

**THESE DE DOCTORAT DE  
L'UNIVERSITE DES ANTILLES ET DE LA GUYANE**

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITE DES ANTILLES ET DE LA GUYANE**

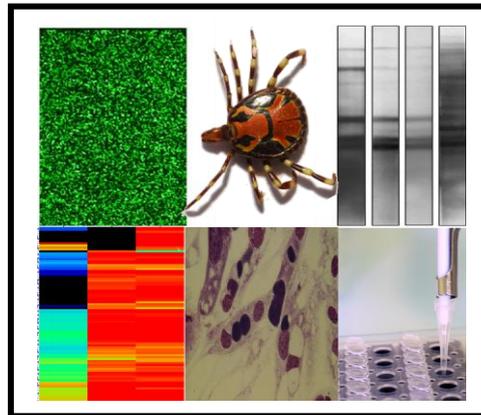
Spécialité  
**Microbiologie Moléculaire**

Ecole Doctorale Pluridisciplinaire de l'Université des Antilles et de la Guyane

Présentée et soutenue publiquement par

**Loïc EMBOULÉ**

**Analyse du transcriptome des souches virulentes et atténuées d'*Ehrlichia ruminantium* et application au développement de vaccin de deuxième génération**



**Devant le jury composé**

Présidente  
Rapporteur  
Rapporteur  
Examineur  
Examineur

**Pr France DAIGLE**, Professeur Université de Montréal  
**Pr Philippe BROUQUI**, Faculté de médecine de Marseille, la Timone  
**Dr Eric DUCHAUD**, CR INRA, Université Paris VI  
**Pr Charles DOZOIS**, Institut Armand Frappier  
**Pr Abel HIOL**, IUT Saint Claude, Université des Antilles et de la Guyane

Directeur de thèse  
Co-directeur  
Encadrant

**Dr Thierry LEFRANCOIS**, Directeur adjoint UMR 15 INRA-CIRAD  
**Pr Olivier GROS**, Université des Antilles et de la Guyane  
**Dr Nathalie VACHIERY**, Chercheur, UMR15

Analyse du transcriptome des souches virulentes et atténuées  
d'*Ehrlichia ruminantium* et application au développement de  
vaccin de deuxième génération

## Remerciements

Les travaux présentés dans ce mémoire ont été réalisés dans le cadre de recherches menées principalement en Guadeloupe, au laboratoire CIRAD UMR15/ UMR CIRAD-INRA 1309 « contrôle des maladies animales exotiques et émergentes ». Une partie des travaux a aussi été réalisée, au sein du Département de Microbiologie et d'Immunologie de l'Université de Montréal ainsi qu'à l'Institut de Pharmacologie Moléculaire et Cellulaire à Sophia Antipolis. Ces travaux ont bénéficié de l'appui financier du CIRAD, de la Région Guadeloupe, du Fonds Social Européen, du projet européen EPIGENEVAC INCO-CT-2005-003713 et du Fonds France-Canada pour la Recherche.

Je tiens tout d'abord à remercier Nathalie Vachery pour m'avoir permis d'accomplir ce travail sous sa direction, ainsi que pour son aide, ses conseils précieux et sa disponibilité. Ta confiance, ta rigueur scientifique et surtout ton amitié à mon égard ont été des sources de motivation. Tu m'as fait partager tes connaissances et ton amour profond de la recherche. Les moments passés avec toi à réfléchir sur les stratégies ou les résultats restent parmi les meilleurs moments de mon travail de thèse. « Je n'oublie pas les fous rires ». Ces quelques lignes ne sont pas suffisantes pour t'exprimer toute ma gratitude.

Merci à Thierry Lefrançois, responsable de notre équipe en Guadeloupe, pour m'avoir co-encadré et conseillé tout au long de ces années de thèse. J'ai été ravi de bénéficier de ton expérience et de ta large culture scientifique. Ton calme à toute épreuve, tes conseils bienveillants et tes qualités humaines m'ont été d'un grand soutien surtout dans les moments difficiles.

Je remercie le Professeur Philippe Brouqui et le Docteur Eric Duchaud qui m'ont fait l'honneur et le plaisir d'accepter d'être les rapporteurs de ce travail.

Merci aussi au Professeur Charles Dozois et au Professeur Abel Hiol pour avoir accepté de juger cette thèse. Je suis très honoré de votre présence dans mon jury.

Merci aussi au Professeur France Daigle, présidente du jury, pour m'avoir accueilli pendant 3 mois dans son laboratoire à l'Université de Montréal au Canada au cours de ma formation à la méthode SCOTS. Grâce à toi j'ai pu disposer de tous les moyens nécessaires à la réalisation de mes travaux de thèse. Ton aide et ta collaboration dans mes recherches ont été précieuses. A travers toi je tiens également à remercier toute ton équipe qui m'a fait découvrir votre merveilleuse culture ainsi qu'un pays exceptionnel.

Je remercie tout particulièrement Bernard Mari ainsi que Pascal Barbry de l'IPMC à Sophia Antipolis pour m'avoir accueilli dans leur laboratoire. Merci de m'avoir fait partager vos connaissances sur les puces à ADN et surtout sur l'analyse statistique des données. Je tiens à remercier sincèrement toute l'équipe pour l'aide précieuse et chaleureuse qu'ils m'ont apportée, avec une attention toute particulière à Virginie Magnone qui m'a montré avec beaucoup de patience la technique d'hybridation.

Merci au Professeur Olivier Gros qui a accepté d'être mon garant à l'Université des Antilles et de la Guyane en étant co-directeur de la thèse. Ta participation aux comités de thèse, mais

surtout les longues discussions que nous avons pu avoir m'ont toujours éclairé et aiguillé ces dernières années. J'ai réellement apprécié les enseignements que j'ai pu assurer sous ta responsabilité dans le cadre des vacances puis en tant qu'ATER.

Je remercie le Docteur Dominique Martinez pour m'avoir accueilli à l'UMR15 du CIRAD et d'avoir participé à mes comités de thèse. Merci pour vos conseils si précieux.

Merci également à Damien Meyer pour m'avoir apporté son aide et ses conseils scientifiques. L'intérêt que tu as porté à ce travail à travers nos discussions et tes fameux « brainstorming » m'a toujours stimulé.

Je remercie sincèrement les autres membres de mon comité de thèse que je n'ai pas encore cités. Il s'agit de Pierre-Yves Teycheney, Jean-Christophe Bambou et Didier Mbegue-Mbegue, qui ont su nous aider à définir les axes prioritaires à travers leurs commentaires et conseils avisés. Jean-Christophe tu es un véritable grand frère et un exemple merci pour tous tes conseils.

Un grand merci à toute l'équipe de l'UMR15 du CIRAD Guadeloupe en particulier Christian Sheikboudou et Valérie Pinarello, mes acolytes. Christian ou Monsieur culture, Monsieur protéine, Monsieur QSE, ton aide le samedi et le dimanche et tous les autres jours m'a été d'un grand secours. Tu as été là du début jusqu'à la fin dans la joie et la bonne ou mauvaise humeur. Même si on a beaucoup rigolé, j'ai vraiment appris énormément grâce à toi. « mési onpil ». Valérie, ma partenaire de choc, merci pour ton excellence et ta rigueur dans le travail. Ton amitié a été capitale dans la réalisation de cette thèse et notre complicité a été un véritable moteur. J'ai adoré défiler avec le « Bagad Karukéra ». Merci à Yolaine Vincent pour m'avoir aidé dans toute la partie administrative (mission, dossiers, congrès etc), ton efficacité me laisse admiratif. L'animation et l'ambiance du labo ont toujours été divertissantes avec de nombreux rebondissements ; il n'y a jamais eu de monotonie ! Merci à Rosalie Aprelon pour m'avoir si souvent arrêté le thermocycleur et pour avoir partagé ma passion pour la musique « gwo ka ». Merci à tous les agents du CIRAD Guadeloupe qui m'ont toujours encouragé en particulier mes amis de Neufchâteaux et Roujol. Merci à tous les VCAT et les stagiaires avec qui j'ai partagé de très bons moments en particulier Sylvie, Nico et Lucie bien sûr !!

Je remercie également Marie-Noëlle Sylvestre qui m'a beaucoup aidé et encouragé la dernière année. Tu as été très compréhensive et surtout indulgente avec moi. Assurer les enseignements sous ta responsabilité a été un vrai plaisir. Je n'oublie pas Clara Allayous qui m'a vraiment soutenu dans la dernière ligne droite avec les TP d'immunologie.

Merci aux amis thésards et néo-docteurs avec lesquels j'ai pas mal échangé aux cours de la thèse notamment Ludovic, Vanessa, Carole, ma cousine Danitza et le Docteur S. Faucher. Je n'oublie pas mes Québécoises préférées Marthe et Christine qui ont totalement adopté la Guadeloupe « Pouuuuuuuu ».

Une petite pensée pour les Professeurs Estelle Jumas-Bilak et Hélène Marchandin avec lesquelles j'ai découvert le monde de la recherche au sein de l'EA3755 à la faculté de Pharmacie de Montpellier. Je n'oublie pas notre merveilleux voyage à Madrid pour mon premier congrès FEMS. Il y a aussi Isabelle et surtout Agnès avec lesquelles j'ai beaucoup rigolé et travaillé bien sûre !

Merci à Murielle, Quévin, Luc, Sonny, Olivier, Nico « kaké » et Marion, amis de longue date. Le chemin a été long mais vous avez toujours été là pour moi, malgré la distance. Nous sommes comme des frères et sœurs ; notre amitié est sacrée et votre absence me pèse énormément.

Merci aux amis et membres de la « tribu gwo ka du vendredi » qui m'ont soutenu surtout pendant la phase de rédaction, en particulier Lina, Servane, Laure et Stéphanie. Grâce à vous le « Bik a tanbou » est devenu une institution en Guadeloupe. Les soirées de détente m'ont fait un bien fou.

Je remercie également tous ceux qui m'ont aidé au cours de cette thèse, ceux qui m'ont dépanné pour du matériel ou autre et que je n'ai pas cité. Je ne vous en remercierai jamais assez.

Enfin je tiens à remercier tous les êtres chers de ma grande famille. Je voudrais donc tout particulièrement remercier ma mère qui a toujours cru en moi. Merci « manman Lydia » pour tes encouragements et ton soutien constant durant mes études, je te dédie cette thèse. Je sais que rien n'a été facile, tu m'as très tôt fait prendre conscience de la valeur du travail. Merci à mon frère Matthew qui m'est très cher et à Jacky. Une tendre pensée pour mes taties qui m'ont toujours gâté sans oublier les tontons, cousins et cousines.

Ce travail n'aurait pas été possible non plus sans tout le soutien de Roxane : ton amour et ta présence à mes côtés ont été très précieux pour tout ce que j'ai entrepris ces dernières années. Je sais que cette année n'a pas été facile mais tu m'as toujours réconforté avec tes attentions tendres à mon égard. Merci de m'offrir le plus beau cadeau dont on puisse rêver.

## Résumé

La bactérie intracellulaire obligatoire *Ehrlichia ruminantium* (ER) est l'agent pathogène responsable de la cowdriose ou heartwater chez les ruminants domestiques et sauvages. Pour élucider la pathogenèse de la bactérie dans les cellules endothéliales, l'analyse du transcriptome a été entreprise en comparant le niveau d'expression des souches Gardel virulente et atténuée dont les génomes ont été séquencés. La sélection des transcrits bactériens a été réalisée grâce à l'adaptation de la méthode SCOTS à notre modèle. Les cDNA obtenus par SCOTS ont été hybridés sur des puces à ADN pangénomique d'ER spécifiques des souches Gardel et Welgevonden. Cette étude a révélé 239 gènes différentiellement exprimés entre souche virulente et atténuée dans les cellules infectées aux 3 stades évolutifs étudiés. 116 gènes ont été détectés au temps le plus tardif de l'infection (corps élémentaires libres). Parmi les gènes surexprimés par la souche virulente, certains sont connus chez d'autres bactéries intracellulaires (*R. conorii*) pour leurs implications dans l'échappement à la réponse de l'hôte : réparation de l'ADN (helicase, ligase), échappement au stress oxydatif et osmotique (thioredoxine), ainsi que dans l'inhibition de la fusion des endosomes (SNARE) et dans le ralentissement de l'apoptose (AnkA). Parmi les gènes surexprimés chez la souche atténuée, certains peuvent induire une réponse immunitaire protectrice. En combinant les données génomiques des groupes de gènes potentiellement co-régulés ont été identifiés notamment le « cluster » *recB*. La validation des résultats par qRT-PCR et l'analyse du protéome d'ER sont en cours, ce qui nous offre une perspective d'analyse fonctionnelle.

**Mots clés :** *E. ruminantium*, intracellulaire, diversité génétique, SCOTS, transcriptome, vaccin

## Summary

The obligatory intracellular bacterium *Ehrlichia ruminantium* (ER) is a pathogen responsible for heartwater which affects wild and domestic ruminants. To elucidate the pathogenesis of ER in endothelial cells, a high throughput analysis of its transcriptome was done. A comparison of gene expression between two sequenced strains was done: Gardel virulent and attenuated. Selective capture of ER transcribed sequences was done using the SCOTS method which has been adapted to our model. The cDNA were hybridized on pangenomic ER microarrays specific of both Gardel and Welgevonden strains. 239 genes were identified to be differentially expressed between virulent and attenuated Gardel strains in host cells within the three developmental stages studied. 116 genes were detected at the late stage corresponding to free elementary bodies. Among over-expressed genes identified for virulent strain, some genes have been described previously in other intracellular bacteria (*R. conorii*) to be involved in evasion of the host cell immune response. They were mainly involved in DNA repair (helicase, ligase), diminution of the oxidative (thioredoxin) and osmotic stress, the inhibition of endosome fusion (SNARE) and diminution of apoptosis (AnkA). Some interesting genes over-expressed for attenuated strain could induce a protective immune response. Genomic and transcriptomic comparison of data from both strains allowed identification of several clusters of genes potentially co-regulated such as the cluster *recB*. The qRT-PCR validation of differential gene expression identified by microarrays is currently done. In parallel, ER proteomic analysis paves the way for further functional analysis.

**Key words:** *E. ruminantium*, intracellular, genetic diversity, SCOTS, transcriptome, vaccine

## Sommaire

<b>Résumé .....</b>	<b>5</b>
<b>Liste des figures .....</b>	<b>10</b>
<b>Liste des tableaux .....</b>	<b>12</b>
<b>CHAPITRE I. Introduction.....</b>	<b>14</b>
<b>1. Les bactéries intracellulaires obligatoires .....</b>	<b>14</b>
<b>2. Notre modèle <i>Ehrlichia ruminantium</i> .....</b>	<b>16</b>
2.1. Caractéristiques génomiques d' <i>Ehrlichia ruminantium</i> .....	17
2.1.1. Souches d' <i>E. ruminantium</i> Gardel et Welgevonden virulentes .....	17
2.1.2. Souche d' <i>E. ruminantium</i> Gardel atténuée.....	19
2.2. Morphologie et cycle de développement d' <i>Ehrlichia ruminantium</i> .....	19
<b>3. Etude des mécanismes connus intervenants dans la pathogenèse des bactéries intracellulaires .....</b>	<b>21</b>
3.1. Internalisation, invasion et mobilité par polymérisation d'actine .....	21
3.2. Systèmes de sécrétions (SSTs).....	24
3.2.1. Les systèmes de sécrétion Sec-dépendant.....	25
3.2.2. Les systèmes de sécrétion Sec-indépendant.....	26
3.3. Signalisation cellulaire, survie dans la vacuole intracytoplasmique et lyse cellulaire .....	28
3.3.1. Les protéines à répétition ankyrines ou ankyrin-repeated proteins .....	28
3.3.2. Inhibition de la fusion phagosome-lysosome.....	29
3.3.3. L'échappement au stress oxydatif.....	30
3.3.4. Le stress osmotique.....	31
3.3.5. Système d'acquisition du fer, régulation et homéostasie.....	32
3.4. Autres effecteurs .....	34
<b>4. L'aire de la post-génomique: Profil d'expression des <i>Rickettsiales</i> .....</b>	<b>34</b>
4.1. Innovations scientifiques : caractérisation des agents pathogènes intracellulaires .....	34
4.2. La technique d'expression <i>in vivo</i> ou <i>In Vivo</i> Expression Technique (IVET).....	35
4.3. La mutagenèse à l'aide de transposon portant un marqueur-signature ou Signature tagged mutagenesis (STM).....	36
4.4. Hybridation soustractive suppressive (SSH):.....	37
4.5. Selective Capture of Transcribed Sequences (SCOTS) .....	37
4.6. Puces à ADN.....	39
4.6.1. Principe .....	39
4.6.2. Les limites pour l'analyse de l'expression de gènes de bactéries intracellulaires obligatoires.....	40
<b>5. Etude de l'expression des gènes de la famille multi-génique <i>map</i> d'<i>Ehrlichia ruminantium</i> chez l'hôte et le vecteur .....</b>	<b>42</b>
<b>6. Présentation de la Cowdriose .....</b>	<b>43</b>
6.1. La Cowdriose et agent pathogène .....	43
6.2. Répartition géographique .....	44
6.3. Les espèces affectées .....	45
6.4. Transmission .....	46
6.5. La maladie : symptômes et pathologie.....	47
6.5.1. Symptômes .....	47

6.5.2. Pathologie : les lésions.....	48
6.6. Diagnostics.....	48
6.7. L'immunité .....	49
6.8. Contrôle de la maladie .....	50
<b>7. Problématique et objectifs .....</b>	<b>52</b>
<b>CHAPITRE II. Protocole expérimental et stratégies d'étude: Etapes préalables nécessaire à l'analyse du transcriptome d'<i>Ehrlichia ruminantium</i> .....</b>	<b>55</b>
<b>1. Préparation du matériel biologique : Les souches <i>Ehrlichia ruminantium</i> Gardel virulente et atténuée en culture .....</b>	<b>56</b>
1.1. Contexte.....	56
1.2. Matériels et Méthodes.....	56
1.2.1. Production des échantillons biologiques et obtentions des ADN bactériens et ARN totaux .....	56
1.2.2. Synchronisation de l'infection et protocole de collecte des ARNs .....	58
1.2.3. Viabilité et quantification bactérienne par qPCR .....	59
1.2.4. Colorations et observations.....	60
1.3. Résultats.....	60
1.3.1. Viabilité et quantification bactérienne par qPCR .....	60
1.3.2. Colorations et observations par microscopie.....	62
1.4. Conclusion .....	64
<b>2. Vérification expérimentale de la virulence et de l'atténuation des souches Gardel <i>in vivo</i> .....</b>	<b>64</b>
2.1. Contexte .....	64
2.2. Matériels et méthodes .....	65
2.3. Résultats.....	66
2.4. Conclusion .....	66
<b>3. Evaluation de la stabilité des gènes constitutifs d'<i>Ehrlichia ruminantium</i> : Recherche de gènes normalisateurs pour la validation par PCR en temps réel des gènes d'intérêt détectés en microarrays.....</b>	<b>68</b>
3.1. Contexte.....	68
3.2. Matériels et méthodes .....	68
3.2.1. Les gènes sélectionnés .....	68
3.2.2. Dessin des amorces.....	69
3.2.3. Extraction d'ARN puis conversion en ADN-complémentaire (cDNA) .....	70
3.2.4. Test des amorces et efficacité de la qPCR.....	70
3.2.5. Quantification du nombre de copies d'ADN d'ER par qPCR Taqman <i>map1</i> .....	71
3.2.6. Mesure du niveau d'expression des gènes candidats .....	71
3.3. Résultats.....	72
3.3.1. Conditions optimales de qPCR et efficacité .....	72
3.3.2. Quantification du nombre de copies d'ADN d'ER par qPCR Taqman <i>map1</i> .....	73
3.3.3. Evaluation de la stabilité des gènes constitutifs par qRT-PCR .....	74
3.4. Conclusion .....	76
<b>4. Adaptation de la méthode SCOTS à la bactérie <i>Ehrlichia ruminantium</i> et validation sur puces à ADN pour l'analyse du transcriptome .....</b>	<b>76</b>
4.1. Contexte.....	76
4.2. Validation de la méthode SCOTS .....	77

4.3. Résultats complémentaires : Hybridation de l'ADN génomique des souches Gardel et Sénégal (virulente et atténuée) sur les puces à ADN.....	92
<b>5. Conclusion.....</b>	<b>95</b>
<b>CHAPITRE III. Analyse du transcriptome de la bactérie <i>Ehrlichia ruminantium</i>.....</b>	<b>98</b>
<b>1. Contexte.....</b>	<b>99</b>
<b>2. Matériels et méthodes .....</b>	<b>99</b>
2.1. Production des échantillons biologiques <i>in vitro</i> : souches Gardel virulente et Gardel atténuée	99
2.2. Capture sélective des transcrits d' <i>Ehrlichia ruminantium</i> par SCOTS.....	100
2.3. Puces à ADN pangénomique d' <i>Ehrlichia ruminantium</i> : description et analyse statistique.....	100
2.4. Acquisition et analyse de l'image .....	100
2.5. Analyse statistique des résultats.....	101
2.6. Annotation des CDS différemment exprimés .....	101
2.7. Validation des gènes d'intérêt par qRT-PCR .....	101
<b>3. Résultats .....</b>	<b>102</b>
3.1. Comparaison des cinétiques de croissance bactérienne pour l'analyse du transcriptome .....	102
3.2. Analyse du transcriptome d' <i>E. ruminantium</i> par microarrays .....	103
3.2.1. Analyse globale et répartition des catégories de gènes.....	103
3.2.2. Identification des gènes différemment exprimés par <i>Ehrlichia ruminantium</i> au cours des différentes phases du cycle de développement .....	108
3.2.3. La famille multigénique <i>map</i> .....	125
3.2.4. Le cluster <i>recB</i> .....	125
3.2.5. Transcriptome d' <i>E. ruminantium</i> à partir des transcrits SCOTS 0x à T3.....	126
3.3. Validation des CDS d'intérêts par qRT-PCR.....	127
<b>CHAPITRE IV. Discussion générale .....</b>	<b>129</b>
<b>1. Expression des gènes impliqués dans les fonctions métaboliques .....</b>	<b>131</b>
1.1. Biogénèse des ribosomes et la traduction .....	131
1.2. La production d'énergie .....	133
1.3. Réparation et réplication de l'ADN .....	133
<b>2. Expression des gènes impliqués dans la survie intracellulaire : échappement aux systèmes de défense de l'hôte. ....</b>	<b>134</b>
2.1. Les stress oxydatifs, thermiques et osmotiques .....	134
2.2. Acquisition de fer, magnésium et autres éléments.....	135
2.3. Transporteurs ABC: SST1 .....	135
2.4. Inhibition de la fusion phagolysosomiale .....	136
2.5. Résistance aux antibiotiques .....	136
<b>3. La famille multigénique <i>map</i> .....</b>	<b>137</b>
<b>4. Le cluster <i>recB</i> .....</b>	<b>137</b>
<b>5. Virulence bactérienne .....</b>	<b>139</b>
<b>6. Les gènes de fonction inconnue .....</b>	<b>140</b>
<b>CHAPITRE V. Conclusions et perspectives .....</b>	<b>142</b>
<b>ANNEXES.....</b>	<b>147</b>

<b>REFERENCES BIBLIOGRAPHIQUES.....</b>	<b>152</b>
<b>Posters présentés dans les congrès internationaux au cours de la thèse.....</b>	<b>165</b>

## Liste des figures

<b>Chapitre I</b> .....	
Figure 1 : Structuration de l'ordre des Rickettsiales [reproduit de Marcelino, (2007) adapté de Dumler <i>et al.</i> , (2001); Kocan <i>et al.</i> , (2003)].....	16
Figure 2 : Observation par microscopie optique de <i>morula</i> d' <i>E. ruminantium</i> (indiquées par des flèches) dans des cellules endothéliales aortiques bovines (BAE) infectées <i>in vitro</i> . (N. Vachiéry).....	20
Figure 3 : Cycle de développement d' <i>E. ruminantium</i> dans l'hôte et le vecteur (D. Martinez).....	21
Figure 4 : Représentation détaillée des différents systèmes de sécrétions des bactéries Gram négatif (KEGG database).....	28
Figure 5: Selective Capture of Transcribed Sequences (SCOTS) (d'après Daigle <i>et al</i> 2001).....	39
Figure 6 : Schéma représentatif des méthodes d'hybridations avec les puces à ADN.....	41
Figure 7 A: Probabilité de présence de la maladie en fonction de la présence de la tique. (Allsopp <i>et al</i> , 2010).....	44
Figure 7 B: Indice de présence de la tique <i>A. variegatum</i> sur les continents américain et africain (Estrada-Pena <i>et al</i> 2006).....	45
Figure 8 : Photo de tiques appartenant à l'espèce <i>Amblyomma variegatum</i> (N. Vachiéry).....	47
<b>Chapitre II</b> .....	
Figure 9 : Répartition des réplicats pour l'analyse du transcriptome d' <i>E. ruminantium</i> utilisation des méthodes microarrays et qRTPCR.....	57
Figure 10 : Protocole expérimental mis en place pour l'extraction d'ARN d' <i>E. ruminantium</i> .....	59
Figure 11 : Estimation de la viabilité d' <i>E. ruminantium</i> ERGp39 (A) et ERGp243 (B) par la cytométrie en flux.....	61
Figure 12 : Quantification absolue par qPCR à l'aide d'une sonde taqman spécifique du gène <i>map1</i> .....	62
Figure 13 : Observation de l'évolution du cycle infectieux des souches d' <i>E. ruminantium</i> par coloration (RAL 555, L. Emboulé).....	63
Figure 14 : Observation de l'évolution des symptômes au cours de l'infection.....	67
Figure 15 : Quantification absolue de ERGp36 (A) et ERGp38 (B) par qPCR à l'aide d'une sonde taqman spécifique du gène <i>map1</i> .....	73
Figure 16 : Coefficient de corrélation entre les deux génomes Gardel virulent et atténuée.....	92
Figure 17 : Coefficient de corrélation entre l'hybridation génomes de Sénégal virulent et atténuée sur les puces pangénomiques d' <i>E. ruminantium</i> Gardel virulent.....	93
Figure 18 : Coefficient de corrélation entre les deux génomes Gardel virulent et Sénégal virulent.....	93
Figure 19 : Coefficient de corrélation entre les deux génomes Gardel virulent et Sénégal atténuée.....	94

