibility. *Microbiol. Rev.* **51**, 381-395.

Nowicki B., Selvarangan R., and Nowicki S. (2001). Family of *Escherichia coli* Dr adhesins: decay-accelerating factor receptor recognition and invasiveness. *J. Infect. Dis.* **183**, S23-S27.

Nunes-Duby S.E., Kwon H.J., Tirumalai R.S., Ellenberger T., and Landy A. (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.* **26**, 391-406.

Ochman H., and Wilson A.C. (1987). Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J. Mol. Evol.* **26**, 74-86.

Ogawa N., and Miyashita K. (1999). The chlorocatechol-catabolic transposon Tn*5707* of *Alcaligenes eutrophus* NH9, carrying a gene cluster highly homologous to that in the 1,2,4-trichlorobenzene-degrading bacterium *Pseudomonas* sp. Strain P51, confers the ability to grow on 3-chlorobenzoate. *Appl. Environ. Microbiol.* **65**, 724-731.

Ohnishi M., Kurokawa K., and Hayashi T. (2001). Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? *Trends in Microbiol.* **9**, 481-485.

Otto B.R., van Dooren S.J., Nuijens J.H., Luirink J., and Oudega B. (1998). Characterization of a hemoglobin protease by the pathogenic *Escherichia coli* strain EB1. *J. Exp. Med.* **188**, 1091-1103.

Ou H.Y., Chen L.L., Lonnen J., Chaudhuri R.R., Thani A.B., Smith R., Garton N.J., Hinton J., Pallen M., Barer M.R., and Rajakumar K. (2006). A novel strategy for the identification of genomic islands by comparative analysis of the contents and contexts of tRNA sites in closely related bacteria. *Nucleic Acids Res*. **34**, 1-11.

Paape M.J., Wergin W.P., Guidry A.J., and Schultze W.D. (1981). Pagocytic defense of the ruminant mammary gland. *Adv. Exp. Med. Biol.* **137**, 555-578.

Pansegrau W., Lanka E., Barth P.T., Figurski D.H., Guiney D.G., Haas D., Helinski D. R., Schwab H., Stanisich V.A., and Thomas C.M. (1994). Complete nucleotide sequence of Birmingham IncPa plasmids: compilation and comparative analysis. *J. Mol. Biol*. **239**, 623–663.

Parreira V.R., and Gyles C.L. (2003). A novel pathogenicity island integrated adjacent to the *thrW* tRNA gene of avian pathogenic *Escherichia coli* encodes a vacuolating autotransporter toxin. *Infect Immun*. **71**, 5087-5096.

Passey S., Bradley A., and Mellor H. (2008). *Escherichia coli* isolated from bovine pastitis invade mammary cells by a modified endocytic pathway. *Vet. Microbiol.* **130**, 151-164.

Patel S.K., Dotson J., Allen K.P., and Fleckenstein J.M. (2004). Identification and molecular characterization of EatA, an autotransporter protein of enterotoxigenic *Escherichia coli*. *Infect. Immun.* **72**, 1786-1794.

Perna N.T., Plunkett G. 3rd, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Pósfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.D., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., and Blattner F.R. (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **409**, 529-533.

Petsch D., and Anspach F.B. (2000). Endotoxin removal from protein solutions. *J. Biotechnol*. **76**, 97-119.

Picard B., Garcia J.S., Gouriou S., Duriez P., Brahimi N., Bingen E., Elion J., and Denamur E. (1999). The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect. Immun.* **67**, 546-553.

Poh R.P.C., Smith A.R.W., and Bruce I.J. (2002). Complete characterization of Tn*5530* from *Burkholderia cepacia* strain 2a (pIJB1) and studies of 2,4-dichlorophenoxyacetate uptake by the organism. *Plasmid* **48**, 1-12.

Polard P., and Chandler M. (1995). Bacterial transposases and retroviral integrases. *Mol. Microbiol.* **15**, 13-23.

Possoz C., Ribard C., Gagnat J., Pernodet J.L., and Guérineau M. (2001). The integrative element of pSAM2 from *Streptomyces*: kinetics and mode of conjugal transfer. *Mol. Microbiol.* **42**, 159-166.