Figure 20 : Heat map : Les gènes les plus marquants permettant la différenciation des souches Sénégal (SP5), Gardel (GP40) et Welgevonden (WP11).....	95
<b>Chapitre III</b> .....	
Figure 21 A : Pourcentages des catégories des 239 gènes différentiellement exprimés de T1 à T3 entre la souche Gardel virulente (ERGp39) et la souche Gardel atténuée (ERGp243).....	106
Figure 21 B : Gènes différentiellement exprimés entre Gardel atténuée et virulente : répartition selon leur fonction connue ou supposée.....	107
Figure 22 : Répartition et fonction des gènes différentiellement exprimés entre Gardel atténuée et virulente à 24 hpi.....	109
Figure 23 : Catégorie de gènes différentiellement exprimés entre les souches Gardel virulentes et atténuées d' <i>Ehrlichia ruminantium</i> au moment de la phase de croissance exponentielle (T2).....	116
Figure 24 : Catégorie de gènes différentiellement exprimés entre les souches Gardel virulentes et atténuées d' <i>Ehrlichia ruminantium</i> moment de la lyse des cellules (T3).....	123
Figure 25 : Nombre de gènes de la famille <i>map</i> surexprimés à chaque phase (T1/T2/T3).....	125
Figure 26 : Représentation schématique du cluster <i>recB</i> .....	126
<b>Chapitre IV</b> .....	
Figure 27 : Schéma global de la biologie d' <i>Ehrlichia ruminantium</i> .....	141

## Liste tableaux

<b>Chapitre I</b> .....	
Tableau 1 : Caractéristiques générales des génomes d' <i>E. ruminantium</i> Gardel (Erga), Welgevonden Guadeloupe (Erwe) et Welgevonden d'Afrique du Sud (Erwo).....	18
<b>Chapitre II</b> .....	
Tableau 2 : Evaluation pathologique avec indices en fonction de l'intensité des signes cliniques observés chez une chèvre naïve ou vaccinée infectée avec une dose d' <i>E. ruminantium</i> vivante.....	65
Tableau 3 : Suivi de l'hyperthermie et de la période d'incubation.....	67
Tableau 4 : Gènes constitutifs utilisés comme normalisateurs dans d'autres modèles bactériens.....	69
Tableau 5 : Conditions optimales de qRT-PCR pour chaque gène candidat.....	72
Tableau 6 : résultats des efficacités des qRT-PCR.....	72
Tableau 7 : Quantification du nombre de copies d'ADN chez ERGp38 et ERGp36 au cours du temps.....	73
Tableau 8 : Nombre de copies de cADN quantifié pour la souche ERGp38.....	74
Tableau 9A : Stabilité de l'expression des gènes (R) chez Gp38 au cours du temps.....	75
Tableau 9B : Stabilité de l'expression des gènes (R) chez Gp36 au cours du temps.....	75
<b>Chapitre III</b> .....	
Tableau 10 : Cinétique et comparaison des différents stades de développement.....	103
Tableau 11 : Répartition des gènes différentiellement exprimés en fonction des stades de développement entre ERGp39 et ERGp243.....	104
Tableau 12 : Gènes surexprimés par la souche virulente ERGp39 à 24 hpi.....	110
Tableau 13 : Gènes sous-exprimés par la souche virulente ERGp39 à 24 hpi.....	111
Tableau 14 : Gènes surexprimés par la souche virulente ERGp39 à T2.....	115
Tableau 15 : Gènes sous-exprimés par la souche virulente ERGp243 à T2.....	118
Tableau 16 : Liste de gènes relevant surexprimés par la souche virulente ERGp39 à T3.....	120
Tableau 17 : Liste de gènes relevant sous-exprimés par la souche virulente ERGp39 à T3.....	121
Tableau 18 : Listes de gènes relevant communs aux différents temps (T1/T2 et T2/T3). A : Comparaison T1 et T2. B : Comparaison T2 et T3.....	124
Tableau 19 : Stabilité de l'expression des gènes (R) chez ERGp243 au cours du temps.....	128
Tableau 20 : Liste des gènes choisis pour la validation de gènes de ménages sur la base des résultats microarrays.....	128

# **CHAPITRE I.**

## **Introduction**

## **1. Les bactéries intracellulaires obligatoires**

Certains microbes pathogènes, virus et bactéries, sont transmis à l'homme et aux animaux par l'intermédiaire de vecteurs hématophages (insectes et arthropodes), qui représentent des réservoirs principaux de maladies sévères telles que la fièvre jaune, la dengue, la fièvre de la vallée du Rift et le typhus épidémique. Les pathogènes interagissent avec les hôtes et les vecteurs de multiples façons grâce à des processus biologiques simples ou complexes. Beaucoup d'entre eux ont un développement extracellulaire, alors que d'autres vont envahir les organes et se multiplier à l'intérieur de cellules spécifiques de l'hôte et du vecteur (mammalien, arthropode), par des mécanismes d'attachements, de pénétrations et d'invasions. Les alpha-protéobactéries (AP) récemment remaniées par Cavalier-Smith (2002), ont des modes de vie variés et originaux. Elles sont capables, en particulier, d'établir des relations étroites avec la cellule eucaryote (Teyssier *et al.*, 2004). Ainsi les AP regroupent des symbiotes des végétaux (*Rhizobium*, *Mesorhizobium*, ...) et des symbiotes d'insectes, de cloportes ou de nématodes (*Wolbachia*), des pathogènes de la cellule végétale (*Agrobacterium*), mais surtout des pathogènes des animaux, intracellulaires facultatifs (*Bartonella*, *Brucella*), ou intracellulaires obligatoires (*Anaplasma*, *Ehrlichia*, *Rickettsia*, ...). L'ordre des *Rickettsiales*, dont les membres présentent une grande divergence génétique, regroupe cette catégorie de bactéries « intracellulaires obligatoires ». Les *Rickettsiales* sont responsables de l'émergence et de la réémergence d'importantes maladies infectieuses dans le monde : notamment les ehrlichioses et les anaplasmoses (La Scola and Raoult, 1997; Richards, 2004; Dumler, 2005; Parola *et al.*, 2005). La plupart a évolué vers un mode de vie spécialisé commun, pathogène des animaux et/ou des hommes, à l'exception de certains membres du genre *Wolbachia* qui ont un comportement plutôt symbiotique. Ces modes de vie sont systématiquement associés à un chromosome unique et de petite taille (inférieur ou égale à 1,5 Mb) résultant des différents phénomènes de réduction génomique qui ont eu lieu au cours de leur évolution principalement liés à l'adaptation au mode de vie intracellulaire stricte. Cependant, la découverte récente de plasmides chez plusieurs espèces de bactéries appartenant au genre *Rickettsia*, *R. amblyommii*, *R. helvetica*, *R. massiliae* et *R. peacockii* isolées dans les tiques du genre *Ixodes* a supplanté le dogme historique qui voulait que ces bactéries ne possèdent pas de plasmides (Ogata *et al.*, 2005; Baldrige *et al.*, 2008).

Les outils moléculaires notamment l'étude de l'ARN ribosomal 16S, ont permis de réviser l'ordre des *Rickettsiales* qui est maintenant réduit à 2 familles (Dumler *et al.*, 2001) (Figure 1). Les *Rickettsiaceae*, dont tous les représentants se multiplient à l'état libre dans le cytoplasme et le noyau de la cellule hôte (Parola *et al.*, 2005), et la famille des *Anaplasmataceae* dont les représentants se multiplient au sein d'une vacuole intracytoplasmique dans la cellule hôte eucaryote. Bien qu'il n'existe pas de relation phylogénétique entre l'ordre des *Rickettsiales* et des *Chlamydiales*, leurs modes de vie particuliers ont été fréquemment comparés de par leur cycle cellulaire complexe.

Parmi les maladies transmises par les *Rickettsiales* on retrouve :

(i) Les rickettsioses causées par les bactéries du genre *Rickettsia*, regroupant le groupe des typhus (*Typhus Group* : TG ou groupe typhus, avec des maladies telles que le typhus épidémique dû à *R. prowazekii* et typhus murin dû à *R. typhi*), transmis par les poux et les puces et le groupe des fièvres pourprées (*Spotted Fever Group* : SFG, fièvre pourprée des montagnes rocheuses dû à *R. rickettsii*), transmis par les tiques à l'exception de *Rickettsia akari* (ii) le typhus des broussailles dû à *Orientia tsutsugamushi* appartenant au genre *Orientia*. (iii) Les ehrlichioses et anaplasmoses causées par les bactéries appartenant à la famille des *Anaplasmataceae* et appartenant aux genres *Anaplasma* et *Ehrlichia*.

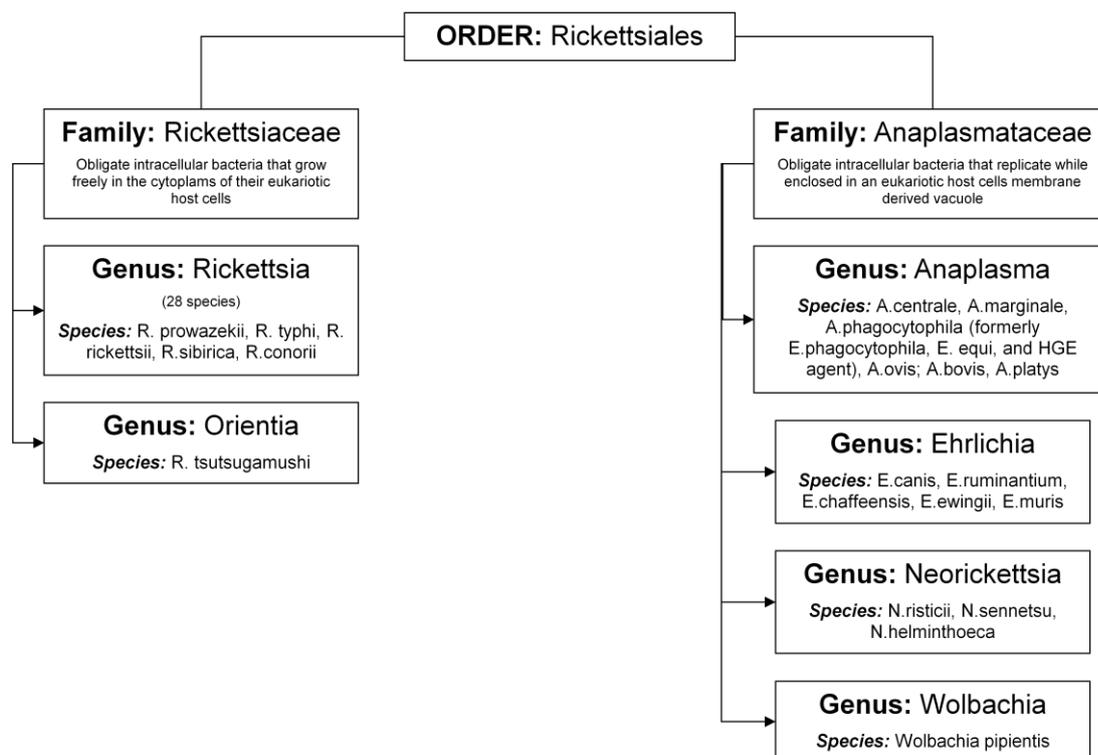


Figure 1 : Structuration de l'ordre des *Rickettsiales* [reproduit de Marcelino, (2007) adapté de (Dumler *et al.*, 2001).

## 2. Notre modèle *Ehrlichia ruminantium*

Notre modèle d'étude, la bactérie *Ehrlichia ruminantium*, a été récemment repositionnée, sur les bases de l'analyse phylogénétique des séquences codant pour l'ARNr 16S et l'opéron *groESL*, au sein du genre *Ehrlichia* contenant les espèces *Ehrlichia canis*, *chaffeensis*, *ewingii* et *muris* (Figure 1) (Dumler *et al.*, 2001). Les bactéries appartenant au genre *Ehrlichia* sont des petites bactéries (0,2-0,25µm) pathogènes intracellulaires strictes qui parasitent les cellules sanguines, épithéliales et endothéliales de nombreux mammifères et de l'homme. Elles sont Gram négatif mais, tout comme les *Rickettsiaceae*, elles diffèrent de la majorité des bactéries Gram négatif par l'absence de production d'endotoxines et par leur transmission essentiellement vectorielle. Largement répandues dans le monde, les *Ehrlichia* sp. sont les agents étiologiques de maladies infectieuses de l'homme, des carnivores, des rongeurs et des ruminants.

## 2.1. Caractéristiques génomiques d'*Ehrlichia ruminantium*

### 2.1.1. Souches d'*E. ruminantium* Gardel et Welgevonden virulentes

Les génomes de trois souches de référence ont été entièrement séquencés ((Collins *et al.*, 2005); (Frutos *et al.*, 2006a)). Il s'agit de la souche d'origine, Welgevonden (Erwo) isolée en Afrique du Sud (Du Plessis, 1985), d'un isolat de cette souche maintenu dans un environnement cellulaire différent en Guadeloupe, la souche Welgevonden (Erwe) et de la souche Gardel (Erga) isolée en Guadeloupe (Uilenberg *et al.*, 1985; Frutos *et al.*, 2006a). L'analyse génomique de ces trois souches a montré que le génome d'*E. ruminantium* avait une taille d'environ 1,5 mégabases (1499891 pb, 1513727 pb et 1516355 pb respectivement pour Erga, Erwe et Erwo) avec de 920 à 958 CDS suivant les souches (Tableau 1). L'analyse en génomique comparée de ces bactéries a permis d'identifier 3 régions RAD (regions of accumulated differences) où des troncations de gènes ont été mises en évidence de même qu'un petit nombre de gènes spécifiques de souches. Globalement, 57 CDS (coding sequences) ou gènes uniques (sans orthologue dans l'une ou l'autre souche), 18 gènes tronqués (permettant de différencier les souches ERga du groupe ERwe/ ERwo), ainsi que des gènes fragmentés et mutés, ont été identifiés. Parmi eux, 7 seulement correspondent à des réarrangements majeurs tels que des délétions partielles ou complètes ou des mutations extensives. Les autres correspondent à une dégénérescence partielle des séquences. Parmi ces 18 CDS tronqués, huit sont spécifiques de Erga et dix de Erwe/Erwo et seulement 2 correspondent à des gènes connus, *map1-2* (tronqué dans Erga) et *ftsA* (tronqué dans Erwe/Erwo). Parmi les gènes tronqués, la majorité est affectée par un seul codon stop ou quelques mutations. L'analyse comparative des mutations basée sur les taux de substitutions synonymes *versus* non synonymes entre les souches a permis d'analyser les pressions de sélection exercées sur les gènes. Sur l'ensemble des CDS identifiés, 181 montrent un nombre significatif de substitutions synonymes qui indique l'existence d'une forte pression de sélection dirigée vers le maintien de la séquence protéique (Frutos *et al.*, 2006a). Seulement 3 paires de CDS ont un nombre significatif de substitutions non synonymes indicateur de l'existence de pseudogènes ou de changements fonctionnels.

De plus, 30 % des CDS du génome d'*E. ruminantium* demeurent de fonction inconnue (Frutos *et al.*, 2006a). Le contenu en G+C est faible avec une valeur moyenne de 27,5%. Le génome d'*E. ruminantium* présente aussi un biais G-C très prononcé d'un brin à l'autre avec l'utilisation d'un jeu de codons différent sur chacun des deux brins, une caractéristique connue uniquement chez *Borrelia burgdorferi*, l'agent de la maladie de Lyme et *Treponema pallidum*, l'agent de la syphilis (Frutos *et al.*, 2006a). Ce qui frappe en premier lieu avec le génome d'*E. ruminantium* est le faible pourcentage de séquences codantes qui est de 62% à 63% associé à la présence de longs espaces inter-géniques. Bien que ce fait ne soit pas unique, cette valeur reste néanmoins l'une des plus faibles parmi l'ensemble des génomes bactériens. A l'intérieur de ces espaces inter-géniques du génome d'*E. ruminantium*, on retrouve un grand nombre de séquences répétées en tandem (tandem repeated sequences) (Collins *et al.*, 2005; Frutos *et al.*, 2006b; Frutos *et al.*, 2007). A ceci s'ajoute le fait que le nombre de copies est variable de façon active (Collins *et al.*, 2005), ce qui révèle la présence d'un mécanisme actif de plasticité génomique du point de vue de la taille (Frutos *et al.*, 2006b; Frutos *et al.*, 2007).

Tableau 1 : Caractéristiques générales des génomes d'*E. ruminantium* Gardel (Erga), Welgevonden Guadeloupe (Erwe) et Welgevonden d'Afrique du Sud (Erwo)

Caractéristiques	Génome		
	ERga <sup>a</sup>	ERwe <sup>a</sup>	ERwo <sup>b</sup>
Taille du génome (pb)	1499920	1512977	1516355
% G+C	27.51 %	27.48 %	27.5 %
Nombre de CDS prédits	950	958	920
% de génome codant	64.4 %	63.8 %	63.1 %
Taille moyenne des CDS	1007	998	1032
Nombre d'ARNt	36	36	36
Nombre d'ARNr	3	3	3

<sup>a</sup> d'après Frutos *et al.*, 2006

<sup>b</sup> d'après Collins *et al.*, 2005

### 2.1.2. Souche d'*E. ruminantium* Gardel atténuée

La souche Gardel atténuée a été entièrement séquencée dans le cadre d'un projet annexe. Il existe une forte colinéarité entre la souche Gardel virulente (données non publiées) et la souche Gardel atténuée, passage 230 avec une taille identique des deux génomes (données non publiées). Il n'y a pas de macro-réarrangements chromosomiques entre les deux souches. Par contre, il existe des différences ponctuelles (2797) ainsi qu'un nombre important de délétions/insertions (insertions 1780, délétions 500, substitutions 517) qui affectent 44% du génome. Pour l'heure, il n'y a pas de mise en évidence de mécanismes d'atténuation par l'analyse génomique *in-silico*. Cependant, seules des vérifications *in situ* par PCR et des analyses fonctionnelles permettront d'évaluer l'importance de ces mutations dans les mécanismes d'atténuation et l'impact biologique associé.

## 2.2. Morphologie et cycle de développement d'*Ehrlichia ruminantium*

*E. ruminantium*, est une bactérie Gram négatif, aérobie, de 0,2 à 2,7  $\mu\text{m}$ , souvent polymorphe, se présentant sous une forme coccoïde ou ellipsoïdale. Elle est présente dans les vacuoles intracytoplasmiques et est capable d'infecter les cellules endothéliales, les cellules de la lignée myéloblastique et les cellules de la lignée monoblastique des ruminants sauvages et domestiques (bovin, caprin et ovin). Elle infecte aussi des macrophages de souris en conditions expérimentales. Les études de microscopie électronique révèlent l'existence de trois stades différents de développement:

- une forme pléomorphe plus grande (1,2 $\mu\text{m}$ -2,8 $\mu\text{m}$ ) présentant des filaments nucléiques dispersés, les corps réticulés (van Vliet *et al.*, 1992). Les corps réticulés, visibles 2 à 4 jours après infection sont considérés comme des formes végétatives non infectieuses
- une forme intermédiaire (0,9 $\mu\text{m}$ -2,2 $\mu\text{m}$ ) issue de la modification des corps réticulés et ayant un ADN plus condensé, les corps intermédiaires.
- une forme plus petite (0,6-1,2 $\mu\text{m}$ ) ayant les filaments nucléiques condensés. Ils sont appelés corps élémentaires ou corps initiaux denses (ou dense-core forms). Ce sont des éléments ovoïdes, limités par une double membrane. Le corps élémentaire constitue la forme libre et infectieuse de la bactérie. Il adhère et pénètre dans la cellule endothéliale par un processus qui

ressemble à la phagocytose où il se développe et se multiplie au sein d'une vacuole cytoplasmique bordée d'une double membrane dérivant de la cellule hôte. Dans la vacuole, les organismes se divisent par fission binaire jusqu'à former une large colonie dénommée *morula* (figure 2). Après environ 4 à 6 jours de développement *in vitro*, la lyse de la cellule hôte conduit à la libération de centaines de corps élémentaires qui peuvent initier de nouveaux cycles infectieux (Jongejan *et al.*, 1991).

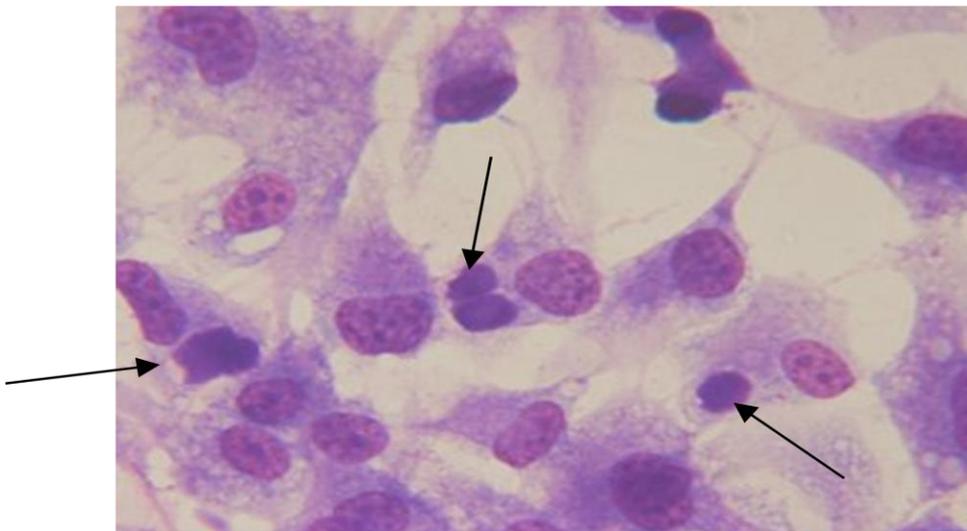


Figure 2 : Observation par microscopie optique de *morula* d'*E. ruminantium* (indiquées par des flèches) dans des cellules endothéliales aortiques bovines (BAE) infectées *in vitro*. (N. Vachiéry)

L'infection du vecteur se fait lors du repas sanguin. Initialement, dans les tiques, *E. ruminantium* se développe dans les cellules épithéliales de l'intestin, puis gagne l'hémolymphe d'où elle diffuse vers les glandes salivaires (Figure 3). La salive de la tique représente donc le principal matériel biologique infestant pour les hôtes réceptifs. Beaucoup de zones d'ombre persistent en ce qui concerne les mécanismes environnementaux, cellulaires et moléculaires qui orientent l'acquisition de la bactérie par le vecteur hôte. Les mécanismes qui permettent le passage de la bactérie de l'intestin vers les glandes salivaires et donc sa transmission au cours de l'infection des animaux sensibles demeurent encore mal connus à ce jour.

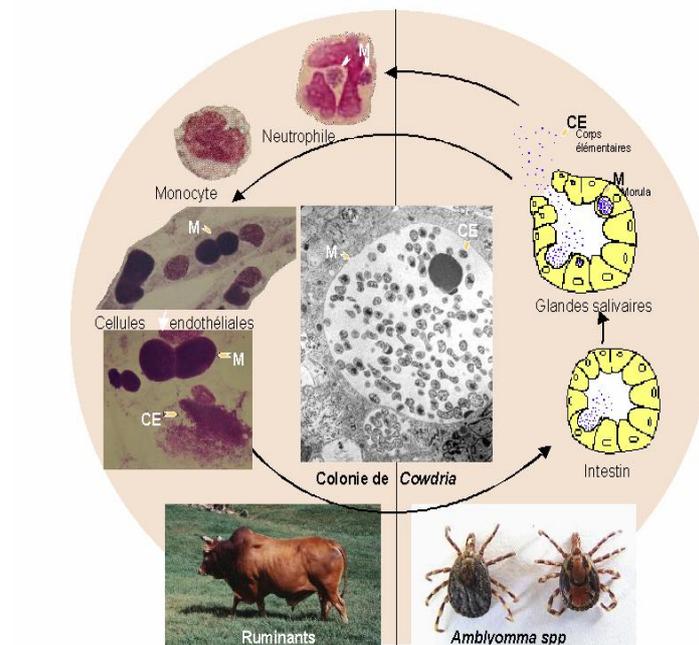


Figure 3 : Cycle de développement d'*E. ruminantium* dans l'hôte et le vecteur (D. Martinez)

### 3. Etude des mécanismes connus intervenants dans la pathogenèse des bactéries intracellulaires

De nombreuses études ont permis de mieux connaître les mécanismes de pathogenèse de bactéries libres ou intracellulaires non obligatoires ; par contre les déterminants génétiques responsables de la pathogenèse des *Rickettsiales* sont encore mal connus. Grâce à la disposition de nombreux génomes de *Rickettsiales*, la génomique comparative a permis d'identifier des gènes potentiellement impliqués dans leur pathogenèse. Ces gènes codent entre autres pour des protéines qui permettent l'invasion, la survie de la bactérie dans la cellule hôte et l'échappement aux systèmes de défense endogènes.

#### 3.1. Internalisation, invasion et mobilité par polymérisation d'actine

Beaucoup de bactéries intracytoplasmiques peuvent infecter une cellule ou se déplacer de cellule en cellule, facilitant leur survie, en polymérisant l'actine de la cellule hôte. L'actine est une protéine extrêmement conservée et abondante dans le cytosquelette des cellules eucaryotes et est essentielle dans l'architecture et la motilité cellulaire. L'actine forme des filaments en polymérisant et participe à de nombreuses fonctions cellulaires essentielles

comme la division cellulaire, la formation d'extensions membranaires comme les filopodes et les lamellipodes, les jonctions entre cellules et la phagocytose. Plusieurs familles de protéines régulent la dynamique du cytosquelette d'actine. Certaines favorisent la nucléation, d'autres la polymérisation, la stabilité des filaments ou leur dépolymérisation. Ainsi certaines protéines bactériennes telles que ActA chez *Listeria monocytogenes* et IcsA (VirG) chez *Shigella flexneri* (Domann *et al.*, 1992) peuvent induire la polymérisation de l'actine en recrutant des protéines du cytosquelette, notamment les protéines du complexe Arp2/3 et N-WASp. IcsA est une protéine de membrane externe de 1 102 acides aminés. La caractéristique majeure de cette protéine est la présence de répétitions riches en glycine dans la première partie de la protéine. Cette protéine est distribuée de manière polaire à la surface de la bactérie. Lorsqu'on exprime IcsA chez *E. coli* K12, les bactéries forment des comètes similaires à celles de *Shigella* (Goldberg and Theriot, 1995; Kocks *et al.*, 1995). ActA, est une protéine de surface ancrée dans la membrane par une région hydrophobe. Cette protéine de 610 acides aminés se caractérise par la présence d'une région centrale constituée de répétitions riches en résidus proline. Comme IcsA, ActA est distribuée de façon polaire à la surface de la bactérie. Lorsqu'on exprime ActA dans la bactérie non pathogène *Listeria innocua*, elle est capable de se déplacer, en polymérisant l'actine, dans un milieu constitué d'extraits d'œufs de xénopes (Marchand *et al.*, 1995; Theriot *et al.*, 1994). ActA et IcsA agissent de façon différentes mais le mécanisme général est similaire. En effet, la protéine IcsA recrute une protéine cellulaire N-WASp, qui a la capacité de recruter le complexe Arp2/3 (complexe de sept protéines). Ce complexe devient alors actif, c'est-à-dire qu'il est capable de catalyser l'étape limitante de nucléation dans le phénomène de polymérisation de l'actine (Egile *et al.*, 1999). La protéine ActA mimerait l'action de la protéine N-WASp en recrutant directement le complexe Arp2/3 par son domaine amino-terminal. Une des propriétés importantes du complexe Arp2/3 sous sa forme activée, est qu'il permet non seulement la mise en route de la polymérisation de l'actine, mais encore le branchement de filaments d'actine (Cossart, 2000).

Ogata et collaborateurs ont identifié une protéine (RickA) chez *R. conorii* (RC0909) qui possède un domaine similaire à ActA capable d'activer Arp2/3 et donc d'induire la polymérisation d'actine (Gouin *et al.*, 2004), même si celle-ci semble être un peu moins efficace. Les analyses en génomique comparative ont montré que les *Rickettsiales* ne possédaient pas toutes de tels gènes.

Bien que la stratégie d'internalisation de la bactérie ne soit pas clairement décrite, l'importance du cytosquelette dynamique d'actine sous-jacent à la membrane plasmique dans ces processus a été démontrée chez les *Rickettsiales* (Gouin *et al.*, 2004). Des voies de signalisation complexes utilisant des récepteurs membranaires et entraînant une augmentation de la phosphorylation de protéines sur des résidus tyrosine par des *tyrosine kinases* (une des principales caractéristiques de l'activation d'une voie de transduction du signal), aboutissent souvent à une réorganisation du cytosquelette d'actine. Des études récentes menées par Wakeel et collaborateurs chez *E. chaffeensis*, ont montré l'implication des protéines répétées en tandem TRPs (TRP47 et TRP120) (Tandem repeat proteins) dans l'internalisation de la bactérie (Wakeel *et al.*, 2010). L'association de ces protéines TRP47 exprimées uniquement à la surface des corps élémentaires à la tyrosine kinase FYN laisse supposer que celle-ci est recrutée par TRP47 pour induire ou faciliter l'entrée de la bactérie dans la cellule hôte. La protéine TRP47 semble être similaire à la protéine Tarp (translocated actin-recruiting phosphoprotein) de la bactérie *Chlamydia trachomatis* serovars L2 qui est phosphorylée par la protéine Ab1 kinase, induisant le mouvement d'actine au niveau du site d'entrée de la bactérie. La protéine TRP47 possède son homologue chez *E. ruminantium* (CDS Erum1110, ERGA CDS 01040).

Les protéines cellulaires CAPs (adenylate cyclase-associated proteins) ont été initialement identifiées chez les levures en tant que composants du complexe adénylate cyclase, intervenant notamment dans la régulation de l'AMP cyclique (Wakeel *et al.*, 2010). Les levures déficientes pour les protéines CAPs ne peuvent pas réorganiser leur cytosquelette. Dans les cellules animales, leur rôle dans la régulation de l'actine en réponse aux signaux cellulaires est largement conservé. CAP1 est une protéine monomérique hautement conservée et fixant l'actine (Wakeel *et al.*, 2010). Elle contient un domaine de fixation à l'actine (C-terminal), à adénylyl-cyclase et cofilin (N-terminal), et un domaine central « Src homology 3 » (SH3) et profilin, qui jouent un rôle central dans la modulation de l'actine (Hubberstey and Mottillo, 2002). Les études moléculaires chez les levures ont montré l'implication de CAP1 dans les mécanismes d'endocytose et de trafics cellulaires. Dans les cellules mammaliennes, CAP1 interagirait avec la protéine mAbp1-dynamine pour induire l'endocytose (Kessels *et al.*, 2001). mAbp1-dynamine est une protéine cellulaire « SH3 domaine dépendant » fixant l'actine.

Chez les bactéries du genre *Ehrlichia*, la protéine bactérienne TRP47, effecteur multifonctionnel, interagit également avec la protéine CAP1 (adenylate cyclase-associated protein 1) intervenant dans la fixation et la régulation des molécules d'actines de la cellule hôte, et aurait un rôle essentiel dans l'infection, la survie, l'invasion cellulaire et la mobilité de la bactérie dans la cellule (Wakeel *et al.*, 2010). Ces protéines sont impliquées dans divers processus du cycle cellulaire. A la fin du cycle cellulaire infectieux, les bactéries du genre *Ehrlichia* induiraient la lyse cellulaire par l'intermédiaire d'interactions entre les protéines TRP47 et CAP1 (Wakeel *et al.*, 2010). Sous l'action de la protéine TRP47, la protéine CAP1 permettant le transport d'actine au niveau de la mitochondrie, induirait une altération des mitochondries suivie d'une lyse cellulaire (apoptose) par voie mitochondriale caspase indépendant (Wang *et al.*, 2008).

### 3.2. Systèmes de sécrétion (SSTs)

Les bactéries pathogènes ont développé des stratégies sophistiquées pour infecter les cellules eucaryotes. En effet, l'interaction de bactéries pathogènes avec les cellules hôtes dépend de facteurs sécrétés dans le milieu extracellulaire qui facilitent l'adhésion, la virulence... On distingue deux types de transport de molécules effectrices : la sécrétion qui définit le transport actif de protéines du cytoplasme de la bactérie, à travers les membranes internes et externes, à la surface de la bactérie ou dans le milieu extracellulaire, et la translocation qui se réfère au transport de protéines à travers la membrane de la cellule cible. Les systèmes de sécrétions sont composés de plusieurs protéines associées qui sont également présentes à la surface des bactéries et agissant pour certaines (type III et IV) comme des seringues moléculaires injectant ainsi différents effecteurs de la bactérie dans la cellule. On pourrait diviser les différents systèmes de sécrétion en deux groupes selon qu'ils soient dépendants ou non du système Sec. Le système Sec transporte des protéines contenant une séquence N-terminale clivable du cytoplasme vers l'espace périplasmique (Figure 4) (Economou, 1999). Actuellement, six voies de sécrétions ont été identifiées chez les bactéries Gram négatif. Il existe trois voies séquentielles, les systèmes de sécrétion de type II, V et VI, nécessitant l'exportation d'intermédiaires périplasmiques et trois voies de sécrétion en une seule étape, les systèmes de sécrétion de type I, III et IV (Figure 4). Les SSTs sont conservés au niveau

structural entre les bactéries Gram négatif et présentent de grandes similitudes avec le corps basal du flagelle (Tampakaki *et al.*, 2004).

### *3.2.1. Les systèmes de sécrétion Sec-dépendant*

- Le système de sécrétion de type II (SST2) correspond au système général de sécrétion de protéines des bactéries pathogènes et non pathogènes. Certaines protéines effectrices contenant un signal N-terminal sont exportées à travers la membrane via ce système. Le SST2, le mieux connu des voies de sécrétion, s'effectue en deux étapes. La première étape consiste en un transport des protéines du cytoplasme vers le périplasme par le système *Sec*. La deuxième étape correspond à l'exportation vers le milieu extérieur à travers la membrane externe par une voie différente de la première. L'exemple le plus étudié est la sécrétion de pullulanase par *Klebsiella oxytoca*. Chez *P. aeruginosa*, le SST2, également appelé Xcp, est impliqué dans la sécrétion de nombreuses enzymes et toxines, telles que l'élastase, l'exotoxine A, la lipase, les phospholipases ou la phosphatase alcaline (Filloux *et al.*, 1990; Filloux *et al.*, 1998). De plus, des protéines transposées par le SST2 telles que les phospholipases ont été identifiées chez les *Rickettsiales* (Winkler and Daugherty, 1989; Walker *et al.*, 2001).

- Le système de sécrétion type V (SST5) est connu comme étant un auto-transporteur. L'auto-transporteur, contient non seulement la séquence N-terminale clivable mais aussi l'information pour le passage à travers la membrane externe, il peut ainsi passer la membrane externe par lui-même (Kostakioti *et al.*, 2005) (Fig. 1). Chez *P. aeruginosa*, l'estérase EstA a été identifiée comme auto-transporteur (Wilhelm *et al.*, 1999) et deux autres homologues ont été identifiés (Ma *et al.*, 2003). Chez *Escherichia coli* et *Salmonella enterica* ce système est codé par certains îlots de pathogénicité. Ces protéines sont responsables de leur propre sécrétion à travers la membrane interne et externe des bactéries Gram négatif. Chez les *Rickettsiales*, beaucoup de protéines membranaires, notamment les protéines OmpA/B, ou Sca 1-3, sont reconnue pour leur appartenance à la famille des auto-transporteurs (Walker and Yu, 2005).

### *3.2.2. Les systèmes de sécrétion Sec-indépendant*

- Le système de sécrétion de type I, ou type de sécrétion ABC, (Figure 4) est constitué de 3 protéines, une protéine de membrane externe (OMP : Outer membrane protein) qui forme un pore externe, une protéine périplasmique (HlyD) et une protéine incluse dans la membrane interne qui sert de transporteur avec une activité ATPase (ABC transporteur, HlyB). Ce système permet entre autre le transport de toxines RTX, protéases, lipases. Il peut également participer au transport d'effecteurs et au recrutement de molécules essentielles au métabolisme de la bactérie. Ce système de sécrétion est connu par exemple chez une souche UPEC d'*E. coli* où il est impliqué dans la synthèse, l'activation et le transport d'alpha-hémolysine (HlyA) (Braun *et al.*, 1993). La bactérie *E. ruminantium* possède plusieurs gènes codant pour les transporteurs ABC (*LolD*, *pstB*...). De plus, des protéines transposées par le système de type I ont été identifiées chez les *Rickettsiales*.

- Système de sécrétion de type III (SST3), par sa composition et son fonctionnement, est le plus complexe des systèmes de sécrétion bactériens connus et est présent dans de nombreux genres de bacilles Gram négatif dont *Yersinia*, *Salmonella*, *Shigella* et les *Escherichia coli* entéropathogènes (Hueck, 1998). La sécrétion de type III est indépendante du système *Sec*, cependant son assemblage requiert probablement le système *Sec*, car plusieurs composants possèdent des séquences signal similaires à celles trouvées dans le système *Sec*. Le SST3 est composé de plus de vingt protéines et la plupart des protéines de la membrane interne sont homologues aux protéines de l'appareil de biosynthèse du flagelle des bactéries Gram négatif et positif (Cornelis, 2006). Ce système est dédié à l'injection d'effecteurs directement du cytoplasme de la bactérie dans le cytosol de la cellule cible. Ces effecteurs ont pour effets la mort cellulaire ou le dysfonctionnement de la cellule en détruisant la membrane, en perturbant les cascades de transduction des signaux ou en désorganisant le cytosquelette (Hueck, 1998). Ils permettent aux bactéries pathogènes d'envahir les tissus de l'organisme infecté, de circonscire ses défenses, de détourner la machinerie cellulaire afin d'acquérir des nutriments et de permettre son métabolisme et sa multiplication (Tseng *et al.*, 2009). Beaucoup de protéines TRP sécrétées par le SST3 ont été rapportées pour leur implication dans l'adhésion à la cellule hôte, la polymérisation d'actine et l'évasion au système immunitaire, notamment chez les *Chlamydiales*. Cependant le système de sécrétion de type III n'a encore jamais été

identifié chez les *Rickettsiales*. Les études complémentaires menées ont également permis d'écartier l'hypothèse d'une sécrétion des TRPs par le SST4 (Wakeel *et al.*, 2010).

- Le système de sécrétion de type IV (SST4) est similaire au SST3. Il correspond à une famille multi-génique codant des protéines membranaires internes, périplasmiques et externes. Il est impliqué dans la translocation de molécules dans les cellules hôtes eucaryotes, l'exportation d'effecteurs protéiques, de facteurs de virulence, d'acides nucléiques et est impliqué dans la pathogénèse de plusieurs bactéries. Ce système a été largement étudié dans le cadre du transfert du plasmide Ti d'*Agrobacterium tumefaciens* dans des cellules de plantes. Ce système est codé par des gènes d'îlots de pathogénicité chez plusieurs bactéries pathogènes comme *H. pylori* (responsable de la sécrétion de la protéine Cag), *Bordetella pertussis*, *Brucella suis* (Schmidt and Hensel, 2004).

Les analyses *in silico*, ont pu mettre en évidence la présence d'un SST4 chez les *Rickettsiales*, incluant les bactéries du genre *Ehrlichia*. Plusieurs ORFs similaires au SST4 virB ont été identifiés chez *R. typhi*, *R. prowazekii* et *R. conorii* (Walker and Yu, 2005). La bactérie *E. ruminantium* possède cinq « clusters » distants de gènes *vir* (*D. Meyer, communication personnelle*). L'ensemble des génomes de *Rickettsiales* séquencés présentent 4 paralogues de la protéine VirB6 dont l'expression et la fonctionnalité ont été démontrées, notamment chez *E. chaffeensis* (Bao *et al.*, 2009). Chez la rickettsie *Anaplasma phagocytophilum*, la protéine anti-apoptotique Ats-1, qui joue un rôle dans le maintien de l'intégrité mitochondriale et donc dans la survie de la bactérie est sécrétée par le SST4.



*Pseudomonas aeruginosa* et les *Rickettsiales* en général) (Wakeel *et al.*, 2010). La fonction de ces protéines atypiques chez les bactéries est encore largement inconnue mais elles sont fortement suspectées de participer aux interactions des bactéries avec leurs hôtes, notamment *via* la modulation transcriptionnelle des gènes de la cellule hôte en se fixant directement à la chromatine. Elles sont impliquées dans le cycle cellulaire, l'organisation du cytosquelette, la toxicité ou encore la réponse inflammatoire (Wakeel *et al.*, 2010). Les *Rickettsiales* séquencées posséderaient au moins deux protéines à motifs Ankyrines dans leur génome, par exemple les orthologues Ank200 et AnkA chez *E. chaffeensis* et *A. phagocytophilum*, qui sont sécrétées par le SST4 (Rikihisa and Lin, 2010). Chez *Ehrlichia ruminantium* l'analyse du génome a pu révéler la présence de quatre protéines à motifs ankyrines, dont les deux CDS orthologues ERGA 04060 / ERWE 04110 et ERGA 06440 / ERWE 06530 chez les souches Gardel et Welgevonden respectivement.

### 3.3.2. Inhibition de la fusion phagosome-lysosome

Semblable au modèle *Chlamydia*, les bactéries du genre *Ehrlichia* modifient leur phagosome infectieux, appelé aussi inclusion ou vacuole, en exprimant très tôt leurs propres protéines à la surface des phagosomes afin d'éviter l'apoptose des cellules. Les bactéries survivent en position intracellulaire dans les vacuoles infestées en inhibant la fusion phagolysosomiale. Les premières études réalisées sur ce phénomène d'échappement supposaient que l'inhibition de la fusion phagosome-lysosome était étroitement liée à l'activité métabolique de la bactérie (Brouqui, 1997; Wells and Rikihisa, 1988). Ainsi, chez *E. risticii* et *E. sennetsu*, en bloquant la synthèse des protéines bactériennes par l'action des tétracyclines, on restaure la fusion phagolysosomiale (Wells and Rikihisa, 1988; Brouqui, 1997). Bien que le mécanisme précis ne soit pas clair, il est probable que les protéines bactériennes exprimées à la surface de ces phagosomes infectieux soient responsables du blocage de la fusion avec les compartiments d'endocytose. Dans les cellules eucaryotes, la fusion des membranes intracellulaires est provoquée par des membres de la famille des protéines SNARE. Ces protéines sont conservées chez tous les eucaryotes et sont présentes à la surface de tous les compartiments de sécrétion. Lors de la fusion membranaire, des t-SNARE (t pour target) présents sur les organites cibles vont interagir spécifiquement avec les V-SNARE présents sur les vésicules

plasmiques. Au cours des dernières années, un nombre croissant de SNARE-like protéines ont été identifiées notamment chez *Chlamydia* et *Legionella* (Paumet *et al.*, 2009). Ces protéines existent chez *Ehrlichia* et sont responsables de l'inhibition de la fusion phagolysosomiale car elles inhibent directement l'action des protéines SNARE eucaryotes.

### 3.3.3. L'échappement au stress oxydatif

Lorsqu'une cellule eucaryote (monocytes, macrophages, neutrophiles) est infectée par une bactérie, elle augmente sa consommation en oxygène (poussée respiratoire) produisant ainsi des dérivés oxygénés, les anions superoxydes ( $O_2^-$ ), les peroxydes d'hydrogène ( $H_2O_2$ ) (ROS, *Reactive Oxygen Species*) en réponse à l'agression. La formation du  $H_2O_2$  se fait par dismutation spontanée ou par le biais d'enzymes, superoxydes dismutases (SOD). L' $H_2O_2$  et  $O_2^-$  ne sont pas des antibactériens puissants mais ils peuvent se transformer en radicaux libres qui vont agir sur l'intégrité de la membrane cellulaire ou bactérienne. Les anions superoxydes en association avec des produits de dégradation de la L-arginine (autre mécanisme antibactérien) vont former des peroxy-nitrites qui sont hautement toxiques et ont une action antimicrobienne importante.

Les pathogènes intracellulaires possèdent des systèmes anti-oxydants qui vont être capables de détecter une modification de l'homéostasie redox, puis de mettre en place la réponse la plus appropriée pour la corriger. Le mécanisme de défense contre les superoxydes, composé de superoxyde dismutases (SOD) existe chez les bactéries. Les *Rickettsiales* y compris *Ehrlichia ruminantium* possèdent la superoxyde dismutase (SodB) qui est une enzyme anti-oxydante majeure. Leur expression est induite en présence de  $H_2O_2$  (Barloy-Hubler *et al.*, 2004). Chez *E. coli*, la lutte contre les peroxydes, les anions superoxydes et les composés non radicalaires implique deux régulons différents : le régulon OxyR et le régulon SoxR. Le régulon OxyR contrôle la plupart des anti-oxydants en réponse aux peroxydes et à l'oxyde nitrique et SoxR contrôle la réponse aux anions superoxydes. OxyR, est un régulateur transcriptionnel de la famille de LysR, et un senseur du stress oxydatif grâce à deux protéines particulières (Trx1 et Trx2), qui peuvent être oxydées par  $H_2O_2$ . La forme oxydée d'OxyR induit l'expression de ses gènes cibles et réprime sa propre expression. OxyR peut donc être soit un activateur, soit un répresseur, selon son état oxydatif (Zeller and Klug, 2006).

Les thiorédoxines (Trx) et glutarédoxines (Grx) médiées par les régulateurs OxyR et SoxR, agissent comme des anti-oxydants en facilitant la réduction de protéines oxydées (Holmgren, 1989). L'importance de Trx et Grx chez divers pathogènes a été démontrée notamment chez *Salmonella enterica* serovar Typhimurium. La délétion de gènes appartenant au complexe Trx chez cette bactérie diminue de façon significative sa virulence et sa capacité de réplication dans la cellule hôte. La délétion de gènes appartenant au complexe Grx est corrélée à une diminution de la résistance au stress oxydatif, ce qui montre leur implication dans divers processus de survie (Bjur *et al.*, 2006). Chez *Rickettsia conorii* isolée dans les escarres de patients infectés, l'expression de Trx (*trxB2*) et Grx (*grxC1*) augmente en réponse au stress oxydatif généré par la cellule (Renesto *et al.*, 2008). Les Trx jouent un rôle majeur dans la synthèse de l'ADN et la réparation des protéines. Chez *E. coli* de nombreuses fonctions cellulaires semblent liées au gène *trxI* parmi lesquels la division cellulaire, la régulation transcriptionnelle, la transduction d'énergie ou encore la protection des protéines contre la dégradation (maintien de la conformation).

### 3.3.4. Le stress osmotique

Les bactéries possèdent des moyens de protection contre les stress osmotiques auxquels elles sont soumises au sein de la cellule hôte. Les variations osmotiques étant délétères pour les bactéries, celles-ci possèdent des systèmes de transport et/ou de synthèse d'osmo-protectants ou solutés compatibles qui assurent le maintien de l'homéostasie par leur accumulation dans le cytoplasme ou leur rejet selon le type de stress osmotique (Smith, 1996). Les solutés compatibles, tels que la carnitine, la proline ou la glycine bêtaïne, sont des composés organiques qui peuvent être accumulés à forte concentration dans le cytoplasme sans interférer avec les processus cellulaires et qui rétablissent l'osmolarité cellulaire (Patchett *et al.*, 1994). Certaines bactéries sont capables de synthétiser la glycine bêtaïne, soit par méthylation de la glycine, soit par oxydation de la choline (Landfald and Strom, 1986). Cependant, la plupart possèdent et utilisent des systèmes de transport qui permettent une importation rapide et en masse de ce composé depuis le milieu environnant. Le système de transport le mieux caractérisé est le système de transport *proU* à haute affinité pour la glycine bêtaïne d'*E. coli*. De nombreux transporteurs proline-bêtaïne (ProP) ont été mis en évidence

chez les *Rickettsiales* en particulier *R. conorii* avec 11 copies de *proP*, *R. typhi* (7 copies), chez *R. prowazekii* (7 copies) ou encore *E. ruminantium* (4 copies). Leur induction a été démontrée en réponse au stress osmotique chez *R. conorii* en condition d'infection *in vivo* (Renesto *et al.*, 2008).

### 3.3.5. Système d'acquisition du fer, régulation et homéostasie

Le fer est un élément essentiel pour la plupart des organismes vivants, y compris pour les bactéries. Notamment, de nombreuses enzymes ont un site actif fer-dépendant et la déplétion en fer induit une inhibition de l'activité enzymatique (Paradkar *et al.*, 2008). Quelque soit l'environnement colonisé par les bactéries, sa concentration en fer est très faible, excepté quelques milieux tels que l'intestin, le colon, ou l'estomac (pH acide). Face à cette carence en fer, la plupart des bactéries utilisent plusieurs mécanismes visant à acquérir le fer de diverses sources. Le fer peut se trouver sous deux formes : la forme réduite ferreuse et soluble ( $\text{Fe}^{2+}$ ) qui diffuse librement à travers les membranes ou la forme oxydée et insoluble ( $\text{Fe}^{3+}$ ) plus difficile d'accès. Dans l'organisme hôte, le fer n'existe pas sous sa forme libre et est généralement complexé à des protéines telles que la transferrine, la lacto-ferrine et la ferritine soit par une liaison directe avec les acides aminés, soit par liaison indirecte par l'intermédiaire de l'hème *via* les hémoprotéines (hémoglobine, myoglobine). La séquestration du fer peut être considérée comme un mécanisme de défense chez les cellules eucaryotes infectées par la bactérie. Afin de surmonter cette carence en fer, et de pouvoir se développer chez l'hôte, les bactéries ont développé deux principaux mécanismes qui leur permettent d'acquérir le fer de l'hôte. Ainsi les bactéries pathogènes possèdent plusieurs mécanismes faisant intervenir notamment les sidérophores qui ont la propriété de se fixer au fer ferrique ( $\text{Fe}^{3+}$ ), et aux protéines de l'hôte liées au fer (Collins, 2003). Les entérobactines synthétisées par les bactéries entériques, dont fait partie *Salmonella*, sont parmi les chélateurs de fer les plus puissants. Caractéristiques des bactéries Gram négatif, les hémophores sont des protéines capables de chélater l'hème de plusieurs sources d'hémoprotéines avec une forte affinité et permettent sa reconnaissance au niveau des récepteurs spécifiques présents à la surface bactérienne (Cescau *et al.*, 2007). L'hémophore HxuA, sécrété par *Haemophilus influenzae*, lui permet d'acquérir l'hème libre ou lié à l'hémopexine. Un deuxième système d'acquisition

d'hème basé sur l'hémophore, HasA (heme acquisition system) a été identifié chez plusieurs bactéries, parmi lesquelles *Serratia marcescens*, *Pseudomonas aeruginosa*, *P. fluorescens*, *Yersinia pestis*, et *Yersinia enterocolitica* (Wandersman and Delepelaire, 2004). Chez la bactérie *E. ruminantium* de nombreux gènes participant au métabolisme de l'hème ont été identifiés : *HemA*, *HemB*, *HemC*, *HemE*, *HemF*, *HemH*. Chez les bactéries Gram négatif, la source de fer doit traverser les deux membranes (externe et cytoplasmique) constituant la paroi cellulaire en passant par le périplasma, afin d'être utilisée par la bactérie. L'internalisation des systèmes de fer complexés dans le cytoplasme fait intervenir des récepteurs spécifiques au niveau de la membrane externe, puis des transporteurs de type ABC au niveau de la membrane interne (Faraldo-Gomez and Sansom, 2003; Rodionov *et al.*, 2006; Wandersman and Delepelaire, 2004).