Pradella S., Hans A., Spröer C., Reichenbach H., and Beyer S. (2002). Characterization, genome size and genetic manipulation of the myxobacterium *Sorangium cellulosum* So ce56. *Arch. Microbiol.* **178**, 484-492.

Puyet A., Greenberg B., and Lacks S.A. (1990). Genetic and structural characterization of *end*A. A membrane-bound nuclease required for transformation of *Streptococcus pneumoniae*. *J. Mol. Biol.* **213**, 727-738.

Qadri F., Svennerholm A.M., Faruque A.S.G., and Sack R.B. (2005). Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clin. Microbiol. Rev.* **18**, 465-483.

Radström P., Sköld O., Swedberg G., Flensburg J., Roy P.H., and Sundström L. (1994). Transposon Tn*5090* of plasmid R751, which carries an integron, is related to Tn*7*, Mu, and the retroelements. *J. Bacteriol.* **176**, 3257-3268.

Rainard P., and Riollet C. (2006). Innate immunity of the bovine mammary gland. *Vet. Res.* **37**, 369-400.

Rajakumar K., Sasakawa C., and Adler B. (1997). Use of a novel approach, termed island probing, identifies the *Shigella flexneri she* pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins. *Infect. Immun.* **65**, 4606-4614.

Rakin A., Noelting C., Schubert S., and Heesemann J. (1999). Common and specific characteristics of the high-pathogenicity island of *Yersinia enterocolitica. Infect. Immun.* **67**, 5265-5274.

Raynaud S., Vernozy-Rozand C., Boscher P., Picant P., Mathieu B., Degand C., Poutrel B., Heuchel V., and Chatelin Y.M. (2005). Shiga-toxin producing *Escherichia coli* (STEC) prevalence, origin, circulation and persistence in French dairy herds. *Renc. Rech. Ruminants* **12**, 379-382.

Rayssiguier C., Thaler D.S., and Radman M. (1989). The barrier to recombination between Escherichia coli and Salmonella typhimurium is disrupted in mismatch-repair mutants. *Nature* **342**, 396-401.

Rayssiguier C., Dohet C., and Radman M. (1991). Interspecific recombination between *Escherichia coli* and *Salmonella typhimurium* occurs by the RecABCD pathway. *Biochimie* **73**, 371-374.

Recchia G.D., and Hall R.M. (1995). Gene cassettes: a new class of mobile element. *Microbiology* **141**, 3015-3027.

Redfield R.J. (2001). Do bacteria have sex? *Nat. Rev. Genet.* **2**, 634-639.

Redford P., and Welch R.A. (2002). Extraintestinal *Escherichia coli* as a model system for the study of pathogenicity islands. *Curr. Top. Microbiol. Immunol.* **264**, 15-30.

Reid S.D., Herbelin C., Bumbaugh A.C., Selander R.K., and Whittam T.S. (2002). Multilocus sequence typing (MLST) of pathogenic *Escherichia coli*. http://www.shigatox.net/stec/mlst-new/

Ritter A., Gally D.L., Olsen P.B., Dobrindt U., Friedrich A., Klemm P., and Hacker J. (1997). The Pai-associated *leu*X specific tRNA5(Leu) affects type 1 fimbriation in pathogenic *Escherichia coli* by control of FimB recombinase expression. *Mol. Microbiol.* **25**, 871-882.

Robins-Browne R., and Hartland E.L. (2002). Advances in pediatric gastroenterology and hepatology. *Escherichia coli* as a cause of diarrhea. *J. Gastroenterol. Hepatol*. **17**, 467-475.

Rosselló-Mora R., and Amann R. (2001). The species concept for prokaryotes. FEMS Microbiol. Rev. **25**, 39-67.

Russo T.A. (2002). Capsule and lipopolysaccharide. In : Donnenberg M.S. (Ed.), *Escherichia coli* : virulence mechanisms of a versatile pathogen. Academic Press : London, 379-404.

Sakellaris H., Luck S.N., Al-Hasani K., Rajakumar K., Turner S.A., and Alder B. (2004). Regulated site-specific recombination of the *she* pathogenicity island of *Shigella flexneri. Mol. Microbiol.* **52**, 1329-1336.

Salyers A.A., Shoemaker N.B., Stevens A.M., and Li L.Y. (1995). Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiol. Rev*. **59**, 579-590.

Samuels A.L., Lanka E., and Davies J.E. (2000). Conjugative junctions in RP4-mediated mating of *Escherichia coli. J. Bacteriol.* **182**, 2709-2715.

Sanchez-Carlo V., McDonald J.S., and Packer R.A. (1984). Virulence factors of *Escherichia coli* isolated from cows with acute mastitis. *Am. J. Vet. Res.* **45**, 1775-1777.

Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K., Yoshino C., Horino A., Shiba T., Sasaki T., and Hattori M. (2002). The complete genomic sequence of *Mycoplasma penetrans*, an intracellular bacterial pathogen in humans. *Nucleic Acids Res*. **30**, 5293-5300.

Scaletsky I.C., Silva M.L., and Trabulsi L.R. (1984). Distinctive patterns of adherence of enteropathogenic *Escherichia coli* to HeLa cells. *Infect. Immun.* **45**, 534-536.

Schepers A.J., Lam T.J.G.M., Schukken Y.H., Wilmink J.B.M., and Hanekamp W.J.A. (1997). Estimation of variance components of somatic cells counts to determine thresholds for uninfected quarters. *J. Dairy Sci.* **80**, 1833-1840.

Schilling J.D., Lorenz R.G., and Hultgren S.J. (2002). Effect of trimethoprim-sulfamethoxazole on recurrent bacteriuria and bacterial persistence in mice infected with uropathogenic *Escherichia coli*. *Infect. Immun.* **70**, 7042-7049.

Schmid K., Schupfner M., and Schmitt R. (1982). Plasmid-mediated uptake and metabolism of sucrose by *Escherichia coli* K12. *J. Bacteriol.* **151**, 68-76.

Schmidt H., and Hensel M. (2004). Pathogenicity islands in bacterial pathogenesis. *Clin. Microbiol. Rev.* **17**, 14-56.

Schubert S., Rakin A., Karch H., Carniel E., and Heesemann J. (1998). Prevalence of the 'high-pathogenicity island' of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. *Infect. Immun.* **66**, 480-485.

Schubert S., Rakin A., Fischer D., Sorsa J., and Heesemann J. (1999). Characterization of the integration site of Yersinia high-pathogenicity island in *Escherichia coli*. *FEMS Microbiol. Lett.* **179**, 409-414.

Schwarz F.V., Perreten V., and Teuber M. (2001). Sequence of the 50-kb conjugative multiresistance plasmid pRE25 from *Enterococcus faecalis* RE25. *Plasmid.* **46**, 170-87.

Seegers H., Menard J.L., and Fournichon C. (1997). Mammites en élevage bovin laitier : importance actuelle, épidémiologie et plans de prévention. *Renc. Rech. Ruminants* **4**, 233-242.

Selander R.K., Caugant D.A., Ochman H., Mussser J.M., Gilmour M.N., and Whittam T.S. (1986). Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl. Environ. Microbiol.* **51**, 873-884.

Selsted M.E., Tangy Q., Morris W.L., McGuire P.A., Novotny M.J., Smith W., Henschen A.H., and Cullor J.S. (1993). Purification, primary structures, and antibacterial activities of beta-defensins, a new family of antimicrobial peptides from bovine neutrophils. *J. Biol. Chem.* **268**, 6641-6648.

Shiga K. (1898). Ueber den Dysenterie-bacillus (*Bacillus dysenteriae*). *Zentralbl. Bakteriol. Orig.* **24**, 913-918.

Shuster D.E., Harmon R.J., Jackson J.A., and Hemken R.W. (1991). Suppression of milk production during endotoxin-induced mastitis. *J. Dairy Sci*. **74**, 3763-3775.

Silver S., and Ji G. (1994). Newer systems for bacterial resistances to toxic heavy metals. *Environ. Health Perspect.* **102**, 107-113.