Il est important que la bactérie contrôle l'homéostasie en fer étant donné son besoin vital en fer mais aussi la toxicité du fer à l'état ferreux. La concentration en fer se fait par une régulation fine des systèmes d'acquisition et de stockage impliquant des protéines régulatrices bien déterminées (Hantke, 2001). Chez la plupart des bactéries Gram négatif, la transcription des gènes impliqués dans le transport du fer est réalisée par l'intermédiaire du régulateur transcriptionnel Fur (Ferric uptake regulator) qui présente une activité dépendante du fer (Rodionov *et al.*, 2006). Lorsque la bactérie est dans un environnement riche en fer ferreux, Fur lie l'ion ferreux au niveau C-terminal et acquiert une conformation qui augmente fortement son affinité pour sa cible d'ADN. La liaison de Fur à l'ADN inhibe la transcription des gènes impliqués dans la synthèse et le transport des sidérophores (Rodionov *et al.*, 2006). Par contre, quand la concentration de fer est faible, Fur libère le fer et quitte l'ADN permettant à l'ARN polymérase d'accéder aux promoteurs et par conséquent les gènes en question seront transcrits (Rodionov *et al.*, 2006). Chez les  $\alpha$ -protéobactéries, de nombreux régulons intervenant dans la régulation et l'acquisition du fer, le transport de l'hème, la biogénèse et le transport des sidérophores ont été identifiés en particulier Fur, IscR, Irr ou ICEs, RirA, Mur et MntR (Rodionov *et al.*, 2006). Chez les *Rickettsiales*, le régulateur transcriptionnel IscR (*Isc* cluster ou « IscR box») semble intervenir dans la répression des gènes qui participent à la biogénèse des systèmes d'acquisitions de fer.

### 3.4. Autres effecteurs

Certaines enzymes bactériennes ayant des activités membrano-lytiques (lyse des membranes de la cellule hôte) ou hémolytique (lyse des globules rouges) ont été détectées chez certaines *Rickettsiales* (Winkler and Daugherty, 1989; Walker *et al.*, 2001). Il s'agit de la phospholipase D et de *PAT-1* (patatin-like phospholipase) (chez *Rickettsia typhi*, *Rickettsia conorii*). L'expression de ces gènes est corrélée à la capacité qu'ont les bactéries du genre *Rickettsia* à s'échapper rapidement du phagosome cytoplasmique, leur permettant ainsi de survivre librement dans le cytoplasme de la cellule hôte. Cependant, ces gènes ne sont pas présents chez toutes les rickettsies notamment chez *Ehrlichia ruminantium* (Walker and Yu, 2005).

## 4. L'aire de la post-génomique: Profil d'expression des *Rickettsiales*

### 4.1. Innovations scientifiques : caractérisation des agents pathogènes intracellulaires

L'importance économique et la gravité des maladies provoqués par les *Rickettsiales*, sur les animaux et l'homme, nécessite de développer des stratégies de lutte efficace (contrôle ou éradication des vecteurs, production de vaccins de seconde génération). Cependant, le caractère intracellulaire obligatoire représente un obstacle important dans l'étude de la pathogenèse mais également un frein dans la lutte contre ce type de bactérie. Ainsi, des approches innovantes (biotechnologie, génomique, analyse du transcriptome, culture cellulaire...) doivent sans cesse être développées et améliorées dans le but d'élucider les mécanismes de virulence de ces agents pathogènes. Ces dernières années, les améliorations concernant les conditions de culture *in vitro*, l'isolement de souches, la conservation des vaccins, mais aussi le séquençage de plusieurs *Rickettsiales* ont engendré des applications importantes au niveau de la caractérisation moléculaire (Bekker *et al.*, 2005), en génomique comparative (Frutos *et al.*, 2006a; Frutos *et al.*, 2006b; Fuxelius *et al.*, 2007), transcriptomique avec les puces à ADN (La *et al.*, 2007; Renesto *et al.*, 2008), protéomique (Ge and Rikihisa, 2007; Huang *et al.*, 2008), permettant ainsi des avancées significatives sur la connaissance des mécanismes qui interviennent dans la pathogenèse et la virulence de ces organismes.

Ainsi, plusieurs méthodes d'identifications des gènes exprimés par les bactéries pathogènes dans des conditions d'infections *in vivo* ou *in vitro* sont couramment utilisées aujourd'hui. Certaines d'entre elles sont applicables à l'analyse transcriptomique utilisant la méthode des puces à ADN ou microarrays.

#### 4.2. La technique d'expression *in vivo* ou *In Vivo* Expression Technique (IVET)

La technique d'expression *in vivo* (IVET) a été élaborée pour identifier par sélection positive les gènes spécifiquement induits chez un micro-organisme lorsqu'il infecte un animal ou une plante hôte (Mahan *et al.*, 1993). Le système est basé sur un plasmide contenant un opéron hybride dépourvu de promoteur où les gènes *lacYZ* sont fusionnés à un gène *purA* ou *thyA* situé en aval du site de clonage unique. Cet opéron hybride a été construit dans un plasmide suicide. Le clonage des fragments d'ADN pathogènes dans le site BglIII produit un ensemble de fusions transcriptionnelles dirigées par les promoteurs qui sont présent dans l'ADN cloné. L'ensemble de ces fusions est transféré dans une souche du même pathogène déficiente pour le gène *purA* ou *thyA*, dont on sélectionne l'intégration dans le chromosome. Les pathogènes recombinants sont utilisés pour infecter un animal traité avec un antibiotique. Les fusions qui contiennent un promoteur actif dans l'animal permettent la transcription du gène *purA* ou *thyA* et donc la survie de la bactérie. On teste *in vitro* le niveau de la  $\beta$ -galactosidase chez les pathogènes survivants, isolés à partir de l'animal traité. Les clones qui contiennent des fusions avec des gènes spécifiquement induits dans l'animal traité n'expriment pas *lacZ* sur les milieux de culture et on peut donc identifier les gènes auxquels appartiennent les promoteurs en séquençant la portion d'ADN en amont de *purA*. Le système IVET a été initialement conçu pour étudier un modèle murin de la typhoïde « Typhimurium ». Depuis sa première description l'IVET a été utilisé pour une grande quantité de pathogènes Gram négatifs et Gram positifs dont *Mycobacterium* et *Actinobacillus pleuropneumoniae* dans un modèle porcin (Fuller *et al.*, 1999).

Cette technique a pour avantage d'être relativement facile à réaliser et ne nécessite pas d'appareillages ou matériels particulièrement coûteux. Elle permet d'identifier des gènes induits lors de l'infection. Cependant elle ne permet pas d'identifier globalement tous les gènes induits *in vivo* puisqu'elle repose sur le séquençage des bactéries isolées post-infection.

Un des désavantages du système IVET est sa tendance intrinsèque à sélectionner des promoteurs exprimés de façon permanente *in vivo* et à un niveau élevé. De plus, l'insertion du plasmide peut désactiver un gène essentiel pour l'infection dans quel cas, la souche ne pourra pas être sélectionnée et le promoteur ne sera pas identifié ce qui représente un désavantage non négligeable. Un certain nombre de modifications de la technique permettant d'identifier des gènes qui ne sont exprimés que de façon transitoire au cours de l'infection chez l'animal (opéron *tnpR*) a été apporté (Camilli and Mekalanos, 1995; Handfield and Levesque, 1999).

#### 4.3. La mutagenèse à l'aide de transposon portant un marqueur-signature ou Signature tagged mutagenesis (STM)

La technique, Signature tagged mutagenesis (STM) est une variante de l'utilisation de la mutagenèse à l'aide de transposons, qui a été utilisé pour identifier des gènes de virulence chez les bactéries. Le principe de base est de créer à l'aide de transposons, un grand nombre de mutants différents d'un organisme pathogène (banque de mutants), puis d'identifier les mutants capable de survivre *in vitro*, mais pas *in vivo*. Il s'agit donc d'un criblage de tout le génome pour trouver les gènes spécifiques d'un environnement donné. Les mutants sont produits par l'insertion de transposons qui ont chacun une étiquette composée d'une séquence unique placée entre deux séquences connues, pour pouvoir amplifier l'étiquette par PCR. Ainsi chaque mutant peut être identifié par Southern blot en utilisant la séquence de l'étiquette comme sonde. En comparant le profil d'hybridation de la banque de mutant avant et après infection *in vivo*, il est possible d'identifier les mutants qui n'ont pas survécu. Après clonage de la région d'ADN contenant le transposon, il est possible d'identifier le gène muté par séquençage. L'utilisation de la STM a été décrite pour la première fois pour l'identification de gènes de virulence de *Salmonella Typhimurium* dans un modèle murin de la fièvre typhoïde (Hensel *et al* 1995). Grâce à cette technique, on peut tester un grand nombre de mutant différents et simultanément dans le même animal, à condition que la banque de mutant soit assez complexe. Par contre, puisque l'identification est basée sur la technique de southern blot et du séquençage, qui ne sont pas des méthodes à haut débit, le nombre de gènes identifiés est limité. Pour palier à ces problèmes, la STM a été utilisée conjointement aux puces à ADN (Sasseti *et al.*, 2001, 2003). Cette variante appelée TraSH, utilise des

transposons contenant le promoteur pour la polymérase à ARN T7 dirigé vers l'extérieur. Il est donc possible de transcrire l'ADN entourant le transposon et de marquer l'ARN obtenu avec des fluorochromes. L'ARN marqué est ensuite hybridé aux puces à ADN. Ainsi, le signal obtenu avec l'ADN de la banque de mutants récolté *in vivo* est comparé au signal *in vitro* pour identifier les gènes qui lorsqu'ils sont mutés, ne permettent pas au mutant de survivre lors de l'infection. Cette méthode a été utilisée pour identifier les gènes nécessaires à l'infection de macrophages murins et à l'infection persistante de la souris par *Typhimurium* (Chan *et al.*, 2005; Lawley *et al.*, 2006).

#### 4.4. Hybridation soustractive suppressive (SSH)

La méthode d'hybridation soustractive suppressive (SSH) est une technique puissante qui permet de comparer l'expression de deux populations de pathogènes (deux souches ou une même souche dans des conditions différentes) et d'identifier ainsi des gènes qui sont exprimés dans une population et pas dans l'autre. Cette technique est appropriée pour l'identification de messagers uniques ou rares. En outre, elle est unique, car elle est basée sur une amplification sélective des séquences exprimées différemment (Diatchenko *et al.*, 1999; Gurskaya *et al.*, 1996). Avec un facteur d'enrichissement allant de 1000 à 5000, la SSH est une méthode de choix pour rechercher et cloner des séquences rares. Cependant la mise en œuvre de la méthode peut être limitée dans certains cas par la quantité nécessaire de matériel de départ.

#### 4.5. Selective Capture of Transcribed Sequences (SCOTS)

La technique SCOTS a été initialement développée par Graham and Clark-Curtiss en 1999 afin d'identifier des gènes de *Mycobacterium tuberculosis* exprimés lors de l'infection de macrophages humains en culture (Graham and Clark-Curtiss, 1999). Cette technique est aujourd'hui largement utilisée et a permis de révéler le profil global d'expression génique de la bactérie *Salmonella Typhi* à partir de macrophages humains infectés (Faucher *et al.*, 2006; Faucher, 2007).

D'une façon générale, cette technique repose sur la capture spécifique des transcrits bactériens (Figure 5). L'ARN total du pathogène est d'abord isolé et converti en ADNc en utilisant des

amorces spéciales. La partie 3' est un nonamère aléatoire (N9) tandis que la partie 5' est une séquence « étiquette » connue qui n'a pas d'homologie dans le génome des espèces étudiées. Ainsi, la conversion de l'ARN en ADNc se fait de façon aléatoire et ajoute la séquence étiquette à chacun des transcrits. Les transcrits peuvent ensuite être amplifiés par PCR en utilisant comme amorce la séquence étiquette. Une fois l'ADNc synthétisé, il s'agit de capturer les transcrits bactériens et de se débarrasser des transcrits eucaryotes et des transcrits bactériens codants pour l'ARN ribosomal (ARNr). Pour y arriver, l'ADN génomique (ADNg) de la bactérie est biotinylé et pré-hybridé avec un excès d'ADN ribosomal cloné (ADNr) bloquant ainsi les sites d'hybridation à l'ARNr. Vient ensuite l'étape de capture des transcrits, où l'ADNg ainsi préparé est hybridé avec les ADNc provenant du pathogène. Seuls les ADNc bactériens pourront s'hybrider à l'ADNg. Les ADNc codant pour l'ARNr ne seront pas capturés puisque les zones homologues sur l'ADNg sont déjà bloquées par l'ADNr. Les hybrides ADNc-ADNg sont récupérés à l'aide de billes magnétiques couvertes de streptavidine. Ensuite, les ADNc sont élués et amplifiés par PCR. Deux autres rondes de capture sont effectuées pour augmenter la quantité de transcrits spécifiques du pathogène et diminuer les contaminants (transcrits eucaryotes et ARNr) (Faucher, 2007). Les ADNc ainsi obtenus sont alors clonés et séquencés, permettant donc d'identifier les gènes bactériens exprimés pendant l'infection. Alternativement, les ADNc obtenus par SCOTS peuvent être hybridés aux puces à ADN pour identifier globalement les gènes exprimés pendant l'infection. La méthode d'enrichissement SCOTS a été utilisée pour plusieurs applications différentes notamment pour l'analyse de l'expression par micro-arrays (Faucher *et al.*, 2006; Poirier *et al.*, 2008), la production de banque de clones (Daigle *et al.*, 2001; Dozois *et al.*, 2003). Par exemple, la méthode d'enrichissement qui consiste à effectuer un second SCOTS en bloquant l'ADNg avec les ADNc provenant d'un premier SCOTS d'une condition (*in vitro*), peut permettre d'identifier des gènes exprimés seulement *in vivo*. Daigle et collaborateurs ont eu recours à cette approche pour identifier des gènes de *Salmonella enterica serovar typhi* exprimés différemment dans le milieu de culture des macrophages humains et dans les macrophages eux-mêmes (Daigle *et al.*, 2001). Les études réalisées sur *Helicobacter pylori* au cours de l'infection de la muqueuse gastrique (Graham *et al.*, 2002) ou encore sur *Salmonella Typhi* au cours de l'infection des macrophages humain en cultures (Faucher *et al.*, 2006; Faucher, 2007) ont illustré tout l'intérêt de l'utilisation SCOTS pour

une analyse par microarrays dans le cas de pathogènes d'intracellulaires. La sensibilité de détection des gènes exprimés par le pathogène augmente avec les rondes de capture SCOTS, avec une efficacité élevée puisque jusqu'à 75% des gènes sont détectés. L'avantage majeur de SCOTS est que cette technique ne requiert pas de modification génétique de la bactérie étudiée, contrairement à l'IVET et à STM. Cependant avant notre étude, aucune technique d'hybridation sélective ou soustractive n'avait été utilisée pour un pathogène intracellulaire obligatoire de la famille *Rickettsiales*.

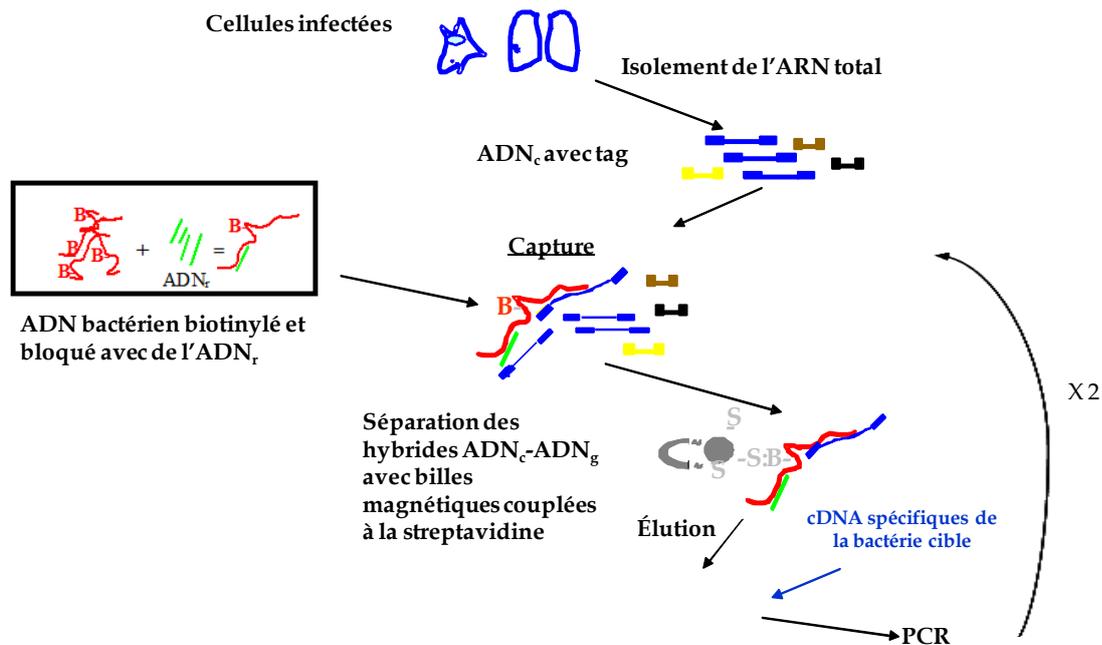


Figure 5: Selective Capture of Transcribed Sequences (SCOTS) (d'après Daigle *et al.*, 2001)

#### 4.6. Puces à ADN

##### 4.6.1. Principe

Le développement de la génomique a conduit à l'émergence de nouvelles technologies, comme les puces à ADN, qui permettent d'appréhender simultanément l'expression de l'ensemble des gènes d'un organisme par l'étude de son transcriptome. Aujourd'hui les puces ADN sont utilisées comme outils de choix pour obtenir le profil d'expression global d'un

pathogène dans plusieurs conditions au cours de son cycle infectieux ou encore dans des environnements différents. Pour la majorité des puces qui sont actuellement utilisées la technique d'analyse est basée sur une mesure relative de l'expression des gènes entre deux conditions. L'idée conceptuelle de la puce à ADN est très simple. Il s'agit de greffer sur une lame de verre des fragments « synthétiques » d'ADN (les sondes) représentatifs de chacun des gènes étudiés pour un organisme donné (Ramsay, 1998; Rockett *et al.*, 2001). Tous les gènes de la bactérie à l'étude sont préalablement déposés sur une lame de verre et servent de sonde. Selon ce que l'on veut comparer, on peut greffer quelques gènes voir un génome entier sur la lame. Par la suite on hybride sur cette lame des transcrits obtenus d'une condition test et d'une condition contrôle marquées par des fluorochromes différents, généralement les Cyanine 3 et 5 (Cy3 et Cy5) (hybridation compétitive). On peut aussi hybrider séparément des transcrits obtenus d'une condition test et comparer ainsi plusieurs conditions hybridées indépendamment entre elles à l'aide d'un seul et même fluorochrome, en général le Cy3 qui est plus stable (Figure 6). Après hybridation des transcrits à chaque sonde, la mesure des intensités des signaux de fluorescence relatifs à chacun des fluorochromes permet finalement de calculer pour chaque gène le rapport des intensités, qui évalue son expression différentielle.

#### 4.6.2. Les limites pour l'analyse de l'expression de gènes de bactéries intracellulaires obligatoires

A l'origine, les puces à ADN ont été appliquées au monde bactérien pour étudier l'expression génique dans des conditions *in vitro* (Conway and Schoolnik, 2003). Renesto et collaborateurs ont réalisé la seule étude transcriptomique dans des conditions *in vivo* pour une *Rickettsie*. Il s'agit d'une étude montrant les différentes réponses transcriptionnelles de *R. conorii* dans les escarres de patients infectés (Renesto *et al.*, 2008). Les difficultés rencontrées au cours de l'analyse du transcriptome de telles bactéries sont nombreuses. Les rickettsies sont des bactéries intracellulaires obligatoires donc difficiles à manipuler. Leur culture *in vitro* sur cellules eucaryotes peut les soumettre à des stress liés à la manipulation ou aux modifications environnementales, modifiant ainsi très rapidement leurs profils d'expressions. L'obtention d'ARN messager (ARNm) bactérien en qualité et quantité suffisante est un pré-requis pour

l'analyse du transcriptome. De plus, l'ARN bactérien possède une demie-vie courte et peut rapidement être dégradé. De ce fait, l'ARN doit être stabilisé pendant l'extraction, ce qui s'avère difficile pour les bactéries intracellulaires (La *et al.*, 2007). Enfin, l'ARN bactérien se compose en majorité d'ARN ribosomal (82%), alors que les ARN messagers représentent 2% des ARN totaux. Lors des extractions en condition d'infection, l'ARN eucaryote est considérablement plus abondant que l'ARN bactérien. Ce qui a pour effet de diminuer sensiblement le signal obtenu après hybridation en augmentant le bruit de fond sur les lames.

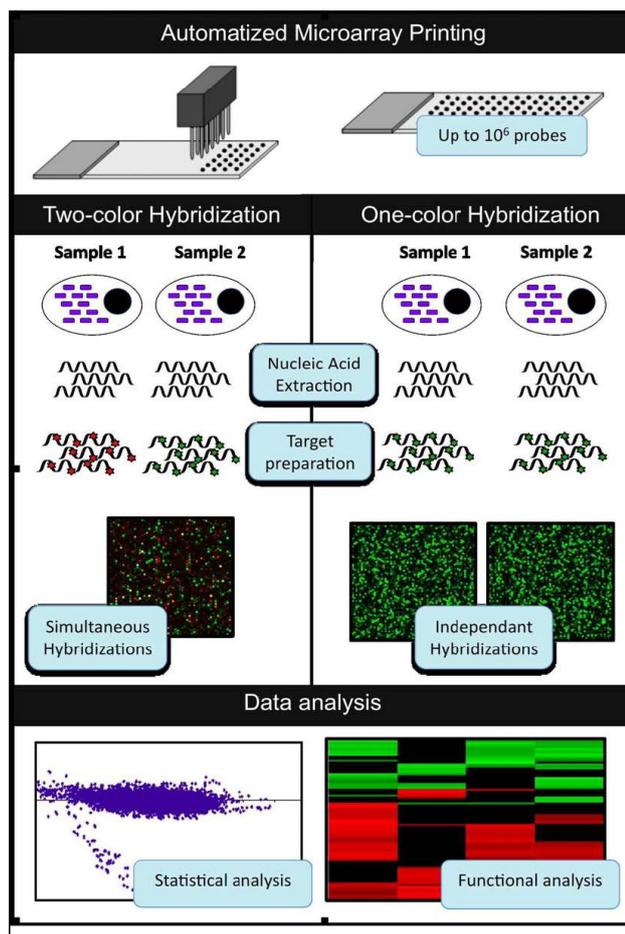


Figure 6 : Schéma représentatif des méthodes d'hybridations avec les puces à ADN ([www-lmgm.biotoul.fr](http://www-lmgm.biotoul.fr))

Ainsi la purification des ARNm bactériens et l'élimination du contaminant eucaryote sont des pré-requis pour effectuer des analyses microarrays efficaces.

Ces dernières années, la technologie des puces a été considérablement améliorée (augmentation du nombre de spots, reproductibilité...) ainsi que le nombre de génomes à disposition permettant ainsi d'étudier avec des lames beaucoup plus spécifiques les mécanismes complexes intervenant dans les interactions hôtes et bactéries (Cummings and Relman, 2000). Les méthodes permettant d'éliminer les contaminants ARN provenant de la cellule hôte ont également été développés. Tout d'abord des méthodes de lyses cellulaires sont appliquées. On utilise soit une lyse mécanique (sonication) (Nicholson *et al.*, 2003) ou une lyse chimique (guanidine thiocyanate) pour détruire les membranes cellulaires (Schnappinger *et al.*, 2003). L'une des méthodes les plus utilisées actuellement pour enlever l'ARN eucaryote est la technologie d'hybridation soustractive développée par Ambion (MicobEnrich Kit). Cette technique est basée sur l'hybridation spécifique de l'ARN 18S, 28S et des ARNm eucaryotes polyadénylés avec d'oligonucléotides couplés à des billes magnétiques (Leroy and Raoult, 2010). Cette méthode de purification a été utilisée avec succès pour purifier l'ARN bactérien des cellules eucaryotes infectées permettant d'obtenir des profils d'expressions des différentes bactéries (*R. conorii*, *Burkholderia pseudomallei*) dans des conditions d'infections *in vivo* (Tuanyok *et al.*, 2006; Renesto *et al.*, 2008). Avec la technique SCOTS qui sera détaillée dans le chapitre II et qui permet d'éliminer en plus le contaminant ribosomale procaryote, l'analyse par puce ADN offre de nouvelles perspectives dans l'étude des transcriptomes de différentes *Rickettsiales*, telle que la bactérie *E. ruminantium*.

## **5. Etude de l'expression des gènes de la famille multi-génique *map* d'*Ehrlichia ruminantium* chez l'hôte et le vecteur**

Une première étude de l'expression de gènes d'*E. ruminantium* a porté sur la famille multi-génique *map* (16 paralogues de gènes codant pour des protéines membranaires externes) qui comprend le gène codant pour la protéine membranaire MAP1 (Major antigenic protein) (Van Heerden *et al.*, 2004; Bekker *et al.*, 2005). Cette protéine est fortement immunogène et induit une réponse humorale importante (van Vliet *et al.*, 1994). Même s'il a été prouvé que MAP1 n'est pas une protéine induisant une réponse immunitaire protectrice, cette famille multi-génique pourrait avoir un intérêt dans une approche vaccinale. L'étude a consisté à étudier

l'expression différentielle des gènes du locus MAP, dans les cellules endothéliales de l'hôte ruminant (bovins) et dans différentes populations de cellules de tiques. Cette étude montre que si tous les gènes *map* sont exprimés dans les cellules endothéliales, seuls certains sont exprimés dans les cellules de tiques (Bekker *et al.*, 2005). De plus, le pool de gènes exprimés dans les cellules de tiques varie suivant l'espèce et l'origine des cellules. Le profil d'expression des gènes est cependant conservé entre les 3 souches d'*E. ruminantium* testées (Gardel, Welgevonden, Senegal).