Smith J.S., ans Boeke J.D. (1997). An unusual form of transcriptional silencing in yeast ribosomal DNA. *Genes Dev.* **11**, 241-254.

Smith M.C., and Thorpe H.M. (2002). Diversity in the serine recombinases. *Mol. Microbiol.* **44**, 299-307.

Snyder L. (2007). Molecular Genetics of Bacteria, 3rd Edition, ASM Press.

Sojka W.J. (1965). *Escherichia coli* in domestic animals and poultry. Part I: General characteristics and biochemical behaviour of *Escherichia coli*. Commonwealth Agricultural Bureaux: Farnham Royal, 1-63.

Solomon J.M., and Grossman A.D. (1996). Who's competent and when: regulation of natural genetic competence in bacteria. *Trends Genet.* **12**, 150-155.

Sordillo L.M., Shafer-Weaver K., and Derosa D. (1997). Symposium : bovine immunology. Immunobiology of the mammary gland. *J. Dairy Sci*. **80**, 1851-1865.

Sota M., Endo M., Nitta K., Kawasaki H., and Tsuda M. (2002). Characterization of a class II defective transposon carrying two haloacetate dehydrogenase genes from *Delftia acidovorans* plasmid pUO1. *Appl. Environ. Microbiol.* **68**, 2307-2315.

Sparling P.F. (1966). Genetic transformation of *Neisseria gonorrhoeae* to streptomycin resistance. *J. Bacteriol.* **92**, 1364-1371.

Springael D., Peys K., Ryngaert A., Van Roy S., Hooyberghs L., Ravatn R., Heyndrickx M., van der Meer J.R., Vandecasteele C., Mergeay M., and Diels L. (2002). Community shifts in a seeded 3-chlorobenzoate degrading membrane biofilm reactor: indications for involvement of in situ horizontal transfer of the clc-element from inoculum to contaminant bacteria. *Envirn. Microbiol.* **4**, 70-80.

Sreedhara Swamy K.H., and Goldberg A.L. (1982). Subcellular distribution of various proteases in *Escherichia coli. J. Bacteriol*. **149**, 1027-1033.

Srinivasan V., Gillespie B.E., Lewis M.J., Nguyen L.T., Headrick S.I., Schukken Y.H., and Olivier S.P. (2007). Phenotypic and genotypic antimicrobial resistance patterns of *Escherichia coli* isolated from dairy cows with mastitis. *Vet. Microbiol.* **124**, 319-328.

Stackebrandt E., and Goebel B.M. (1994). Taxonomic note: a place for DNA-DNA reassociation and rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* **44**, 846-849.

Stokes H.W., and Hall R.M. (1989). A novel family of potentially mobile DNA elements encoding site-specific gene integration functions: integrons. *Mol. Microbiol.* **3**, 1669-1683.

Stordeur P., Marlier D., Blanco J., Oswald E., Biet F., Dho-Moulin M., and Mainil J. (2002). Examination of *Escherichia coli* from poultry for selected adhesin genes important in disease caused by mammalian pathogenic *E. coli*. *Vet. Microbiol.* **84**, 213-241.

Sukupolvi S., and O'Connor C.D. (1990). TraT lipoprotein, a plasmid-specified mediator of interactions between gram-negative bacteria and their environment. *Microbiol. Rev.* **54**, 331-341.

Sullivan J.T., and Ronson C.W. (1998). Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proc. Natl. Acad. Sci. USA* **95**, 5145-5149.

Sullivan J.T., Trzebiatowski J.R., Cruickshank R.W., Gouzy J., Brown S.D., Elliot R.M., Fleetwood D.J., McCallum N.G., Rossbach U., Stuart G.S., Weaver J.E., Webby R.J., De Bruijn F.J., and Ronson C.W. (2002). Comparative sequence analysis of the symbiosis island of *Mesorhizobium loti* strain R7A. *J. Bacteriol*. **184**, 3086-3095.

Szalo I.M., Taminiau B., and Mainil J. (2006). Le lipopolysaccharide d'*Escherichia coli* : structure, biosynthèse et rôles. *Ann. Méd. Vét.* **150**, 108-124.

Szöllosi G.J., Derényi I., and Vellai T. (2006). The maintenance of sex in bacteria is ensured by its potential to reload genes. *Genetics* **174**, 2173-2180.

Tabuchi N, Mitani H, Seino S, and Fukunaga M. (2002). The 44-kb linear plasmid molecule in the relapsing fever agent *Borrelia duttonii* strain Ly serve as a preservation of vmp genes. *Microbiol. Immunol*. **46**, 159-165.

Tatum E.L., and Lederberg J. (1947). Gene recombination in the bacterium *Escherichia coli* . *J. Bacteriol.* **53**, 673-684.

Tauschek M., Strugnell R.A., and Robins-Browne R.M. (2002). Characterization and evidence of mobilization of the LEE pathogenicity island of rabbit-specific strains of enteropathogenic *Escherichia coli. Mol. Microbiol.* **44**, 1533-1550.

Taylor A.L. (1963). Bacteriophage-induced mutation in *Escherichia coli* . *Proc. Natl. Acad. Sci. USA* **50**, 1043-1051.

Top E.M., and Springael D. (2003). The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. *Curr. Opin. Biotechnol.* **14**, 262-269.

Tsuda M., and Iino T. (1987). Genetic analysis of a transposon carrying toluene degrading genes on a TOL plasmid pWW0. *Mol. Gen. Genet.* **210**, 270-276.

Tsuda M., and Iino T. (1988). Identification and characterization of Tn4653, a transposon covering the toluene transposon Tn*4651* on TOL plasmid pWW0. *Mol. Gen. Genet.* **213**, 72-77.

Tsuda M., Ninegishi K., and Iino T. (1989). Toluene transposons Tn*4651* and Tn*4653* are class II transposons. *J. Bacteriol.* **171**, 1386-1393.

Tsuda M., and Iino T. (1990). Naphthalene degrading genes on plasmid NAH7 are on a defective transposon. *Mol. Gen. Genet.* **223**, 33-39.

Tsuda M. (1996). Catabolic transposons in pseudomonads. In: Nakazawa T., Furukawa K., Haas D., Silver S. (eds) Molecular biology of pseudomonads. American Society for Microbiology, Washington, D.C., pp. 219-228.

Tsuda M., and Genka H. (2001). Identification and characterization of Tn*4656*, a novel class II transposon carrying a set of toluene-degrading genes from TOL plasmid pWW53. *J. Bacteriol.* **183**, 6215-6224.

Valente C., Cardaras P., Ciorba A., and Tesei B. (1988). Studies on virulence factors of *Escherichia coli* isolated from cows with acute mastitis. *Arch. Veter. Italiano* **39**, 254-260.

VanBogelen R.A., Kelley P.M., and Neidhardt F.C. (1987). Differential induction of heat shock, SOS, and oxidative stress regulons and accumulation of nucleotides in *Escherichia coli*. *J. Bacteriol.* **169**, 26-32.

van der Lelie D., Chavarri F., Venema G., and Gasson M.J. (1991). Identification of a new genetic determinant for cell aggregation associated with lactose plasmid transfer in *Lactococcus lactis. Appl. Environ. Microbiol.* **57**, 201-206.

van der Meer J.R., and Sentchilo V. (2003). Genomic islands and the evolution of catabolic pathways in bacteria. *Curr. Opin. Biotechnol.* **14**, 248-254.

van der Meer J.R., Zehnder A.J.B., and de Vos WM. (1991). Identification of a novel composite transposable element, Tn*5280*, carrying chlorobenzene dioxygenase genes of *Pseudomonas* sp. Strain P51. *J. Bacteriol.* **173**, 7077-7083.

Vangroenweghe F., Duchateau L., and Burvenich C. (2004). Moderate inflammatory reaction during experimental *Escherichia coli* mastitis in primiparous cows. *J. Dairy Sci*. **87**, 886-895.

van Ham R.C., Kamerbeek J., Palacios C., Rausell C., Abascal F., Bastolla U., Fernández J.M., Jiménez L., Postigo M., Silva F.J., Tamames J., Viguera E., Latorre A., Valencia A., Morán F., and Moya A. (2003). Reductive genome evolution in *Buchnera aphidicola*. *Proc. Natl. Acad. Sci. USA* **100**, 581-586.