## **6. Présentation de la Cowdriose**

### 6.1. La Cowdriose et agent pathogène

Avec les theilérioses et les trypanosomoses, la cowdriose est considérée comme l'une des plus importantes maladies infectieuses transmises par les arthropodes chez les ruminants en zone tropicale. L'agent pathogène est une bactérie intracellulaire stricte récemment reclassée sous le nom d'*E. ruminantium* (Dumler *et al.*, 2001). L'ordre et le genre auxquels appartient cette bactérie est présentée ci-dessus. Les premiers cas cliniques furent rapportés par Louis Trichardt en Afrique du sud remontent à 1838. La maladie affecte les ruminants domestiques et sauvages et est transmise par des tiques du genre *Amblyomma*. En Afrique, la productivité du bétail est sévèrement affectée par les tiques et les maladies qu'elles transmettent (Musoke *et al.*, 1997) et la cowdriose peut causer jusqu'à 90% de perte dans un troupeau, chez les races sensibles. De façon générale on considère que les races originaires des régions d'enzootie ont acquis un certain degré de résistance génétique, dont le niveau est cependant difficile à établir. Cependant, la maladie constitue un des obstacles majeurs à l'introduction en Afrique subsaharienne et aux Antilles de races améliorées à cause de la sensibilité des races exotiques en provenance d'autres continents. Sur le continent nord américain la présence de vecteur (*Amblyomma americanum*) susceptible de transmettre la maladie et la possible migration d'oiseaux porteurs de tiques infectées (au stade nymphal) représente un risque majeur d'introduction de la cowdriose. Un tel scénario aurait un impact économique désastreux dans la région avec des pertes estimées à 792 millions de dollars par an selon l'Office international des epizooties (OIE).

## 6.2. Répartition géographique

L'aire de répartition de la cowdriose s'étend à l'ensemble de l'Afrique sub-saharienne en dehors des zones désertiques et de la forêt dense humide et est corrélée à la présence de la tique vectrice dans ces régions (Figure 7A). Les premiers travaux de recherche sur la maladie (étiologie, mode de transmission, pathogénie...) furent naturellement menés en Afrique du Sud (Hutcheon, 1900; Alexander, 1931). C'est seulement à partir du début du 20<sup>ème</sup> siècle que la maladie a été décrite dans la plupart des pays de l'Afrique sub-saharienne, dans de nombreuses îles (Comores, Sao Tomé, Zanzibar, Madagascar), les îles Mascareignes (Réunion et Ile Maurice). La présence de la maladie dans les Antilles Françaises (Guadeloupe, Marie Galante) et dans l'île d'Antigua est corrélée à l'indice de présence de la tique vectrice qui a une aire de répartition plus large dans la caraïbes (Figure 7B) (Camus *et al.*, 1996).

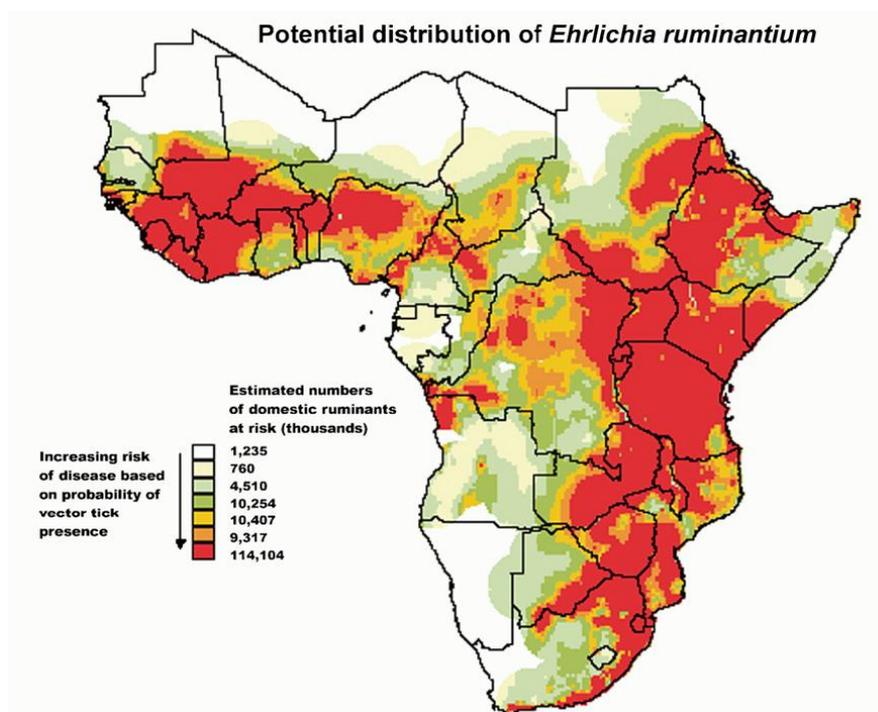


Figure 7 A: Probabilité de présence de la maladie en fonction de la présence de la tique. (Allsopp, 2010)

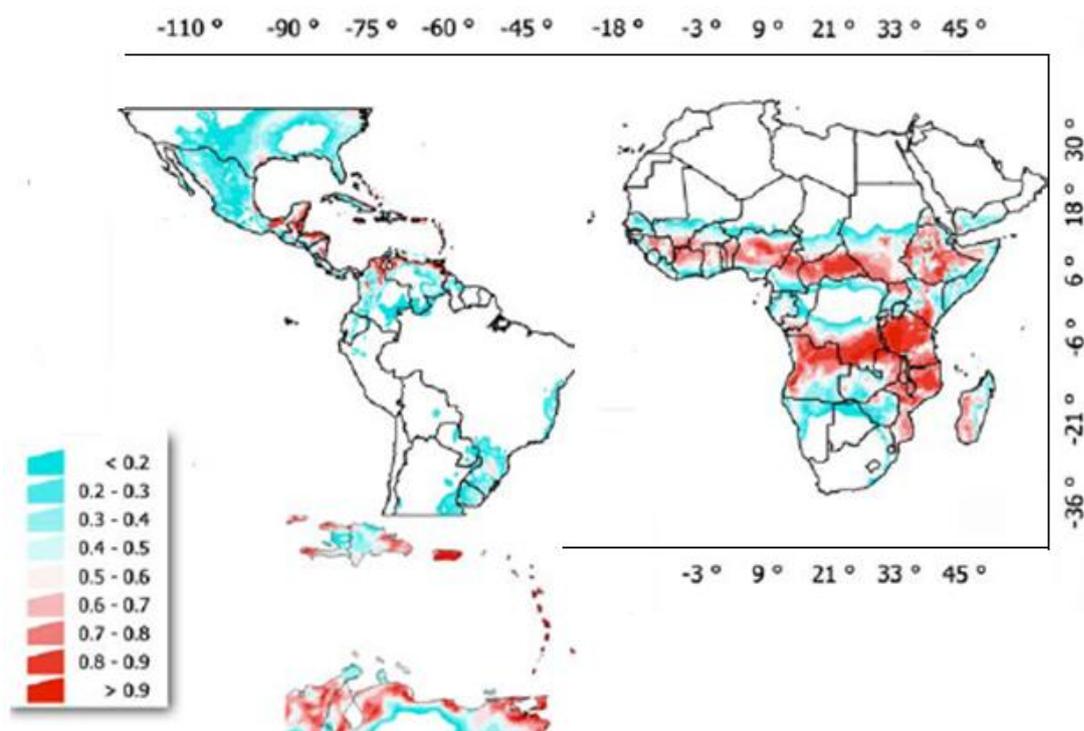


Figure 7 B: Indice de présence de la tique *A. variegatum* sur les continents américain et africain (Estrada-Pena *et al.*, 2007)

### 6.3. Les espèces affectées

La cowdriose est une maladie qui touche principalement les ruminants domestiques : bovins, caprins et ovins. Les ruminants sauvages peuvent être également infectés. Cependant, l'expression clinique est très variable selon les espèces ou les races, allant du portage asymptomatique au tableau clinique fatal (Norval *et al.*, 1990). Ainsi, parmi les 12 espèces de ruminants chez qui l'infection par *E. ruminantium* a été prouvée, huit d'entre eux font l'objet d'un portage sub-clinique (Peter *et al.*, 2002). Le principal réservoir sauvage serait constitué par les buffles (*Syncerus caffer*) qui, après infection, peuvent rester porteurs du germe durant des mois et peut-être durant toute leur vie (Euzéby and Tindall, 2001). D'autres vertébrés ont pu être infectés expérimentalement (Oberem and Bezuidenhout, 1987a). La souris à ce titre constitue un modèle expérimental.

Alors qu'un risque potentiel d'infection chez l'homme existe, *E. ruminantium* étant cultivable sur cellules endothéliales humaines (Totte *et al.*, 1993a), la cowdriose n'est toujours pas reconnue comme étant une zoonose (Kelly *et al.*, 1992). En 2005, trois cas humains d'infection par *E. ruminantium* ont été rapportés en Afrique du Sud (Allsopp *et al.*, 2005) avec détection d'*E. ruminantium* par PCR sur les prélèvements, cependant il n'a pas été prouvé que ce pathogène était la cause de la mort de ces individus.

La connaissance de la gamme d'espèces d'animaux sauvages chez lesquels la transmission d'*E. ruminantium* est possible, est un pré-requis pour mieux apprécier la dynamique de transmission du pathogène chez les animaux domestiques vivant en zone d'enzootie mais également pour évaluer le risque de transmission de la maladie chez les animaux importés des zones indemnes. Des stratégies de régulation de mouvements d'animaux transfrontaliers peuvent alors être plus facilement développées.

#### 6.4. Transmission

Seules les tiques du genre *Amblyomma* (plus de 100 espèces) sont capables de transmettre cette bactérie (Figure 8). Ainsi, en Afrique, la répartition de la cowdriose est superposable à la distribution des tiques du genre *Amblyomma*. A ce jour, seules une douzaine d'espèces seraient impliquées dans la vectorisation du pathogène (Uilenberg, 1982; Walker and Olwage, 1987), de façon naturelle ou expérimentale. Parmi celles-ci, *A. variegatum* (vecteur majeur) et *A. hebraeum* sont les deux espèces les plus importantes sur le continent africain. Dans les caraïbes, *A. variegatum*, la seule espèce présente, est responsable des cas de cowdrioses recensés. Des espèces de tiques présentes sur le continent américain ont pu être infectées expérimentalement et représentent des vecteurs potentiels de la cowdriose : *A. cajannense*, *A. maculatum*, *A. dissimile* / *A. americanum*. La présence de cowdriose dans les Antilles et la capacité des tiques américaines à transmettre la maladie représentent une menace réelle pour le continent américain (Barré *et al.*, 1987).

Les tiques du genre *Amblyomma* sont trixènes (3 hôtes), chacune des trois stases (larve, nymphe, adulte) prenant un repas sanguin sur un hôte différent. Les tiques sont infectées lors du repas sanguin sur un animal malade ou porteur asymptomatique. Elles vont ensuite se détacher de l'animal pour muer avant d'infecter un nouvel hôte. La transmission d'*E.*

*ruminantium* par les tiques survient 1 à 3 jours après le repas infectant pour les nymphes et 2 à 4 jours pour les adultes (Bezuidenhout, 1987; Camus and Barre, 1995). La transmission transovarienne est possible mais elle joue un rôle négligeable par rapport à la transmission transstadial.



Figure 8 : Photo de tiques appartenant à l'espèce *Amblyomma variegatum* (N. Vachiéry)

## 6.5. La maladie : symptômes et pathologie

### 6.5.1. Symptômes

La période d'incubation après une infection naturelle est en moyenne de deux semaines, même si elle peut aller jusqu'à un mois. Différentes formes de la maladie peuvent être observée allant de la forme suraiguë à légère (Van de Pypekamp and Prozesky, 1987) selon la race, l'état immunitaire, la sensibilité à la maladie, l'âge de l'animal, la virulence et la dose infectante de l'isolat. Les principaux signes cliniques caractéristiques d'une forme grave (aiguë) sont une hyperthermie soudaine et brutale (supérieure à 41°C), pour une durée variant de 2 à 6 jours. Cette montée en température s'accompagne généralement de divers troubles tels que l'anorexie, l'apathie, les troubles respiratoires (dyspnée) et les troubles nerveux (convulsions, hypertonie). Sans aucun traitement aux antibiotiques au début de

l'hyperthermie, l'animal meurt en développant des symptômes nerveux caractéristiques de la maladie (pédalage, perte d'équilibre). La forme suraiguë de la cowdriose est relativement rare. Elle intervient sur des animaux très réceptifs et l'animal meurt par convulsion, quelques heures après l'hyperthermie, sans avoir montré le cortège de signes cliniques précédents. Enfin les formes légères rarement mortelles existent chez certains animaux vivants en zone d'endémie ou chez les animaux jeunes.

#### *6.5.2. Pathologie : les lésions*

Les lésions post-mortem caractéristiques de la cowdriose sont généralement des œdèmes au niveau des séreuses (plèvre et péricarde) avec présence d'exsudats (Prozesky, 1987a, b). On note un hydropéricarde plus important chez les petits ruminants que chez les bovins. Cette caractéristique est à l'origine du nom anglais de la maladie: heartwater. D'autres lésions moins caractéristiques peuvent apparaître telles que splénomégalie, hémorragies dans divers organes, congestion hépatique. Enfin au niveau du cerveau, on observe généralement une congestion vasculaire accompagnée parfois d'œdèmes. Les lésions histologiques observées, œdème pulmonaire et péricardique, résultent d'une augmentation de la perméabilité vasculaire (processus inflammatoire aigu), le liquide infiltrant contenant en majorité des macrophages. Cependant il ne semble pas exister de corrélation entre la sévérité des lésions et la présence du pathogène dans les tissus infectés.

#### 6.6. Diagnostics

Les méthodes traditionnelles qui permettent le diagnostic de la cowdriose comprennent :

- la reconnaissance des signes cliniques (hyperthermie élevée, signe nerveux, détresse respiratoire et mort en 2 à 5 jours). Cependant ces symptômes peuvent être confondus avec d'autres maladies d'origine bactérienne (Uilenberg, 1983).
- un diagnostic microscopique qui se fait par recherche de colonie d'*E. ruminantium* dans les capillaires cérébraux à partir d'un frottis de cerveau coloré au GIEMSA. Il faut au préalable effectuer une biopsie au niveau du cerveau de l'animal pour récupérer un fragment de substance cérébrale.

Des méthodes moléculaires plus sensibles ont été développées basées notamment sur l'amplification par PCR de séquences spécifiques d'*E. ruminantium* (gène *map1* et fragment de gène *pCS20*), pour la détection de la bactérie (Mahan *et al.*, 1998; Martinez *et al.*, 2004; Raliniaina *et al.*, 2010). Avec ces nouveaux outils moléculaires, des études épidémiologiques ont pu être menées sur la prévalence du pathogène et la caractérisation des souches (Martinez *et al.*, 2004; Raliniaina *et al.*, 2010). Ces méthodes sont complétées par le diagnostic sérologique : indirect fluorescent antibody (IFAT) (Du Plessis and Malan, 1987), enzyme-linked immunosorbent assays (ELISA) (Sumption *et al.*, 2003).

### 6.7. L'immunité

Lorsqu'il s'agit de pathogènes intracellulaires obligatoires, la réponse cellulaire chez l'hôte est en général plus complexe et fait intervenir une réponse à médiation cellulaire et des médiateurs solubles interleukines (II) et cytokines. Au cours de leur développement intracellulaire, les pathogènes ne sont pas directement accessibles au système immunitaire. Leur présence est signalée à la surface des cellules infectées soit par présentation de peptides en association avec les molécules du CMH (complexe majeur d'histocompatibilité) ou alors par l'expression de protéines excrétées par ce dernier. L'absence d'accès direct de certains effecteurs solubles ou le respect de l'intégrité de la cellule infectée fait intervenir des cellules spécialisées capables de détruire sélectivement les cellules infectées. Ainsi, le rôle primaire de cellules cytotoxiques est de compléter la fonction des anticorps qui sont plus efficaces dans la lutte de pathogènes libres (Kaufmann, 1993).

Les connaissances que nous avons de l'immunité protectrice contre la cowdriose restent encore fragmentaires car les mécanismes mis en jeu ne sont pas bien élucidés. Cependant, les divers travaux ont démontré que l'immunité protectrice contre *E. ruminantium* est principalement à médiation cellulaire (Totte *et al.*, 1997; 1999), avec une augmentation significative des lymphocytes T CD8<sup>+</sup> et un rôle important de l'IFN $\gamma$  sécrété à la fois par les lymphocytes T CD8<sup>+</sup> et CD4<sup>+</sup>. Même si les anticorps produits sont détectables aussitôt qu'apparaissent les premiers symptômes fébriles (Viljoen *et al.*, 1987), ils ne semblent pas jouer un rôle majeur dans la protection. En effet leur taux dans le sang n'est pas corrélé à la

durée de la protection, et le transfert de sérum d'animaux immuns n'induit pas de protection chez des animaux naïfs. Expérimentalement, il a été prouvé que les interférons  $\alpha$  et  $\gamma$  ont un effet inhibiteur puissant sur la croissance d'*E. ruminantium* (Totte *et al.*, 1993b). L'effet antagoniste des Interleukines (Ils) sur l'évolution de la maladie a également été montré. Ces médiateurs chimiques notamment l'Il-1 et l'Il-6 ont un rôle de cofacteur de stimulation des lymphocytes B et T mais sont aussi des cytokines pro-inflammatoires ayant un rôle potentiel dans le processus morbide.

#### 6.8. Contrôle de la maladie

Actuellement il existe quatre grandes stratégies de contrôle de la cowdriose :

- 1) le traitement à base de tétracyclines (Van Amstel and Oberem, 1987). Le traitement n'est réellement efficace que s'il est administré tôt, au début de la phase hyperthermique avant l'apparition des signes cliniques de focalisation (nerveux et cardiaque).
- 2) le contrôle du vecteur par des acaricides, méthode potentiellement nocive pour l'environnement (selon le produit utilisé) et dont le coût élevé dans les stratégies d'éradication limite son utilisation.
- 3) en mettant à profit la résistance naturelle des jeunes animaux et des races indigènes dans les régions d'endémies.
- 4) la vaccination (Van Amstel and Oberem, 1987; Uilenberg, 1990).

Il existe différents types de vaccination mais jusqu'à présent la seule méthode commercialisée est basée sur une infection avec un inoculum infectieux (sang, broyat de tiques) et un traitement aux tétracyclines lors de l'apparition de la fièvre (Oberem and Bezuidenhout, 1987b). Depuis de nombreuses années des essais de développement de vaccin ont été réalisés. L'immunisation utilisant un vaccin inactivé à partir d'*E. ruminantium* cultivés sur des cellules endothéliales bovines a été développé par Martinez et collaborateurs (1994). Le vaccin inactivé confère une bonne protection avec 80 à 100% de protection pour des animaux immunisés et infectés avec une souche homologue et une protection plus faible lors d'une infection hétérologue (Martinez *et al.*, 1994). Le principal avantage du vaccin inactivé est la possibilité de fournir un vaccin multivalent avec un cocktail de plusieurs souches à condition

qu'elles soient isolées et cultivées *in vitro*. Il a été montré récemment qu'un vaccin incluant une souche de référence Gardel et une souche locale permettait une bonne protection de moutons sensibles à la cowdriose lors d'essais sur le terrain au Burkina Faso (Adakal *et al.*, 2010).

Le développement de la culture cellulaire a permis l'atténuation de plusieurs souches d'*E. ruminantium in vitro*, offrant ainsi de nouvelles possibilités vaccinales. Ainsi des souches comme Gardel, ou encore Sénégal ont pu être atténuées après de multiples passages sur cellules endothéliales (200 et 60 passages respectivement) (Jongejan *et al.*, 1993). La souche Welgevonden a aussi pu être atténuée en cultivant la souche dans des macrophages canins (Zweygarth *et al.*, 2005). La méthode de vaccination par inoculation de souches atténuées a montré son efficacité dans le cas d'une infection homologe (100 % de protection) à la fois en conditions contrôlées et lors d'essais sur le terrain (Faburay *et al.*, 2007). Le vaccin atténué comme le vaccin inactivé confère une protection plus faible lors d'infection avec des souches différentes des souches vaccinales. De plus, il n'existe que peu de souches atténuées et un risque de réversion de l'atténuation est possible.

En dehors des améliorations notables au niveau des processus de génération de vaccins (stabilité, conservation, coûts, accessibilité) et des bons niveaux de protection obtenus pour ces différents vaccins dans le cas de protection contre une infection homologe, la grande diversité des souches d'*E. ruminantium* dans les régions enzootiques est une problématique majeure pour l'élaboration d'un vaccin efficace. En effet, au vu de cette grande diversité il serait très difficile d'atténuer toutes les souches d'*E. ruminantium*, certaines n'étant pas atténuables par passage *in vitro*.

Dans ce contexte, l'élaboration d'un vaccin recombinant incluant des gènes ou antigènes ciblés d'*E. ruminantium* et induisant une protection pour un panel de souches semble être une alternative intéressante. Sur ce principe, deux souches d'*E. ruminantium* possédant des phénotypes différents (absence de protection croisée) ont été séquencées pour trouver en outre des gènes candidats à l'élaboration de vaccin recombinant: *E. ruminantium* Gardel (isolée en Guadeloupe à Gardel) et d'*E. ruminantium* Welgevonden (isolé en Afrique du sud). Des travaux ont été réalisés sur les protéines membranaires de surface appartenant à la famille multigénique *map1* (Major antigen protein). Ainsi un plasmide exprimant la protéine MAP1 a été réalisé mais les essais sur les ruminants n'ont pas été concluants surtout en ce qui

concerne la protection en challenge hétérologue (Nyika *et al.*, 2002; Nyika *et al.*, 1998). Récemment, trois gènes ont été sélectionnés pour leur polymorphisme de souches *CpG1*, *CpG2* et *CpG3* (Cowdria polymorphic genes) et testés au sein d'un vaccin ADN plasmidique. Aucune efficacité n'a été montrée lors de la vaccination avec le vaccin ADN. Par contre, lorsqu'une pré-immunisation avec la protéine recombinante CPG1 est effectuée avant l'immunisation plasmidique *CpG1* il y a une protection des animaux contre une infection homologe (Pretorius *et al.*, 2010). Cette protection ne garantit pas une protection efficace sur le terrain si un mélange de *CpG1* de souches différentes n'est pas testé.

## **7. Problématique et objectifs**

Jusqu'à aujourd'hui très peu d'analyses transcriptomiques ont été réalisées chez les *Rickettsiales*. Plus précisément sur notre modèle, l'étude du transcriptome d'*E. ruminantium* n'a pas encore été réalisée, seule l'expression d'un nombre limité de gènes à un stade de développement et dans un environnement donné a fait l'objet de recherche. Certains mécanismes commencent à être élucidés avec l'avènement de méthodes permettant d'étudier le transcriptome et le protéome des pathogènes mais globalement la pathogénèse et en particulier les gènes de virulence exprimés par *E. ruminantium* au cours d'une cinétique d'infection n'ont pas encore été identifiés. Avec l'amélioration de la culture d'*E. ruminantium in vitro* dans les cellules endothéliales bovines, les possibilités d'analyses pour comprendre ce qui se passe au cours du cycle complexe de la bactérie ont augmenté. De plus, la souche Gardel a pu être atténuée *in vitro* au sein de notre laboratoire nous donnant l'opportunité de comparer les niveaux d'expression entre une souche virulente et une souche atténuée au cours du temps et ainsi d'identifier les mécanismes associés à la virulence et à l'atténuation des souches.

L'annotation du génome des deux souches Gardel et Welgevonden d'*E. ruminantium* a permis d'identifier de nombreux gènes dont certains semblent être impliqués dans divers processus permettant la survie de la bactérie dans le milieu intracellulaire, ouvrant ainsi un champ d'application très large à partir des données génomiques (Frutos *et al.*, 2006a). Le fort pourcentage de gènes de fonction inconnue (~30% des CDS) dans un si petit génome (920 à 958 CDS) est révélateur de l'hyperspécialisation de la bactérie à son hôte. Il est possible que

parmi l'ensemble de ces gènes se trouve des candidats qui possèdent un rôle capital dans la virulence d'*E. ruminantium*. De plus la spécificité de l'interaction de chaque souche avec l'hôte eucaryote semble être un élément important à prendre réellement en compte puisque l'annotation du génome de Gardel et Welgevonden a révélé de nombreux gènes uniques pour chaque souche (57 CDS) mais aussi des gènes tronqués permettant de les différencier entre elles.

Dans ce contexte, le but de cette thèse est d'étudier pour la première fois le transcriptome d'*E. ruminantium*. L'objectif principal est de comparer la souche Gardel virulente et la souche atténuée à différents temps de l'infection afin d'identifier à tous les stades de développement les gènes différentiellement exprimés. La comparaison des modifications d'expressions des gènes à l'intérieur des cellules endothéliales bovines entre les souches atténuées et virulentes, d'un stade de développement à un autre, permettra de mieux comprendre la pathogenèse d'*E. ruminantium* en identifiant les gènes cruciaux impliqués dans la virulence et dans l'interaction hôte-pathogène.

Les gènes dont le niveau d'expression est altéré chez la souche atténuée par rapport à la souche virulente feront l'objet d'études approfondies en génomique comparative puisque nous disposons également de la séquence du génome de la souche Gardel atténuée. Des mutations sur la souche atténuée par rapport à la souche virulente pourraient confirmer le rôle dans la pathogenèse de gènes identifiés par l'étude transcriptomique.

Dans un premier temps la méthode SCOTS (Selective Capture Of Transcribed Sequences) est utilisée sur les échantillons biologiques générés *in vitro* de manière à éliminer le contaminant ARN eucaryote de la cellule hôte, l'ARNr procaryote et d'enrichir spécifiquement en ARN messenger bactérien (Daigle *et al.*, 2001). Ensuite, les échantillons générés par SCOTS sont hybridés sur des puces à ADN pangénomiques contenant des sondes oligonucléotidiques spécifiques de la bactérie *E. ruminantium*, permettant ainsi de détecter l'expression de chacun des gènes de Gardel. Le différentiel d'expression des gènes est ensuite validé par PCR en temps réel sur un nombre restreint de gènes.