Vestigian K., and Goldenfeld N. (2005). Global divergence of microbial sequences mediated by propagating fronts. *Proc. Natl. Acad. Sci. USA* **102**, 7332-7337.

Viel A., and Branton D. (1996). Spectrin: on the path from structure to function. *Curr. Opin. Cell Biol*. **8**, 49-55.

Villaseca J.M., Navarro-Garcia F., Mendoza-Hernandez G., Nataro J.P., Cravioto A., and Eslava C. (2000). Pet toxin from enteroaggregative *Escherichia coli* produces cellular damage associated with fodrin disruption. *Infect. Immun.* **68**, 5920-5927.

Vivian A., Murillo J., and Jackson R.W. (2001). The roles of plasmids in phytopathogenic bacteria: mobile arsenals? *Microbiology* **147**, 763-780.

Vulic M., Dionisio F., Taddel F., and Radman M. (1997). Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. *Proc. Natl. Acad. Sci. USA* **94**, 9763-9767.

Vulic M., Lenski R.E., and Radman M. (1999). Mutation, recombination, and incipient speciation of bacteria in the laboratory. *Proc. Natl. Acad. Sci. USA* **96**, 7348-7351.

Wain J., House D., Pickard D., Dougan G., and Frankel G. (2001). Acquisition of virulence-associated factors by the enteric pathogens *Escherichia coli* and *Salmonella enterica*. *Phil. Trans. R. Soc. Lond. B* **356**, 1027-1034.

Waldor M.K., and Mekalanos J.J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**, 1910-1914.

Waldor M.K., Tschape H., and Mekalanos J.J. (1996). A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in *Vibrio cholerae* O139. *J. Bacteriol.* **178**, 4157-4165.

Wang L., Qu W., and Reeves P.R. (2001). Sequence analysis of four *Shigella boydii* O-antigen loci: implication for *Escherichia coli* and *Shigella* relationships. *Infect. Immun*. 69, 6923-6930.

Wang Y., Wang H., Xiang Q., Sun S.X., and Yu S.Y. (2002). Detection of the high-pathogenicity island of *Yersinia enterolitica* in enterotoxigenic and enteropathogenic *E. coli* strains. *Di Yi Jun Yi Da Xue Bao* **22**, 580-583.

Waters V.L. (2001). Conjugation between bacterial and mammalian cells. *Nat. Genet.* **29**, 375-376.

Watts J.L. (1988). Etiological agents of bovine mastitis. *Vet. Microbiol.* **16**, 41-66.

Wayne L.G., Brenner D.J., Colwell R.R., Grimont P.A.D., Kandler O., Krichevsky M.I., Moore L.H., Moore W.E.C., Murray R.G.E., Stackebrandt E., Starr M.P., and Truper H.G. (1987). Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Bacteriol.* **37**, 463-464.

Welch R.A., Burland V., Plunkett G.D. III, Redford P., Roesch P., Rasko D.A., Buckles E.L., Liou S. R., Boutin A., Hackett J., Stroud D., Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T., Mobley H.L.T., Donnenberg M.S., and Blattner F.R. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli. Proc. Natl. Acad. Sci. USA* **99**, 17020-17024.

Wenz J.R., Barrington G.M., Garry F.B., McSweeney K.D., Dinsmore R.P., Goodell G., and Callan R.J. (2001). Bacteremia associated with naturally occurring acute coliform mastitis in dairy cows. *JAVMA* **219**, 976-981.

Wenz J.R., Barrington G.M., Garry F.B., Ellis R.P., and Magnuson R.J. (2006). *Escherichia coli* isolates' serotypes, genotypes, and virulence genes and clinical coliform mastitis severity. *J. Dairy Sci.* **89**, 3408-3412.

Whitman W.B., Coleman D.C., and Wiebe W.J. (1998). Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. USA* **95**, 6578-6583.

Wiking L., and Nielsen J.H. (2004). The influence of oxidation on proteolysis in raw milk. *J. Dairy Res.* **71**, 196-200.

Willetts N.S., and Skurray R. (1987). Structure and function of the F factor and mechanism of conjugaison. In F.C. Neidhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, and H.E. Umbarger (ed.), *Escherichia coli* and *Salmonelle typhimurium*: cellular and molecular biology. American Society For Microbiology, Washington, D.C., pp. 1110-1133.

Willetts N., and Wilkins B. (1984). Processing of plasmid DNA during bacterial conjugation. *Microbiol. Rev*. **48**, 24-41.

Wyndham R.C., Singh R.K., and Straus N.A. (1988). Catabolic instability, plasmid gene deletion and recombination in *Alcaligenes* sp. BR60. *Arch. Microbiol* **150**, 237-243.

Xia X.S., Aathithan S., Oswiecimska K., Smith A.R.W., and Bruce I.J. (2001). A novel plasmid pIJB1 possessing a putative 2,4-dichlorophenoxyacetate degradative transposon Tn*5530* in *Burkholderia cepacia* strain 2a. *Plasmid* **39**, 154-159.

Yamamoto S., Terai A., Yuri K., Kurazono H., Takeda Y., and Yoshida O. (1995). Detection of urovirulence factors in *Escherichia coli* by multiplex polymerase chain reaction. *FEMS Immunol. Med. Microbiol.* **12**, 85-90.

Yamamoto S. (2007). Molecular epidemiology of uropathogenic *Escherichia coli*. *J. Infect. Chemother* **13**, 68-73.

Yang H.H., Vinopal R.T., Grasso D., and Smets B.F. (2004). High diversity among environmental *Escherichia coli* isolates from a bovine feedlot. *Appl. Environ. Microbiol.* **70**, 1528-1536.

Young J.P., Crossman L.C., Johnston A.W., Thomson N.R., Ghazoui Z.F., Hull K.H., Wexler M., Curson A.R., Todd J.D., Poole P.S., Mauchline T.H., East A.K., Quail M.A., Churcher C., Arrowsmith C., Cherevach I., Chillingworth T., Clarke K., Cronin A., Davis P., Fraser A., Hance Z., Hauser H., Jagels K., Moule S., Mungall K., Norbertczak H., Rabbinowitsch E., Sanders M., Simmonds M., Whitehead S., and Parkhill J. (2006). The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol*. **7,** R34.

Zhao K., Chai X., and Marmorstein R. (2004). Structure and substrate binding properties of cobB, a Sir2 homolog protein deacetylase from *Escherichia coli*. *J. Mol. Biol.* **337**, 731-741.

Zinder N.D., and Lederberg J. (1952). Genetic exchange in Salmonella. *J. Bacteriol.* **64**, 679-699.