L'objectif à long terme est d'apporter les éléments pouvant conduire à un meilleur contrôle de la maladie par l'élaboration de vaccin de deuxième génération : vaccin recombinant ou atténué. La connaissance des gènes impliqués dans la pathogenèse et le développement devrait permettre d'élaborer un vaccin recombinant contenant des antigènes d'*E.*

*ruminantium* important au développement de la bactérie ou d'atténuer par ingénierie génétique n'importe quelle souche d'*E. ruminantium*. Ce vaccin de nouvelle génération semble être d'autant plus nécessaire que la bactérie présente une forte diversité génétique, mais aussi phénotypique en fonction des souches. Pour contourner ce problème de diversité l'utilisation d'un vaccin multivalent comprenant plusieurs gènes ou antigènes est une alternative envisageable.

## **CHAPITRE II.**

**Protocole expérimental et stratégie d'étude : Etapes préalables nécessaires à l'analyse du transcriptome d'*Ehrlichia ruminantium***

## **1. Préparation du matériel biologique : Les souches *Ehrlichia ruminantium* Gardel virulente et atténuée en culture**

### 1.1. Contexte

*E. ruminantium* est une bactérie intracellulaire obligatoire à cycle complexe. La difficulté technique majeure pour l'analyse du transcriptome d'*E. ruminantium* comme pour toutes les *Rickettsiales*, est liée à ce caractère intracellulaire obligatoire qui pose deux problèmes : la quantité limitante du matériel biologique et la pureté de ce matériel biologique. La culture de la bactérie *in vitro* dans les cellules endothéliales d'aortes de bovin (BAE : Bovine Aorta endothelial cells) permet d'obtenir différents stades évolutifs d'*E. ruminantium* en quantité suffisante. De ces échantillons biologiques sont extraits d'une part de l'ADN bactérien et d'autre part de l'ARN total incluant l'ARN bactérien et bovin. Avant d'effectuer la validation des puces à ADN d'*E. ruminantium* et l'analyse transcriptomique, plusieurs étapes indispensables ont dû être réalisées au préalable: production d'échantillons biologiques, synchronisation des cultures et détermination des croissances bactériennes et des points de collectes à analyser, adaptation de la méthode de capture sélective des transcrits au modèle *E. ruminantium*. Une fois les résultats d'hybridation sur les puces à ADN obtenus, la validation par PCR quantitative en temps réel (qRT-PCR) est obligatoire. Pour cela, le choix d'un ou plusieurs gènes constitutifs stables permettant la normalisation doit être également fait.

### 1.2. Matériels et Méthodes

#### *1.2.1. Production des échantillons biologiques et obtentions des ADN bactériens et ARN totaux*

L'ARNm d'*E. ruminantium* a été produit à partir de cultures des souches Gardel virulente et atténuée sur les BAE. Pour augmenter le rendement dans la récolte d'ARNm, les cultures ont été réalisées dans des boîtes de culture Falcon 25 cm<sup>2</sup> (2 TC 25cm<sup>2</sup> par point de collecte). *E. ruminantium* est une bactérie extrêmement labile et donc son développement dépend très

fortement de l'état physiologique de la cellule hôte. Or, les cellules endothéliales sont très versatiles et le moindre changement de conditions environnementales perturbe leur croissance. Ainsi pour la réalisation de ce travail de thèse, plusieurs cinétiques ont été réalisées à différents passages pour la souche Gardel virulente (ERGp29, p36 et p39) et Gardel atténuée (ERGp243, 232, 229) de manière avoir des répliques biologiques (représentatif des variations inter expériences), dans des conditions de cultures identiques, mais avec un intervalle de temps dans la production. Le milieu de culture (Glasgow minimal essential medium) contenant 1% de L-glutamine et de pénicilline-streptomycine et de *Tryptose phosphate broth (TPB)* est enrichi en sérum de veau fœtal (10%) et les cellules sont incubées à 37 °C dans une atmosphère contenant 5% de CO<sub>2</sub>. L'un des répliques est utilisé pour l'hybridation sur microarrays après capture par la méthode SCOTS ainsi que pour la validation du différentiel d'expression de certains gènes sans qu'il y ait capture de transcrits par qRT-PCR. Le deuxième réplique est utilisé uniquement pour la validation par qRT-PCR (Figure 9).

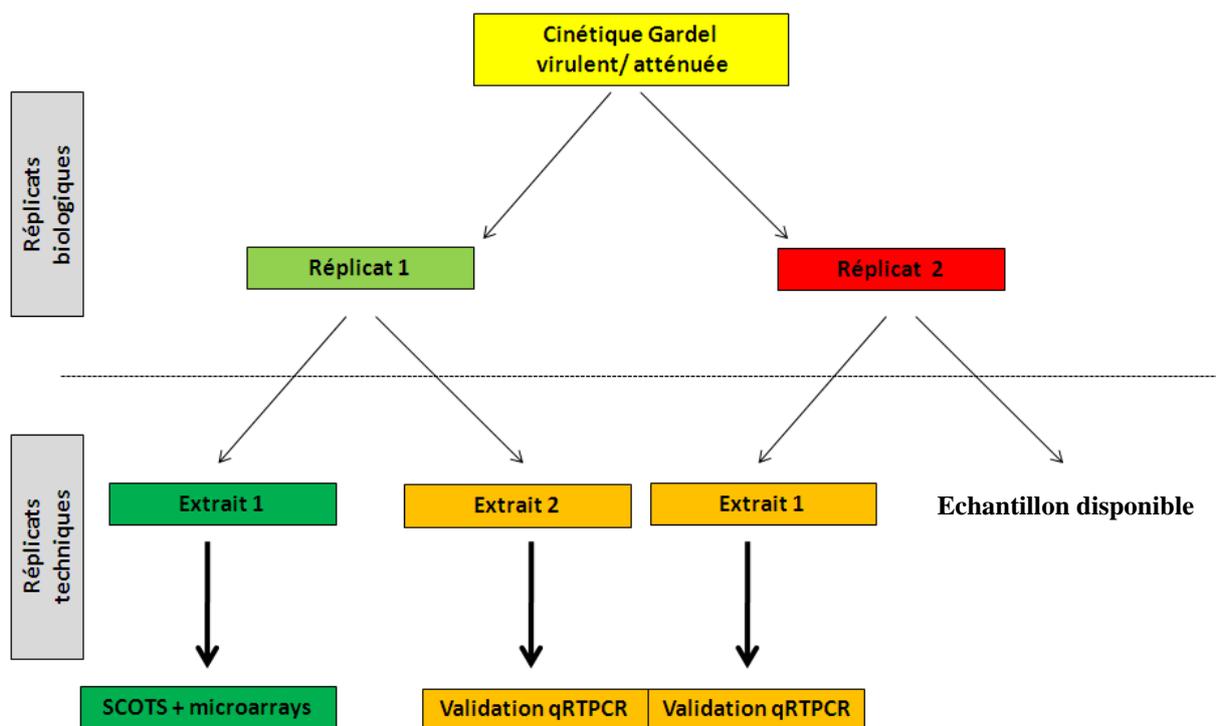


Figure 9 : Répartition des répliques pour l'analyse du transcriptome d'*E. ruminantium* utilisation des méthodes microarrays et qRT-PCR

*1.2.2. Synchronisation de l'infection et protocole de collecte des ARNs*

La première étape a consisté à synchroniser les cultures des souches Gardel virulente et atténuée afin d'obtenir les ARN messagers correspondant à un même stade de développement incluant les stades de corps réticulés, corps intermédiaires et corps élémentaires. Pour Gardel virulente, 5 points de collecte des échantillons d'*E. ruminantium* (lyse complète en 5 jours) ont été réalisés afin de récupérer tous les stades de développement (récolte toutes les 24 heures post infection, hpi) ; soit 24 hpi, 48 hpi, 72 hpi, 96 hpi et 120 hpi. Le même procédé a été appliqué à la souche Gardel atténuée mais seulement 4 points de collecte (24 hpi, 48 hpi, 72 hpi et 96 hpi) ont été réalisés, cette souche lysant de façon complète et synchrone à 96 hpi. Pour l'ensemble des points de collecte, dix pour cent du surnageant contenant les cellules infectées est récolté et conservé en culot sec pour extraction d'ADN permettant ainsi une estimation du nombre de copie d'*E. ruminantium* dans chaque échantillon par PCR quantitative en temps réel (qPCR).

Après 24 h, 48 h et 72 h de cultures post-infection par les souches bactériennes virulentes ou atténuées, les cellules endothéliales sont encore indemnes et les bactéries sont toujours à l'intérieur de la cellule hôte. Les cellules infectées sont trypsinées et rapidement lavées avec du milieu de culture d'infection pour arrêter l'action protéasique de la trypsine. Le surnageant contenant les cellules infectées est centrifugé à 1700 g pendant 5 min à 4°C et le culot cellulaire est rapidement lysé dans du TRIzol (Invitrogen, 2,5 ml pour une Tc 25 cm<sup>2</sup>) et conservé à -80°C pour extraire l'ARN total, qui servira pour l'analyse transcriptomique (Figure 10). A 96 hpi pour la souche virulente, lorsque les plages de lyse sont visibles, le surnageant est collecté et ultracentrifugé avec le tapis cellulaire préalablement trypsiné.

Après 96 hpi ou 120 hpi lorsqu'il y a 80% de lyse, on collecte le surnageant contenant les corps élémentaires et si nécessaire les cellules adhérentes sont grattées. Le surnageant contenant cellules infectées et corps élémentaires est ultracentrifugé pendant 15 min à 20000g, à une température de 4°C. L'ensemble du culot cellulaire est lysé à l'aide du TRIzol et les échantillons sont conservés à -80°C.

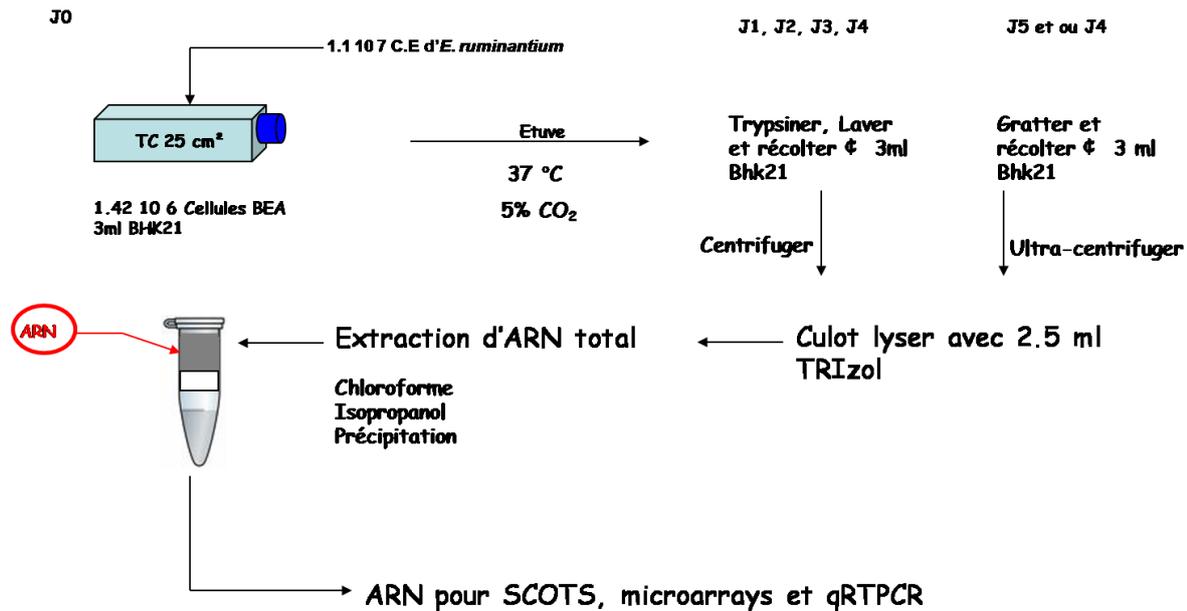


Figure 10: Protocole expérimental mis en place pour l'extraction de l'ARN d'*E. ruminantium*

### 1.2.3. Viabilité et quantification bactérienne par qPCR

Pour chaque infection de cellules par *E. ruminantium*, le pourcentage de viabilité de l'inoculum est mesuré en utilisant un kit de viabilité bactérien (live/dead BacLight Bacterial Viability Kit (Invitrogen, France)) (Vachier *et al.*, 2006). Le pourcentage de bactéries vivantes est estimé par cytométrie en flux (Figure 11 A et B) tandis que la concentration totale en bactéries se fait sur lame de comptage au microscope à fluorescence. Puis un inoculum calibré est utilisé pour infecter les cellules ( $1.1 \times 10^7$  corps élémentaires vivants pour  $1.42 \times 10^6$  cellules).

A chaque point de collecte, on effectue une quantification de la bactérie à partir de l'ADNg extrait par qPCR utilisant une sonde spécifique dirigée contre le gène *map1* d'*E. ruminantium*, dont une seule copie est présente sur le génome. La quantification est réalisée dans un volume réactionnel de 25  $\mu$ L. Le mix est composé du Taqman universal PCR master mix (*Applied Biosystems*), de 900 nM d'amorces sens et anti-sens, de 250 nM de sonde et d'H<sub>2</sub>O avec comme matrice 4  $\mu$ L de l'ADNg extrait. Un plasmide contenant la cible *map1* est utilisé pour la réalisation de la gamme étalon obtenue à partir de dilutions successives. Un témoin négatif constitué d'eau est inclus pour chaque essai. Les conditions d'amplification dans le thermocycleur (ABI-Prism 7000, *Applied Biosystems*) sont : 2' à 50°C, 10' à 95°C, 40

cycles comprenant 15'' à 95°C, 1' à 60°C. La lecture de la fluorescence se fait à chaque cycle lors de la deuxième étape.

Les amorces utilisées ont les séquences suivantes:

- map1gard-ex3F : 5' CACTTGAAGGAATGCCAGTTTCTC 3'

- map1gard-ex3R : 5' CTTAGGATTTGTAGCATTGATTACTGACACT 3'

La sonde *map1* spécifique d'*E. ruminantium* possède la séquence suivante :

- map1gard-ex3m2 : 5' 6-FAM-ATGCCTGCACACACATAT-MGB

Un réplicat technique est effectué sur tous les échantillons quantifiés.

#### 1.2.4. Colorations et observations

Pour vérifier la synchronisation du développement d'*E. ruminantium* et l'absence de contamination, on effectue une coloration RAL 555 (fixation Ethanol, Eosine et Bleu de méthylène). Une observation (avant et après coloration) par microscopie inversée des stades de développement de la bactérie est effectuée en parallèle.

### 1.3. Résultats

#### 1.3.1. Viabilité et quantification bactérienne par qPCR

Pour chaque infection avec une souche Gardel virulente ou atténuée, la viabilité estimée de l'inoculum était toujours supérieure à 50% (entre 52 et 80%) (Figure 11).

Pour les souches virulentes, les cinétiques obtenues ont un profil classique des cinétiques bactériennes avec une phase de latence entre 24 et 48 hpi, une phase de croissance exponentielle entre 48 et 96 hpi et le plateau qui est atteint entre 96 hpi et 120 hpi. Pour les souches atténuées, la phase de latence n'est pas observée ; en effet à 24 hpi on se situe au début de la phase exponentielle de croissance (Figures 12 A, B, C, D). Les résultats de quantification des échantillons montrent une très bonne reproductibilité des expériences avec des profils similaires entre les réplicats biologiques à la fois pour la souche virulente et pour la souche atténuée. (Figure 12 A, B, C, D). Toutefois on observe une différence de quantité d'environ 1 log<sub>10</sub> entre chaque réplicat biologique pour les souches virulentes et atténuées.

Ces différences sont sûrement dépendantes de l'état physiologique des bactéries lors de l'infection de départ.

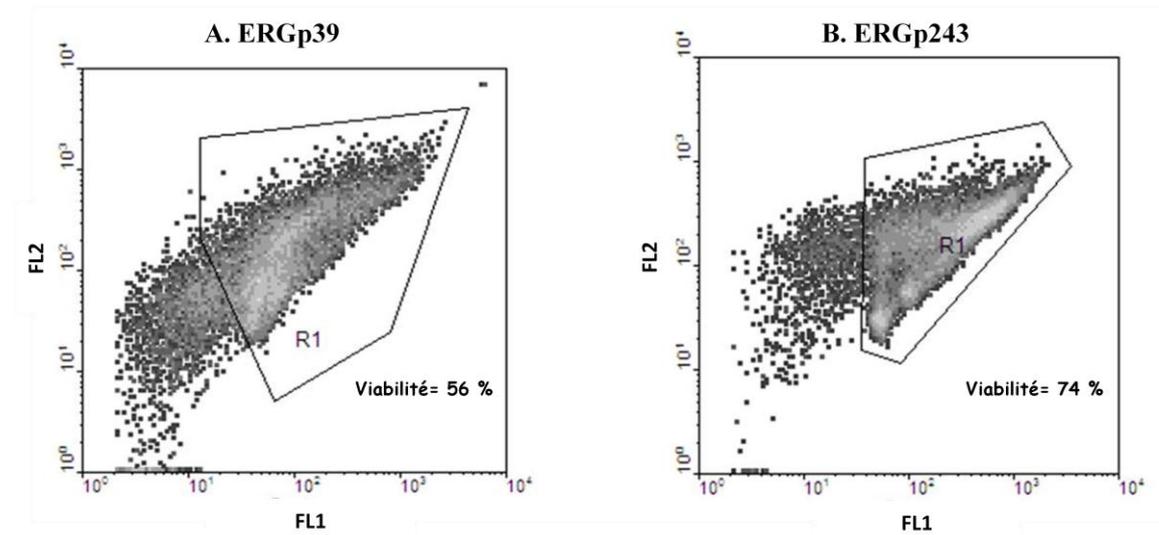


Figure 11. Estimation de la viabilité d'*E. ruminantium* ERGp39 (A, virulente) et ERGp243 (B, atténuée) par la cytométrie en flux

Virulentes

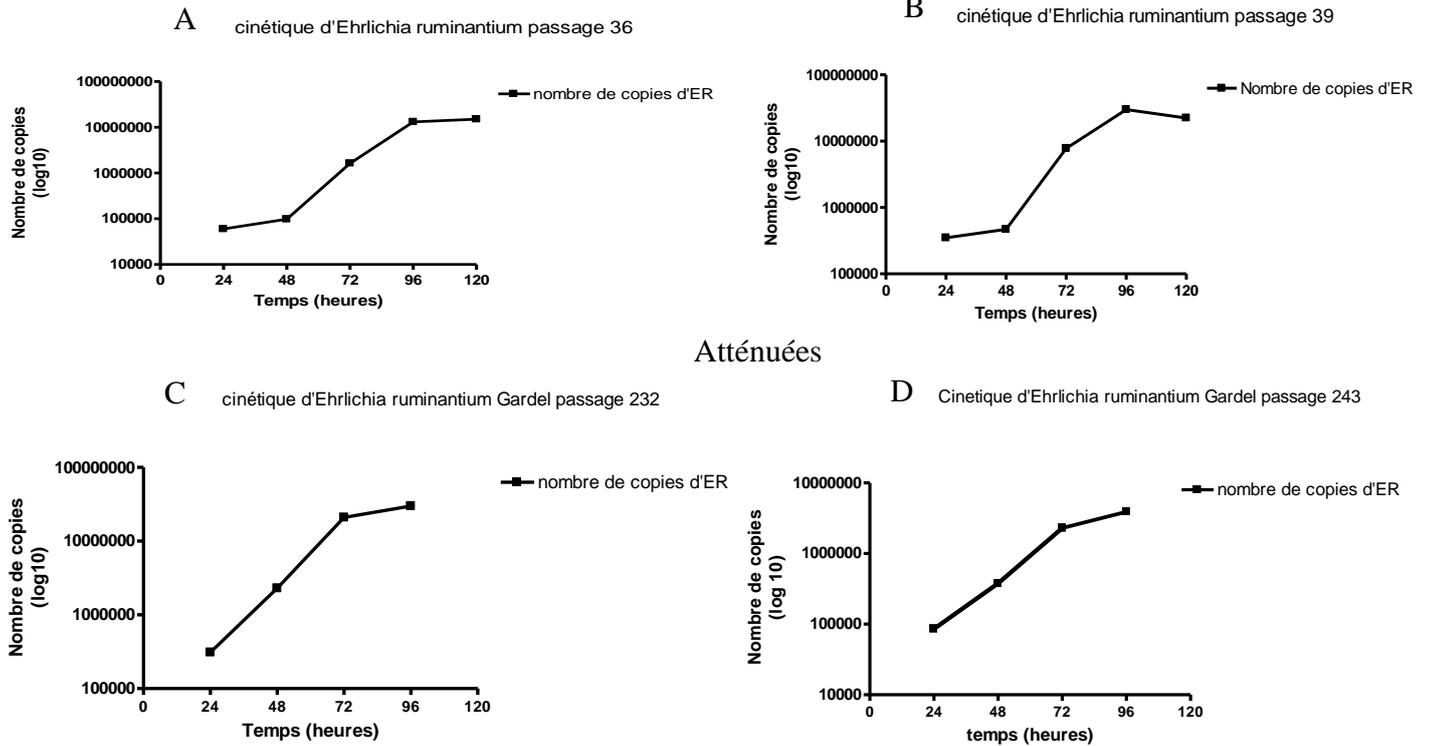


Figure 12 : Quantification absolue par qPCR à l'aide d'une sonde taqman spécifique du gène *map1*

1.3.2. Colorations et observations par microscopie

Les colorations ont permis de suivre l'évolution de l'infection en temps réel. Différentes phases du cycle de développement ont pu être observées notamment l'apparition des *morulas* et la lyse cellulaire. Les observations réalisées sont les suivantes :

- Généralement à 24 hpi les BAE sont bien infectées ; elles sont très gonflées et condensées, il n'y a pas d'effet cytotoxique (Figure 13 A).
- A 48 hpi on observe l'apparition de condensations en forme de strie sur l'ensemble du tapis cellulaire avec parfois l'observation de *morulas* pour la souche atténuée (Figure 13 B).
- A 72 hpi la souche Gardel atténuée présente des *morulas* plus grosses en grande quantité et on commence à détecter les *morulas* chez Gardel virulent (Figure 13 C).

- A 96 hpi de grosses *morulas* sont visibles avant et après coloration chez Gardel virulent et il y a très peu de lyses (<10% du tapis cellulaire). Cependant les cellules sont extrêmement condensées (90% du tapis cellulaire). Pour la souche atténuée, les cellules ont complètement lysées à 96 hpi, avec une lyse proche de 100% du tapis cellulaire pour toutes les boites de culture (synchronisation sur toutes les boites). La coloration permet de mettre en évidence de nombreuses *morulas* par cellule et on observe dans le milieu de culture des corps élémentaires au niveau des points d'éclatement des *morulas*.
- A 120 hpi une lyse supérieure à 80% est observé pour Gardel virulente sur l'ensemble des boites infectées (Figure 13 D).

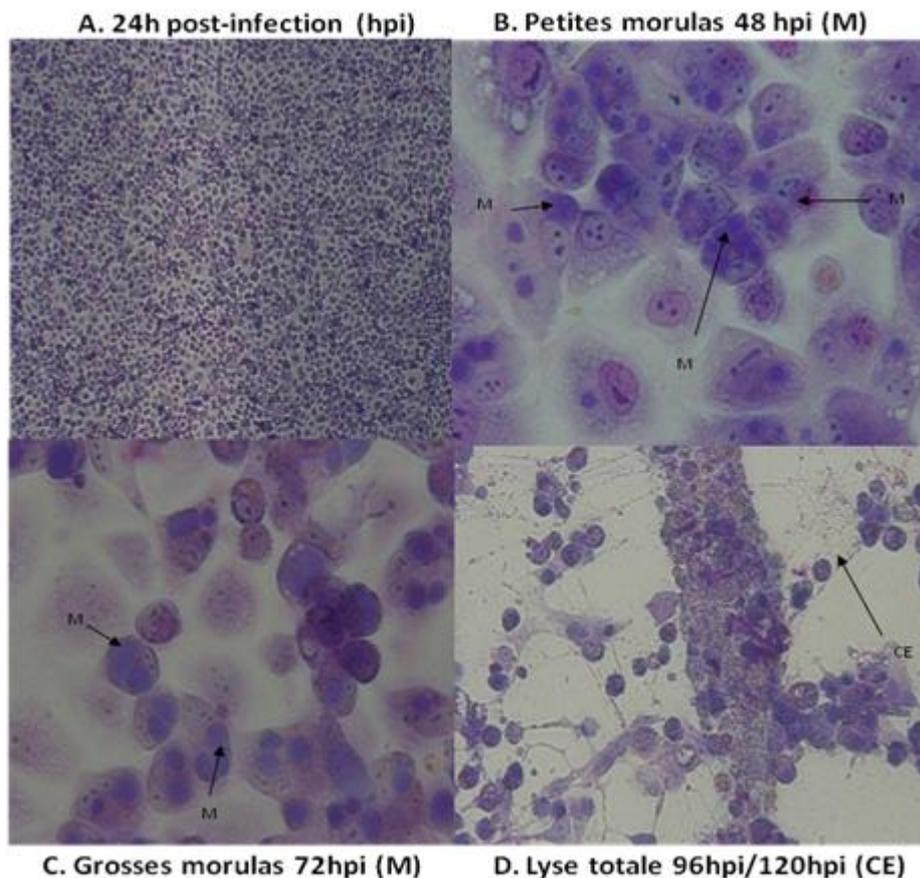


Figure 13 : Observation de l'évolution du cycle infectieux des souches d'*E. ruminantium* par coloration à différentes heures post infection (hpi) (RAL 555, L. Emboulé)  
M= *morulas* ; CE= corps élémentaires

#### 1.4. Conclusion

Cette partie présente le protocole expérimental mis en place afin d'obtenir le matériel biologique aux différentes phases de développement d'*E. ruminantium* nécessaire à l'étude du transcriptome. Plusieurs précautions ont été prises pour vérifier la synchronisation mais aussi la reproductibilité des cinétiques. La calibration de l'inoculum (viabilité et quantification) permet d'obtenir une dose infectieuse identique quelque soit la souche utilisée. Le suivi de l'évolution de l'infection par la coloration permet d'observer différentes phases dans l'évolution du cycle complexe de la bactérie. Ainsi, par ces différentes étapes dans la production du matériel nous nous assurons que nous récoltons plusieurs stades évolutifs de la bactérie, particulièrement les morulas et les corps élémentaires qui sont récoltés à partir des 24 h précédant la lyse cellulaire. Lors de la lyse totale, nous récoltons majoritairement des corps élémentaires (à 120 h et 96 h) pour les deux souches qui correspondent aussi au statut de l'inoculum infectieux T0. De la même manière à 48 hpi, nous récoltons majoritairement des corps réticulés. Pour les temps intermédiaires, il est possible qu'il y ait un mélange des différents stades de développement de la bactérie dans les échantillons collectés. Par conséquent 3 stades concordants peuvent être comparés entre les souches virulentes et atténuées. Les stades correspondant (i) aux 24 hpi suivant l'infection (24 hpi), (ii) le stade évolutifs dans lequel on retrouve un grand nombre de *morula* (48 hpi (atténuée) vs 72 hpi (virulente) dans la phase exponentielle de croissance et (iii) la phase de lyse cellulaire (96 hpi vs 120 hpi). Les échantillons générés selon le même procédé doivent ensuite servir pour réaliser l'analyse du transcriptome par l'intermédiaire de la technique SCOTS, pour une partie et à la validation des résultats obtenus par qRT-PCR, pour une autre partie.