Résumé

L'objectif du travail présenté dans ce mémoire a été de caractériser une souche *E. coli* bovine appartenant au pathotype MPEC (MammoPathogen *Escherichia coli*). Ce dernier est probablement l'un des pathotypes de l'espèce *E. coli* le moins bien caractérisé. Ce manque de connaissances est expliqué par le fait que, jusqu'à très récemment, les souches bovines MPEC étaient considérées comme des souches pathogènes opportunistes. Il était admis, en effet, que ces souches atteignaient les mamelles bovines à la suite d'une contamination de l'environnement (litière, sol, eau…), qu'elles étaient ensuite capables d'y déclencher une inflammation appelée mammite, uniquement par le biais de leur lipopolysaccharide, puis qu'elles étaient rapidement éliminées sous l'action du système immunitaire bovin. Des études récentes ont montré cependant que, tout comme les souches des autres pathotypes de l'espèce *E. coli*, les souches bovines MPEC semblaient être dotées de propriétés de virulence particulières, telles que des propriétés de survie et de persistance dans la mamelle, et d'adhérence et d'invasion des cellules épithéliales mammaires. La souche bovine MPEC qui a été prise comme modèle pour ce travail est la souche *E. coli* P4. Une étude phylogénétique de cette souche a tout d'abord été réalisée, montrant qu'elle appartenait au groupe phylogénétique A de l'espèce *E. coli* et que son génome "core" se rapprochait du génome core de la souche commensale non pathogène *E. coli* K12 MG1655. Une recherche au sein de son génome de gènes codant différents facteurs de virulence connus chez les autres pathotypes de l'espèce *E. coli* a permis ensuite de révéler uniquement la présence du gène *traT*, codant un facteur de résistance au sérum. Pour tenter d'expliquer tout de même la virulence de cette souche, une analyse des parties variables de son génome a alors été entreprise. Il est désormais bien connu en effet qu'au sein d'une espèce bactérienne, les propriétés particulières de certaines souches s'expliquent par un bagage génétique variable différent. La recherche d'une catégorie d'éléments génétiques variables a ainsi été ciblée, celle des îlots génomiques situés en aval des loci d'ARNt. Un criblage de quinze loci d'ARNt connus pour accueillir fréquemment des îlots génomiques, effectué au sein du génome d'*E. coli* P4, a révélé, pour sept d'entre eux (*argW*, *asnT*, *aspV*, *leuX*, *pheV*, *serU* et *thrW*) la présence de telles structures. Le séquençage partiel des régions aval aux loci *asnT*, *leuX*, *pheV*, *serU* et *thrW* et le séquençage complet de la région aval au locus *argW* a montré la présence systématique de séquences nucléotidiques différentes de celles présentes chez *E. coli* K12 MG1655. Si l'analyse du contenu de ces îlots n'a pas encore permis d'expliquer directement le caractère pathogène d'*E. coli* P4, leur mise en évidence est une première de ce type au sein du pathotype MPEC et laisse envisager la découverte d'autres régions génomiques spécifiques à ce pathotype et pouvant, peut-être, expliquer son tropisme et sa nature. Une autre partie du travail réalisé a consisté à évaluer le rôle d'*E. coli* P4 dans la caséinolyse élevée du lait observée lors d'une mammite bovine. Dans ce cadre, une sécrétion apparemment constitutive de quatre protéases bactériennes extracellulaires présentant une activité caséinolytique par zymographie a été mise en évidence. Différents procédés expérimentaux réalisés par la suite ont, cependant, montré que l'intensité de l'activité protéolytique de ces enzymes sur les caséines n'était pas significative. Un rôle direct de la bactérie dans la caséinolyse du lait observée lors d'une mammite ne peut ainsi être soutenu pour l'instant. La découverte de ces protéases extracellulaires propres à *E. coli* P4 pourrait constituer tout de même un point intéressant dans la caractérisation de la virulence de cette souche. Au sein des autres pathotypes de l'espèce *E. coli*, différentes protéases extracellulaires ont en effet été identifiées et leur rôle crucial dans la virulence des souches a été démontré.

Annexe

AUTORISATION DE SOUTENANCE DE THESE
DU DOCTORAT DE L'INSTITUT NATIONAL
POLYTECHNIQUE DE LORRAINE

o0o

VU LES RAPPORTS ETABLIS PAR :
Madame Chantal LE BOUGUENEC, Directeur de Recherche, Institut Pasteur, Paris

Monsieur Christian BURVENICH, Professeur, Université de Ghent, Faculté de Médecine Vétérinaire, Belgique

Le Président de l'Institut National Polytechnique de Lorraine, autorise :

Madame DUFOUR Delphine

NANCY BRABOI
2, AVENUE DE L
FORET-DE-HAY
BOITE POSTALE
F - 5450
VANDŒUVRE CEDE

à soutenir devant un jury de l'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE, une thèse intitulée :

"**Recherche de déterminants génétiques permettant l'adaptation d'une souche *Escherichia coli* à la mamelle bovine**"

en vue de l'obtention du titre de :

DOCTEUR DE L'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE

Spécialité : « **Sciences agronomiques** »

Fait à Vandoeuvre, le 17 octobre 2008
Le Président de l'I.N.P.L.,
F. LAURENT

TEL. 33/03.83.59.59.5
FAX. 33/03.83.59.59.5