## **2. Vérification expérimentale de la virulence et de l'atténuation des souches Gardel *in vivo***

### 2.1. Contexte

Notre étude repose sur la comparaison entre les souches Gardel virulente et atténuée au cours du cycle de développement. Nous avons vu qu'une lyse des cellules était observée *in vitro* aussi bien dans la souche atténuée que dans la virulente. Même si des expériences précédentes

au sein du laboratoire ont montré l'innocuité de la souche Gardel atténuée, ces résultats n'ont pas été publiés. Par conséquent, la vérification *in vivo* de l'atténuation de la souche Gardel est nécessaire pour valider notre modèle expérimental. Les précédents essais vaccinaux menés par Vachiéry et collaborateurs ont permis d'estimer la dose létale quantifiée *in vitro* pour la souche Gardel virulente : elle est de 30000 et 90000 corps élémentaires vivants. La virulence de la souche Gardel a été vérifiée de nombreuses fois car elle est systématiquement utilisée lors des essais vaccinaux réalisés dans le laboratoire en conditions contrôlées.

## 2.2. Matériels et méthodes

Le challenge des chèvres avec les deux souches de Gardel (virulente et atténuée) a été réalisé. Pour cela une chèvre (n°0212) naïve a été infectée avec une dose létale de 90000 CE vivants de Gardel virulente passage 27. Deux autres chèvres ont été infectées simultanément avec une dose létale (90000 CE) (chèvre n° 0614) et une dose correspondant à plus de 10 fois la dose létale minimum (900000 CE) (chèvre n° 0541) avec la souche vivante Gardel atténuée passage 230. Pendant un mois post-infection, les signes cliniques ont été observés et pondérés par des indices correspondant aux critères définis dans le tableau 2.

Tableau 2: Evaluation pathologique avec indices en fonction de l'intensité des signes cliniques observés chez une chèvre naïve ou vaccinée infectée avec une dose d'*E. ruminantium* vivante

Période d'incubation (jours) <sup>(a)</sup>	Intensité de la fièvre (°C)	Signes nerveux	Autres signes	Scores cliniques
PI>13	40 > T ≥ 39.5	tremble	Couché/ Anorexie	1
9 < PI < 11	41 > T ≥ 40	titube		2
PI < 9	T ≥ 41	pédale		3
			Mort/ Euthanasie <sup>(b)</sup>	9

a PI: période d'incubation: nombre de jours entre l'infection et l'hyperthermie (T ≥ 39.5°C)

T: température du corps journalière

b les animaux sont euthanasiés lorsque les scores cliniques prédisent une mort certaine

### 2.3. Résultats

Les résultats obtenus sont les suivants :

- Les deux chèvres infectées avec Gardel atténuée passage 230 avec les deux doses infectieuses ( $9 \cdot 10^4$  et  $9 \cdot 10^5$  CE vivants) montrent une hyperthermie ( $T > 39.5^\circ\text{C}$ ) à j15 et j12 pour une durée de 5 jours et 8 jours (Tableau 3). La chèvre infectée avec la plus faible dose est légèrement ébouriffée avec une perte d'appétit pendant 4 jours. L'autre chèvre a une perte d'appétit pendant deux jours. Les symptômes disparaissent après le 19<sup>ème</sup> jour post infection et les animaux retrouvent un état physiologique normal. Les signes cliniques sont très modérés (Figure 14) et les chèvres survivent quelque soit la dose infectieuse inoculée.
- Au contraire, la chèvre infectée avec la souche virulente Gardel passage 27 présente une soudaine apparition et accumulation de symptômes (manque d'appétit, difficulté à rester debout) avec une montée en température avec une hyperthermie brutale. Même avec une dose 10 fois supérieure à la dose létale minimum, l'animal survit après infection avec la souche atténuée, ce qui prouve la réelle atténuation de la souche Gardel. La chèvre infectée avec la souche virulente a été traitée immédiatement après l'apparition des signes pathognomonique car plusieurs autres expériences d'infection avec une dose similaire de la souche Gardel à des passages précoces (entre 20 et 30) sur des animaux naïfs ont montré que cette souche tuait 100% des animaux.

### 2.4. Conclusion

Ces résultats permettent de valider l'atténuation et la virulence de souches à l'étude et montrent ainsi la pertinence du choix des souches expérimentales.

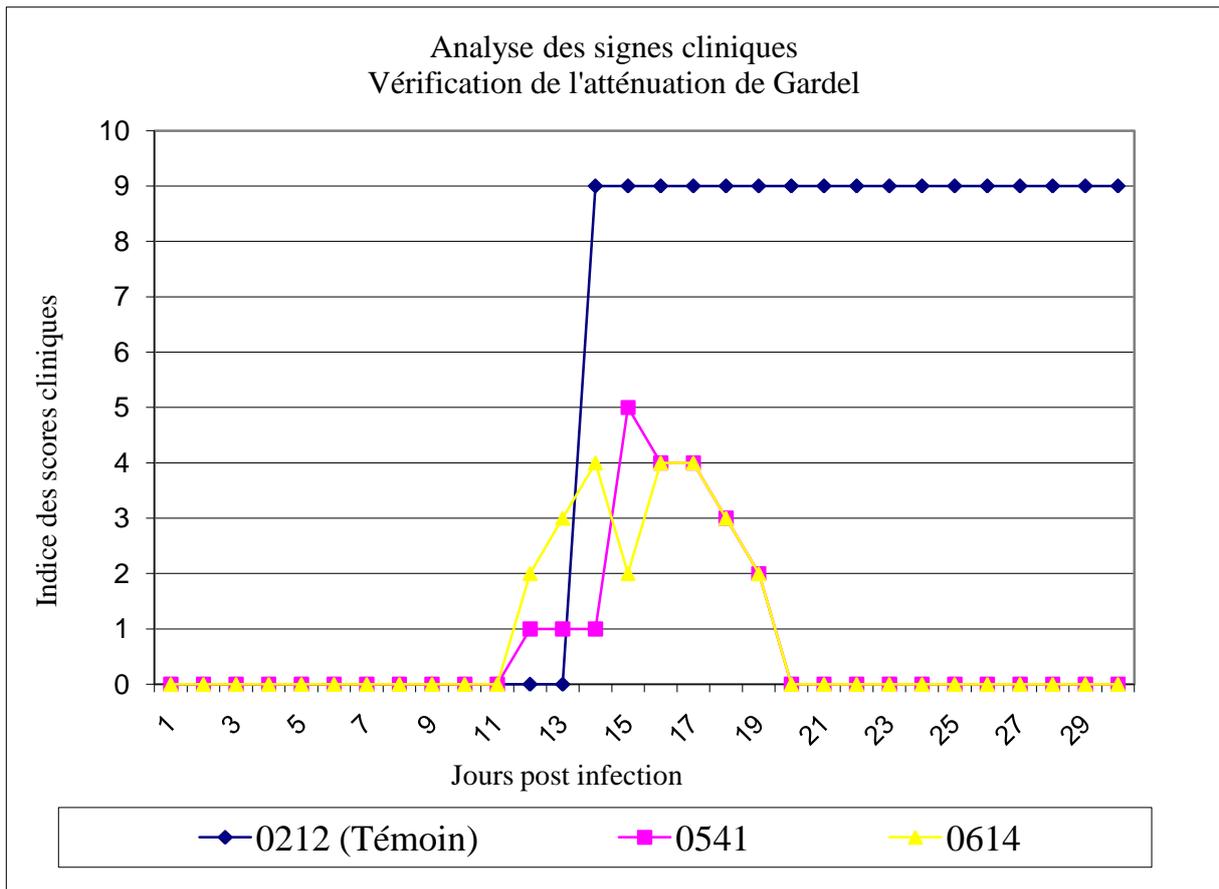


Figure 14 : Observation de l'évolution des symptômes au cours de l'infection

Tableau 3 : Suivi de l'hyperthermie et de la période d'incubation

N° chèvre	Inoculum	Dose	T°moy (°C)	Tmax (°C)	Durée incubation (j)
0541	Gardel p230	9 10 4	40.8	41.3	5
0614		9 10 5	40.7	41.2	8
0212*	Gardel p27	9 10 4	40	40	1

\*Chèvre témoin traitée aux oxytétracyclines à j14

### **3. Evaluation de la stabilité des gènes constitutifs d'*Ehrlichia ruminantium* : Recherche de gènes normalisateurs pour la validation par PCR en temps réel des gènes d'intérêt détectés en microarrays**

#### 3.1. Contexte

L'analyse du transcriptome se fait par l'hybridation des échantillons SCOTS en haut débit en utilisant des puces ADNs, puis par validation des résultats par qRT-PCR sur un nombre limité de gènes d'intérêt. Cette partie du travail consiste en l'identification d'un ou plusieurs gène(s) de ménage d'*E. ruminantium* dont l'expression ne varie pas au cours du cycle d'infection *in vitro*. Ce gène sera par la suite utilisé comme normalisateur lors de la validation par qRT-PCR des résultats obtenus sur puces ADNs, pour des mesures d'expression relative. Une étude bibliographique a permis de sélectionner des gènes de ménage préférentiellement utilisés comme gènes normalisateurs dans l'étude de l'expression de gènes dans d'autres modèles bactériens. Il s'agit des gènes *recA*, *ffh*, *proC*, *rpoD* et *16S* (Tableau 4). De plus deux autres gènes d'*ER* dont la stabilité de l'expression a été évaluée au cours du cycle de développement ont été inclus dans cette étude: les gènes *pCS20* et *map1*. La stabilité des gènes est mesurée en quantifiant les transcrits par rapport au nombre de bactéries présentes dans chaque échantillon.

#### 3.2. Matériels et méthodes

##### 3.2.1. Les gènes sélectionnés

La présence de ces gènes a été vérifiée sur le génome de Gardel et Welgevonden en utilisant le site NCBI (<http://www.ncbi.nlm.nih.gov>). Les 5 gènes choisis ont été retrouvés dans le génome de Gardel ainsi que celui de Welgevonden, en inscrivant le nom du gène recherché dans la zone, « Search gene, GeneID or locus\_tag ». Parmi les 5 gènes initialement sélectionnés, le gène *proC* code pour la pyrroline-5-carboxylate réductase, *rpoD* code pour le facteur sigma de l'ARN polymérase, *ffh* pour une protéine SRP (particule de reconnaissance du signal N-terminal d'adressage dans la membrane plasmique), *recA* pour la recombinaison A et *16S* correspond à l'ARN ribosomal *16S*. Les gènes *map1* et *pCS20* sont généralement

utilisés pour la quantification du nombre d'*E. ruminantium* dans des échantillons biologiques. Le gène *map1* qui code pour une protéine de la membrane externe très polymorphe entre souches, est fortement exprimé tandis ce que *pCS20* qui est très conservé entre souches et est faiblement exprimé.

Tableau 4 : Gènes constitutifs utilisés comme normalisateurs dans d'autres modèles bactériens

Gène	Publications	Séquence des amorces	Bactérie
<i>ffh</i>	(Takle <i>et al.</i> , 2007)	F 5' ATGGGCGATGTGCTTTCAC 3' R 5' TCAAACCCATCGCCTTCTT 3'	<i>Pectobacterium astrosepticum</i>
<i>proC</i>	(Takle <i>et al.</i> , 2007) (Savli <i>et al.</i> , 2003)	F 5' CACAGCTGATGCAGAGCGTC 3' R 5' GAAGAAATAGGCCGGTGCG 3' F 5' CAGGCCGGGCAGTTGCTGTC 3' R 5' GGTCAGGCGCGAGGCTGTCT 3'	<i>Pseudomonas aeruginosa</i>
<i>recA</i>	(Takle <i>et al.</i> , 2007)	F 5' GGTGAGCTGGTTGATCTGGG 3' R 5' GCATTCGCTTTACCCTGACC 3'	<i>Pectobacterium astrosepticum</i>
<i>rpoD</i>	(Savli <i>et al.</i> , 2003)	F 5' GGGCGAAGAAGGAAATGGTC 3' R 5' CAGGTGGCGTAGGTGGAGAA 3'	<i>Pseudomonas aeruginosa</i>

### 3.2.2. Dessin des amorces

Les amorces dérivées pour d'autres modèles n'hybrident pas entièrement sur *E. ruminantium* Gardel et Welgevonden. Des amorces spécifiques de notre modèle ont été dessinées. Pour chaque gène, des amorces donnant un produit d'amplification de ~200 paires de bases (pb), et répondant aux critères d'une qRTPCR( Th de 60°C), ont donc été dérivées à l'aide du logiciel *Primer Express Software v2.0* (Applied Biosystems), en s'assurant d'être dans une zone la plus conservée possible pour les souches Gardel et Welgevonden. Les conditions optimales de dessin des amorces ont été déterminées. Étant donné le faible pourcentage de bases G-C d'*E. ruminantium*, les amorces ont été dérivées dans une zone ayant une forte proportion de G-C. Le nombre de structures en épingles à cheveux (hairpins loops) et de dimères d'amorces a été le plus faible possible. La différence des températures de dissociation de 2 amorces était inférieure ou égale à 3°C et la longueur optimale de l'amorce est de 20 nucléotides. Une ou 2 bases G-C en 5' permet de stabiliser l'hybridation et pas plus d'une base G-C en 3' permet de

minimiser les faux appariements. Pour chaque gène, 2 couples d'amorces ont été dérivés (F1/R1 et F2/R2) et testés. Cela n'a pas été possible pour le gène *proC* car les zones correspondantes aux critères énoncés ci-dessus présentent un polymorphisme (souvent sous forme de SNP- single nucleotide polymorphism) trop important entre la souche Gardel et Welgevonden. Il a été vérifié sur la base de données du NCBI que chaque gène cible n'existe qu'en une seule copie et que les amorces dessinées ne s'hybrident pas à un autre endroit du génome. Pour les gènes *map1* et *pCS20* la quantification a été réalisée à l'aide d'une sonde taqman, spécifique de ces deux séquences chez *E. ruminantium* (Partie I section 1.2.3).

### *3.2.3. Extraction d'ARN puis conversion en ADN-complémentaire (cDNA)*

L'analyse d'expression de gènes a été réalisée lors de deux cinétiques avec des échantillons collectés à différents temps post infection avec la souche Gardel passage 36 et passage 38. Pour chacune des cinétiques, la lyse a été complète et synchrone à 120 heures post infection. Les échantillons de Gp36 et de Gp38 constituent donc des réplicats biologiques. A chaque temps, 1/10 du culot cellulaire est utilisé pour une quantification du nombre de copies d'ADN d'*E. ruminantium* et les 9/10 restants sont conservés à -80°C pour l'extraction de l'ARN total par le TRIzol (Invitrogen). Puis 20 µL d'ARN sont rétro-transcrits en cDNA en utilisant l'amorce aléatoire hexamérique pd(N)<sub>6</sub> du kit « First-Strand cDNA Synthesis kit » (*Amersham Biosciences*).

### *3.2.4. Test des amorces et efficacité de la qPCR*

L'amplification des gènes ciblés a été réalisée par l'intermédiaire de la qRT-PCR en utilisant du Sybergreen (*Applied Biosystems*). Le test des amorces se fait dans 25 µL de milieu réactionnel. Le mix composé de Sybergreen PCR master mix 2x (*Applied Biosystems*), de 900 nM à 150 nM d'amorce sens et anti-sens et d'H<sub>2</sub>O est ajouté à 1 µL d'ADN témoin Gardel passage 27 (Gp27) dilué au 1000<sup>ème</sup>. Un échantillon sans ADN constitué d'eau correspondant au témoin négatif est inclus systématiquement dans chaque essai. Les conditions d'amplification dans le thermocycleur (*ABI-Prism 7000, Applied Biosystems*) sont : dénaturation initiale 10' à 95°C, 40 cycles comprenant 30'' à 95°C et 30'' à 60°C avec un cycle de dissociation : 15'' à 95°C, 20'' à 60°C et 15'' à 95°C. La lecture de la fluorescence

se fait à chaque cycle lors de l'élongation ; la courbe de dissociation en fin de run permet de vérifier la présence d'un seul produit d'amplification. De manière à calculer l'efficacité des qPCR ciblant chaque gène, une gamme étalon à partir de dilutions sériées de l'ADNg d'ERGp27 quantifié a été réalisée. Efficacité=  $(10^{-1/\text{pente}}-1) \times 100$  avec une pente de la droite représentant les Ct en fonction du log de la concentration en ADNg.

### 3.2.5. Quantification du nombre de copies d'ADN d'ER par qPCR Taqman *map1*

Pour la quantification en ADNg d'*E. ruminantium* extrait à partir des cinétiques ERGp38 et ERGp36, nous avons utilisé la méthode de référence utilisant une sonde Taqman spécifique du gène *map1*, comme expliqué précédemment (Partie I, 1.2.3).

### 3.2.6. Mesure du niveau d'expression des gènes candidats

La mesure de l'expression des gènes sélectionnés a été effectuée par qRT-PCR pour chaque temps post infection en parallèle avec une mesure de la quantité de bactéries présentes par échantillon par qPCR. Les gènes dont l'expression varie le moins au cours du cycle de développement seront choisis d'après la stabilité du rapport R calculé selon la formule :

$R = \text{nbre de copies } map1 \text{ dans l'ADN} / \text{nbre de copies du gène cible dans le cDNA}$ . Un facteur de correction (1000x) a été appliqué au calcul de R pour *16S* et *map1*. Une vérification de la contamination de l'ARN par l'ADN génomique (ADNg) est faite en même temps que la quantification du gène cible dans le cDNA en ajoutant dans l'essai pour chaque échantillon, l'ARN non transcrit. Il a été vérifié que la contamination par de l'ADN génomique était négligeable dans les ARN extraits avec un écart de Ct de plus de 5 cycles entre l'ARN et le cDNA correspondant à chaque gène et pour chaque temps.

L'amplification et la quantification a été réalisée selon les conditions optimales établies précédemment. La réalisation d'une gamme étalon avec l'ADN ERGp27 (concentration connue) permet de quantifier les copies de gène amplifiées à partir des cDNA amplifiés. La quantification est faite en dupliqué systématiquement. Tous les échantillons sont utilisés purs sauf pour la quantification de l'expression de l'ARNr *16S* pour lequel les cDNAs sont dilués au 1/1000.

### 3.3. Résultats

#### 3.3.1. Conditions optimales de qPCR et efficacité

Les conditions optimales de concentration en amorces sont présentées dans le Tableau 5. Les couples *16S* F2-R2 et *ffh* F1-R1 n'ont pas été utilisés car ils amplifient soit une séquence non spécifique en plus du produit attendu (*16S* R/F 2) soit ils présentent un bruit de fond dans le témoin négatif dû aux dimères d'amorces (*ffh* F/R 1).

Tableau 5 : Conditions optimales de qRT-PCR pour chaque gène candidat

Couple d'amorces	[Primers] (nM)	T hybridation (C°)
<b>16S F1-R1</b>	150	
<b>ffh2 F2-R2</b>	150	
<b>recA F1-R1</b>	150	60
<b>recA F2-R2</b>	300	
<b>rpoD1 F1-R1</b>	300	
<b>rpoD2 F2-R2</b>	300	

Pour l'ensemble des résultats obtenus les efficacités des PCRs sont bonnes : elles sont comprises entre 95% et 99,8% (Tableau 6).

Tableau 6 : résultats des efficacités des qRT-PCR

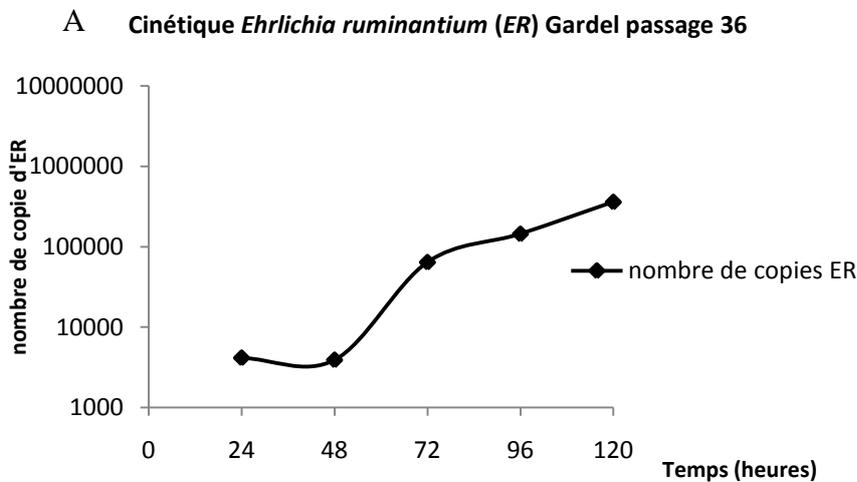
Couple d'amorces	Pente	Efficacité en %
<b>16S1</b>	-3,3	99,2
<b>ffh2</b>	-3,4	95,0
<b>recA1</b>	-3,3	99,8
<b>recA2</b>	-3,3	99,8
<b>rpoD1</b>	-3,4	98,3
<b>rpoD2</b>	-3,3	97,3

3.3.2. Quantification du nombre de copies d'ADN d'ER par qPCR Taqman map1

La quantification de l'ADN (Tableau 7) aux différents stades de développement de la bactérie est représentée figure 15. Après une période de répliation de 48 heures, on observe une augmentation exponentielle du nombre de copies d'*E. ruminantium* ; le maximum ( $1.97 \times 10^6$  copies/  $\mu\text{L}$ ) étant atteint à 120 h, ce qui a déjà été observé pour la cinétique ERGp36 (réplicat technique 1). La quantité de copies d'*E. ruminantium* est bien supérieure à 120 h pour ERGp38 que pour ERGp36 :  $3.6 \times 10^5$  copies pour ERGp36 comparé à  $1.9 \times 10^6$  pour ERGp38.

Tableau 7 : Quantification du nombre de copies d'ADN chez ERGp38 et ERGp36 au cours du temps

	<b>ERGp38 Nbre de copies/<math>\mu\text{L}</math></b>	<b>ERGp36 Nbre de copies/<math>\mu\text{L}</math></b>
<b>24 hpi</b>	9117	4151
<b>48 hpi</b>	14410	3915
<b>72 hpi</b>	321585	64208
<b>96 hpi</b>	1690000	144383
<b>120 hpi</b>	1970000	360556



**B Cinétique *Ehrlichia ruminantium* (ER) Gardel passage 38**

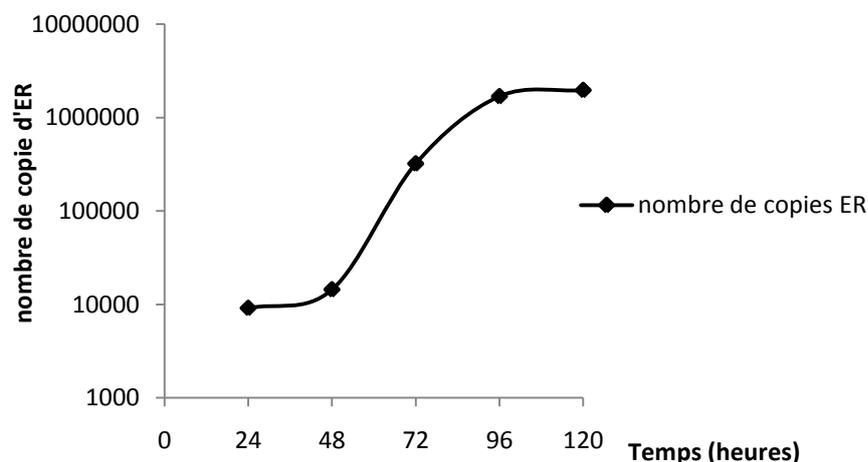


Figure 15 : Quantification absolue de ERGp36 (A) et ERGp38 (B) par qPCR à l'aide d'une sonde taqman spécifique du gène *map1*

3.3.3. Evaluation de la stabilité des gènes constitutifs par qRT-PCR

Les résultats obtenus montrent que le nombre de copies d'ADNc (Tableau 8) augmente au fur et à mesure du développement de la bactérie de 24 à 96 h, quel que soit le gène observé, avec une diminution importante des transcrits de chaque gène à 120 h. La comparaison du nombre de copies de cDNA pour chaque gène et à chaque temps confirme que les gènes *map1* et *16S* sont plus fortement exprimés que les autres, pour une même quantité d'ADNc.

Tableau 8 : Nombre de copies d'ADNc quantifié pour la souche ERGp38

Nbres de copies/ $\mu$ L	24 h	48 h	72 h	96 h	120 h
<b>recA1</b>	434	1462	72007	237772	7823
<b>ffh2</b>	257	586	38065	292049	20738
<b>rpoD2</b>	107	242	18468	166386	16907
<b>16S1</b>	4,82E+05	8,24E+05	2,00E+07	1,35E+08	7,91E+07
<b>pCS20</b>	966	1156	46414	193725	27234
<b>map1</b>	31315	81233	1,46E+06	2,58E+06	1,76E+05

De même, les rapports R représentant les niveaux d'expression des 6 gènes normalisés par la quantité de bactéries, indiquent que *I6S* et *map1* sont fortement exprimés comparés à *recA*, *ffh* et *rpoD*. Cela se traduit par des rapports bruts faibles lorsqu'ils ne sont pas corrigés (Tableau 9A et 9B). La stabilité des rapports d'un temps à l'autre traduit une augmentation de l'expression proportionnellement au nombre de copies d'ADN. Les résultats mettent en évidence une stabilité de l'expression du gène *I6S*, d'un stade de développement à l'autre avec R variant de 13 à 25. L'expression de *pCS20* est stable (R de 9 à 12) jusqu'à 120 hpi où la quantité de transcrits diminue considérablement (R=72). L'expression des autres gènes varient entre les temps sauf pour *map1* entre 48 et 72 h (Tableau 8A, R variant de 18 et 22). Pour *ffh* et *rpoD*, les variations sont plus grandes, surtout entre les stades précoces (24 et 48 h) et la phase exponentielle de croissance (tableau 8A/B). Lors de la répétition biologique (ERGp36), nous avons confirmé une stabilité d'expression de *I6S* entre 96 et 120 hpi contrairement à *ffh* et *rpoD* (Tableau 9B). L'expression de *recA* semble stable entre 96 hpi et 120 hpi pour le réplicat biologique.

Tableau 9A : Stabilité de l'expression des gènes (R) chez Gp38 au cours du temps

<b>R</b>	<i>recA1</i>	<i>ffh2</i>	<i>rpoD2</i>	<i>I6S1*</i>	<i>pCS20</i>	<i>map1*</i>
<b>24 hpi</b>	21	35	85	<b>19/</b> 0,0189	9	<b>29/</b> 0,2911
<b>48 hpi</b>	10	25	60	<b>17/</b> 0,0175	12	<b>18/</b> 0,1774
<b>72 hpi</b>	4	8	17	<b>16/</b> 0,0160	7	<b>22/</b> 0,2203
<b>96 hpi</b>	7	6	10	<b>13/</b> 0,0125	9	<b>66/</b> 0,6550
<b>120 hpi</b>	252	95	117	<b>25/</b> 0,0249	72	<b>1121/</b> 11,21

\* pour *I6S*, indication du rapport multiplié en gras (facteur 1000) et brut

Tableau 9B : Stabilité de l'expression des gènes (R) chez Gp36 au cours du temps

<b>R</b>	<i>recA1</i>	<i>ffh2</i>	<i>rpoD2</i>	<i>I6S1</i>
<b>96 hpi</b>	59 (+/- 15)	48 (+/- 8)	47 (+/-16)	44 (+/- 2)/ 0,0044
<b>120 hpi</b>	67 (+/-1)	109 (+/- 5)	204 (+/- 75)	38 (+/- 4)/ 0,0038

\* pour *I6S*, indication du rapport multiplié en gras (facteur 1000) et brut

### 3.4. Conclusion

Ce travail a permis d'évaluer des gènes décrits comme étant stables chez d'autres modèles bactériens et des gènes spécifiques d'*E. ruminantium* comme *map1* et *pCS20*, pouvant être utilisés comme gènes normalisateur en qRT-PCR. La stabilité du gène *16S* a été démontré au cours des différents stades de développement *in vitro* comme ce qui avait été décrit précédemment pour les bactéries *Pectobacterium astrosepticum*, *Salmonella enterica* et *Listeria monocytogenes* respectivement (Takle *et al.*, 2007). Ce gène semble être un bon candidat pour la validation des résultats de microarrays par qRT-PCR bien que son niveau d'expression soit particulièrement fort. Il sera probablement plus fortement exprimé que les gènes que nous aurons sélectionnés par micro-arrays. Les gènes *recA* et surtout *pCS20* présentent une bonne stabilité d'expression au cours du cycle infectieux *in vitro* excepté à 120 hpi et semblent être des bons candidats pour être utilisés comme gènes normalisateurs notamment pour des comparaisons avec des gènes très faiblement exprimés. L'analyse haut-débit par micro-arrays donnera aussi des indications sur la stabilité de ces gènes au cours du temps et entre souches. Elle permettra de révéler d'autres gènes constitutifs pouvant être utilisé comme normalisateur en qRT-PCR s'ils sont identifiés comme stables en microarray.

## **4. Adaptation de la méthode SCOTS à la bactérie *Ehrlichia ruminantium* et validation sur puces à ADN pour l'analyse du transcriptome**

### 4.1. Contexte

L'utilisation des puces à ADN nécessite de l'ARN messager (ARNm) bactérien de qualité et en quantité suffisante. Or l'isolement de l'ARNm bactérien, présent en faible quantité parmi l'ARN total (eucaryote et procaryote) représente une phase critique. De plus, l'ARNm bactérien ne représente qu'une très faible proportion de l'ARN total bactérien car l'ARNr est présent en grande quantité (> 80%). Pour pallier à ce problème nous avons eu recours à la méthode SCOTS (selective capture of transcribed sequences) initialement développée pour déterminer les gènes exprimés chez *Mycobacterium tuberculosis* lors de l'infection de macrophages humains en culture *in vitro*. Cette technique qui consiste à capturer les transcrits

bactériens parmi l'ARN total a été adaptée à *E. ruminantium*, par le biais d'hybridation spécifique et répétée de l'ADN génomique bactérien biotynilé à l'ADN complémentaire bactérien (comme précisé dans la section 4.5 Chapitre I). Les échantillons (cDNA) obtenus après SCOTS pour la souche Gardel virulente passage 39 isolés à partir de culture sur cellules endothéliales bovines, ont servis à effectuer plusieurs tests de validation de la méthode pour notre modèle. Les tests ont été effectués à plusieurs temps post infection notamment aux temps précoces et tardifs. Cette méthode, déjà éprouvée pour les parasites intracellulaires facultatifs n'avait jamais été appliquée à un pathogène intracellulaire obligatoire auparavant.

#### 4.2. Validation de la méthode SCOTS

1/ Dans un premier temps la capture de cDNA a été vérifiée par PCR ciblant la séquence étiquette *nKpnI*. Ensuite nous avons effectué des *Southern blots* après chaque ronde de SCOTS pour observer l'enrichissement spécifique et l'augmentation de la diversité des cDNA bactériens cibles. Puis nous avons réalisé des PCR quantitatives en temps réel ciblant l'ARNr *16S* sur les échantillons générés au cours des différentes rondes de capture SCOTS pour démontrer l'efficacité de la technique dans la diminution du contaminant ribosomal procaryote.

2/ Nous avons également ciblé 5 gènes *map1*, *rpoD*, *ffh*, *pcs20*, *recA*, pour évaluer par RT-PCR leur niveau ou intensité de détection par SCOTS après chaque ronde, aux différents temps post-infection. Une dilution homogène était appliquée pour chaque temps de collecte post-infection, faible aux temps précoces ( $10^3$  à 24 hpi) et fort au temps tardifs ( $10^5$  à 120 hpi). Globalement les résultats de RT-PCR montrent que sans SCOTS (0x), il n'y a aucune détection quelque soit le gène. On observe également que SCOTS permet une détection de la majorité des gènes dès la 1<sup>ère</sup> ronde de capture entre 24 h et 120 h avec des niveaux d'intensités variables. Certains gènes ne sont détectés qu'à la troisième capture (3x) selon la condition considérée (ex : *recA* et *rpoD* à 120 hpi).

3/ En ciblant 3 des 5 gènes précédemment cités (*map1*, *recA* et *rpoD*), nous avons voulu vérifier par qRT-PCR qu'il n'y avait pas de modification majeure du niveau relatif d'expression de ces gènes entre SCOTS 0x et SCOTS 3x. Les résultats obtenus sans aucune dilution montrent la nécessité de SCOTS aux temps précoces puisque un très faible nombre de

copies de *map1*, *recA* et *rpoD* est détectable à 24 hpi et 48 hpi alors qu'après 3 rondes de SCOTS ces 3 gènes sont en plus grandes quantités, entre 942 et 62000 fois plus, selon le gène. On observe également avant capture des expressions différentielles avec 1 à 2.5 log<sub>10</sub> de différence de niveau d'expression selon les conditions. Après capture on observe une normalisation du niveau d'expression des gènes, surtout au temps tardifs (72 hpi-120 hpi). Il reste cependant toujours 1log<sub>10</sub> d'écart là où les différences d'expressions étaient initialement plus grandes (jusqu'à 2.5log<sub>10</sub>).

4/ Ensuite, les cDNA enrichis par SCOTS ont été hybridés sur les puces à ADN pour observer l'efficacité de la détection. A l'aide des microarrays, nous avons comparé l'efficacité de la méthode SCOTS par rapport à la méthode RiboMinus qui permet d'enlever l'ARN eucaryote. L'analyse des données brutes a démontré la nécessité d'effectuer SCOTS surtout au temps précoces pour lequel la détection est nulle sans SCOTS (SCOTS 0x). Au temps tardif sans SCOTS il y a une détection de 25% des gènes contrairement au 74% détectés après 3 rondes de SCOTS (SCOTS 3x). Les échantillons SCOTS 1x et 2x pour ERGp39 ont également été hybridés sur puce à ADN d'*E. ruminantium*. Les analyses de corrélation entre les échantillons SCOTS 1x, 2x, 3x donnent de très bons résultats (0.98), alors que la corrélation entre SCOTS 0x et SCOTS 3x est de 0.7.

Ce travail a permis de valider l'adaptation de la méthode SCOTS à notre modèle *E. ruminantium*. Il nous permet de réaliser l'étude du transcriptome mais aussi plus largement il a une implication pour l'étude d'autres pathogènes intracellulaires obligatoires.

Pour la réalisation de ce travail plusieurs collaborations ont été effectuées. Nous disposons de toute l'expertise nécessaire pour générer les cinétiques bactériennes. En collaboration avec l'Inria Grenoble (Alain Viari), nous avons développé les outils bioinformatiques permettant l'analyse génomique comparée d'*E. ruminantium* mais aussi les sondes oligonucléotidiques permettant de créer les puces à ADN pangénomiques nécessaires à l'analyse transcriptomique et fonctionnelle. Le professeur France Daigle (Université de Montréal) a apporté l'élément clé qui est la méthodologie de capture spécifique des ARNm bactériens (SCOTS), améliorant de façon conséquente la qualité de l'hybridation sur lame. Enfin, l'IPMC Sophia Antipolis (Bernard Mari et Pascal Barbry) a apporté un savoir faire éprouvé en matière de création et d'utilisation des puces à ADN et d'interprétation des données, ainsi qu'une plateforme transcriptomique opérationnelle et reconnue. La combinaison de ces expertises a permis de

réaliser la première analyse microarrays chez cette bactérie intracellulaire dans des conditions d'infection *in vitro*, ouvrant la voie à un vaste champ d'investigation du transcriptome.

Les résultats de ce travail ont été publiés dans le journal (*BMC Molecular Biology*) sous forme d'article méthodologique disponible sur <http://www.biomedcentral.com/1471-2199/10/111>.

Methodology article

Open Access

## Innovative approach for transcriptomic analysis of obligate intracellular pathogen: selective capture of transcribed sequences of *Ehrlichia ruminantium*

Loïc Emboulé\*<sup>1</sup>, France Daigle<sup>2</sup>, Damien F Meyer<sup>1</sup>, Bernard Mari<sup>3</sup>, Valérie Pinarello<sup>1</sup>, Christian Sheikboudou<sup>1</sup>, Virginie Magnone<sup>3</sup>, Roger Frutos<sup>4</sup>, Alain Viari<sup>5</sup>, Pascal Barbry<sup>3</sup>, Dominique Martinez<sup>4</sup>, Thierry Lefrançois<sup>1</sup> and Nathalie Vachiéry\*<sup>1</sup>

Address: <sup>1</sup>UMR 115 CIRAD-INRA «Contrôle des maladies animales exotiques et émergentes», Site de Duclos, Prise d'Eau 97170, Petit Bourg, Guadeloupe, <sup>2</sup>Département de microbiologie et immunologie, Université de Montréal, C.P 6128 succursale Centre-ville, Montréal, QC H3C3J7, Canada, <sup>3</sup>UMR6097, CNRS-Université de Nice Sophia Antipolis, Institut de Pharmacologie Moléculaire et cellulaire, Sophia Antipolis, 06560, France, <sup>4</sup>UMR 115 CIRAD-INRA «Contrôle des maladies animales exotiques et émergentes», TA 30/C Campus international de Baillarguet 34398 Montpellier Cedex 5, France and <sup>5</sup>Inria Rhône-Alpes Projet HELIX, 655 Av. de l'Europe, 38330 Montbonnot-Saint Martin, France

Email: Loïc Emboulé\* - loic.emboule@cirad.fr; France Daigle - france.daigle@umontreal.ca; Damien F Meyer - damien.meyer@cirad.fr; Bernard Mari - bernard.mari@unice.fr; Valérie Pinarello - valerie.pinarello@cirad.fr; Christian Sheikboudou - christian.sheikboudou@cirad.fr; Virginie Magnone - magnone@ipmc.cnrs.fr; Roger Frutos - roger.frutos@cirad.fr; Alain Viari - Alain.Viari@inrialpes.fr; Pascal Barbry - barbry@ipmc.cnrs.fr; Dominique Martinez - dominique.martinez@cirad.fr; Thierry Lefrançois - thiery.lefrancois@cirad.fr; Nathalie Vachiéry\* - nathalie.vachier@cirad.fr

\* Corresponding authors

Published: 24 December 2009

Received: 9 April 2009

BMC Molecular Biology 2009, 10:111 doi:10.1186/1471-2199-10-111

Accepted: 24 December 2009

This article is available from: <http://www.biomedcentral.com/1471-2199/10/111>

© 2009 Emboulé et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Whole genome transcriptomic analysis is a powerful approach to elucidate the molecular mechanisms controlling the pathogenesis of obligate intracellular bacteria. However, the major hurdle resides in the low quantity of prokaryotic mRNAs extracted from host cells. Our model *Ehrlichia ruminantium* (ER), the causative agent of heartwater, is transmitted by tick *Amblyomma variegatum*. This bacterium affects wild and domestic ruminants and is present in Sub-Saharan Africa and the Caribbean islands. Because of its strictly intracellular location, which constitutes a limitation for its extensive study, the molecular mechanisms involved in its pathogenicity are still poorly understood.

**Results:** We successfully adapted the SCOTS method (Selective Capture of Transcribed Sequences) on the model Rickettsiales ER to capture mRNAs. Southern Blots and RT-PCR revealed an enrichment of ER's cDNAs and a diminution of ribosomal contaminants after three rounds of capture. qRT-PCR and whole-genome ER microarrays hybridizations demonstrated that SCOTS method introduced only a limited bias on gene expression. Indeed, we confirmed the differential gene expression between poorly and highly expressed genes before and after SCOTS captures. The comparative gene expression obtained from ER microarrays data, on samples before and after SCOTS at 96 hpi was significantly correlated ( $R^2 = 0.7$ ). Moreover, SCOTS method is crucial for microarrays analysis of ER, especially for early time points post-infection. There was low detection of transcripts for untreated samples whereas 24% and 70.7% were revealed for SCOTS samples at 24 and 96 hpi respectively.

**Conclusions:** We conclude that this SCOTS method has a key importance for the transcriptomic analysis of ER and can be potentially used for other Rickettsiales. This study constitutes the first step for further gene expression analyses that will lead to a better understanding of both ER pathogenicity and the adaptation of obligate intracellular bacteria to their environment.

## Background

Elucidating molecular mechanisms that drive the adaptation of obligate intracellular pathogens to their host is crucial to understand their pathogenesis. To date, molecular studies on obligate intracellular bacteria can only be performed *ex vivo* at one time or *in vitro* in host cells. Thus, RNA extraction from infected cell cultures leads to low quantities of prokaryotic mRNAs with short half-lives and a high amount of contaminant eukaryotic RNAs [1,2]. Moreover, in prokaryotic RNA, ribosomal RNAs (rRNAs) represent more than 80% of total RNA, whereas mRNAs represent only 2% of total RNAs. Therefore, high throughput gene expression analysis of obligate intracellular bacteria depends strongly on the quality of mRNAs samples, deprived from ribosomal RNAs and host RNAs. Up to recently, no methods were available to obtain purified obligate intracellular bacteria mRNAs from infected cells. Various methods can be used to monitor the complete set of RNA molecules produced by a microorganism, including both targeted and random approaches. Among the latter are differential expression of customized amplification libraries (DECAL) [3] and Selective Capture Of Transcribed Sequences (SCOTS) [4], techniques that combine polymerase chain reaction (PCR) and subtractive hybridization in order to identify genes that are expressed differentially. DECAL method is a powerful technique that permits global comparisons of bacterial gene expression under various growth conditions. It allows direct determination of differential gene expression by comparison of relative intensity with which PCR probes hybridize with individual colonies. However, this method has the disadvantage of being time-consuming and more complex to implement because of the construction of the Customized Amplification Library (CAL). Moreover, this technique does not assure to cover all the genome and several genes could be not detected, thus compelling to construct more complete CALs. Selective capture of transcribed sequences (SCOTS) was initially developed by Graham and Clark-Curtiss in 1999 for the non obligatory intracellular pathogen *Mycobacterium tuberculosis* and allowed to enlighten bacterial gene expression from different growth conditions in macrophages cells. It was also later used for *Salmonella enterica* serovar Typhi [5] and then used successfully for further transcriptomic microarray analysis [6]. Recently, SCOTS was employed to identify the *in vivo* expression of several genes of *Actinobacillus pleuropneumoniae* at different developmental stages post infection [7,8] but was never applied on obligate intracellular pathogens.

The Rickettsia *Ehrlichia ruminantium* (ER), (previously *Cowdria ruminantium*) is the causative agent of heartwater, which affects both wild and domestic ruminants and is transmitted by ticks of the genus *Amblyomma* [9]. Heartwater represents a serious problem for livestock productivity in endemic areas such as sub-Saharan Africa and the West-indies and it poses a severe threat to livestock in the American continent due to migratory birds and the presence of potential indigenous vector ticks [10,11]. The genotypic heterogeneity of the bacterium leads to difficulties for the generation of an efficient vaccine [12-15]. Little is known about the genetic determinants and molecular mechanisms of ER pathogenesis, due to its isolated intracellular location. ER was considered to be a good model to evaluate the feasibility of SCOTS method for obligate intracellular bacteria. Indeed, ER is a gram negative bacterium that belongs to alpha proteobacteria and is an obligate intracellular pathogen that infects the endothelium of all blood vessels. ER has a complex life cycle described as chlamydia-like developmental cycle [16]. In the early stage of the cycle, elementary bodies, which represent the extracellular and infectious forms of the parasite, adhere to host target cells and then are engulfed. They remain within intracytoplasmic vacuoles, where they divided by binary fission to produce intermediate bodies and further reticulated bodies. After 4 to 6 days, the disruption of host cell leads to the release of numerous elementary bodies thus initiating a new infectious cycle. The genomic organization of this microorganism was revealed by the genome sequencing of two strains: Gardel and Welgevonden [17]. Even if comparative genomic studies provided data on the active mechanisms of genome plasticity [18,19], almost 30% of genes had unknown functions and genes involved in virulence, host cell penetration or invasion and intracellular growth processes remain unidentified.

Expression analysis of ER genes during life cycle, between attenuated and virulent strains, will allow the identification of the key factors involved in virulence mechanisms and the development of the bacteria. From the sequencing of Gardel and Welgevonden strains, whole genome ER microarrays were designed in order to validate sample quality obtained by SCOTS method and to evaluate the potential use of this method for further ER whole transcriptomic analysis. *In vitro* model using bovine endothelial cells and virulent Gardel strain allowed us to obtain a sufficient amount of ER RNA to perform such a study.

In this article, we report for the first time the successful adaptation of the SCOTS method to an obligate intracellular bacterium, *ER*. We demonstrate the efficient isolation of specific bacterial transcripts from total RNA after 3 rounds of capture, with low amounts of 16S ribosomal RNA contaminant. Moreover, besides the use of PCR amplifications, the differential gene expression was still detected by qRT-PCR and microarrays analysis. SCOTS method seems to be crucial for the analysis of gene expression especially at early stage of *ER* development during the lag phase.

## Methods

### Extraction of ER RNA from bovine infected cells

The Gardel strain stock was isolated in 1982 in Guadeloupe from a goat infected experimentally with *Amblyomma variegatum* ticks collected from cows [20]. Gardel strain passage 39 and passage 47 were multiplied successively in bovine aorta endothelial (BAE) cells grown in Glasgow minimal essential medium complemented with fetal calf serum, tryptose-phosphate broth, and antibiotics [21] at 37°C, 5% CO<sub>2</sub>, with a weekly passage on fresh cells [22]. BAE cells were infected with a calibrated inoculum of Gardel strain ( $1.1 \times 10^7$  elementary bodies per  $1.42 \times 10^6$  cells). Estimation of bacterial viability and quantity were both carried out by flow cytometry and fluorescent microscopy using live/dead BacLight Bacterial Viability Kit (Invitrogen, France) [23]. The infected cells were incubated at 37°C in a 5% CO<sub>2</sub> atmosphere. The supernatant was renewed only 24 h (6 ml) after infection and the cell monolayer was harvested by trypsinization every 24 h and centrifuged at  $1700 \times g$  for 5 min at 4°C. When 80% cell lysis was observed, after 120 hours post infection (hpi), supernatant and cellular debris were harvested and then ultra-centrifuged at  $20,000 \times g$  for 15 min at 4°C to collect elementary bodies. The pellets were placed in sterile eppendorf and homogenized in 2.5 ml of TRIzol reagent (Invitrogen). The cells lysed immediately and released RNA and DNA in the supernatant. The samples were immediately stored at -80°C before RNA extraction.

### Extraction of total RNA with TRIzol

For each time of collection after cell lysis (24 to 120 hpi), total RNA extraction procedure was carried out with TRIzol reagent (Invitrogen) according to the manufacturer's instructions. RNA pellets were dissolved in 100 µl of DEPC water and treated with turboDNase (Ambion, France) according to manufacturer's protocol in order to remove all contaminant DNA. The removal of bacterial genomic DNA (gDNA) contaminant in RNA samples was verified by PCR targeting *pCS20* gene using primers AB128 and AB129 which amplified specifically a 281 pb of *ER* as described previously [24] (table 1). The quantification of total RNA was performed by fluorimeter using ribogreen reagent (Invitrogen). The yield obtained after

the extraction was between 3 to 25 µg. For each time of collection, total RNA samples were pooled in RNase free water at a final concentration of 0.5 µg/µl.

### ER gDNA production and ER ribosomal DNA (rDNA) cloning for SCOTS method

Genomic DNA (> 50 µg) from Gardel strain passage 40 was extracted from elementary bodies, as previously described [18,25]. After high speed centrifugation ( $20,000 \times g$  during 30 min), the pellet of elementary bodies was resuspended in 350 µl of saline phosphate buffer. 150 µl of DNase (1 µg/ml) was added to remove the contaminant bovine DNA from host cells and the samples were incubated at 37°C for 90 min. The treatment was stopped by adding 25 mM of EDTA. Whole bacterial DNA was obtained using QIAamp extraction kit (Qiagen, France) [26].

In *ER*, the organization of the *m* operon coding the rRNAs is not canonical as the gene coding the 16S rRNA is 900 kb distant from the 23S-5S gene cluster [17]. Both rRNA 16S and 23S - 5S cluster sequences were amplified using specific primers derived from *ER* genome: rRNA 16SGarF- rRNA 16SGarR for 16S and rRNA5GarR; rRNA23SGarR for 5S-23S cluster (table 1). The PCR products were cloned into the high copy number vector pGEM-T Easy (Promega, France). Competent *Escherichia coli* SURE2 cells were transformed with plasmid containing the inserts and isolated on LB-ampicillin-XGal (5-bromo-4-chloro-3-indol-β-D-galactopyranoside)-IPTG (isopropyl-β-D-thiogalactopyranoside) selective medium. Positive clones were selected and the presence of the insert was checked using the specific primers previously cited. Transformants were grown in LB medium with appropriate antibiotic (Ampicillin 100 µg/ml) and plasmid extraction was carried out with plasmid purification Maxi kit (Qiagen) according to manufacturer's protocol. Up to 300 µg of plasmid pellets were homogenized in TE buffer pH 7.5. Both gDNA and rDNA were quantified by fluorometer using picogreen reagent (Invitrogen).

### Selective capture of transcribed sequences (SCOTS)

For each time of the kinetic, 5 µg of total RNA from Gardel strain passage 39 was reverse transcribed by random priming with Superscript II (Invitrogen) according to manufacturer's instructions. The reverse transcription was done using *KpnI*-RNA primers (table 1) containing a defined 5' terminal sequence as a tag and a random nonamer at the 3' end as in conditions previously described by [27] and [5]. Second-strand cDNA was synthesized by using Klenow fragment (Biolabs, France). Then, cDNA was amplified by PCR using the specific primers corresponding to the *KpnI* tag. This corresponded to cDNA before SCOTS (SCOTS 0x). Reactions were prepared using the following PCR conditions: initial denaturation of 3 min at 94°C fol-

**Table 1: Primers used for the detection of specific ER genes**

Primer name	Primer sequence	Target gene or sequence	Product size (bp)	source (references)
<i>ffh-F2<sup>a</sup></i> <i>ffh-R2</i>	5' GGTAGGTCTTCAAGGTGTTGGTAAA 3' 5' AGTTTGAGCTGCAGGACGATATAA 3'	<i>Ffh</i>	121	this work
<i>recAF1<sup>a,b</sup></i> <i>recAR1</i>	5' TTGAAAAGCGTTTGGTCGTG 3' 5' GGGAAACCACCAATACCCAAT 3'	<i>recA</i>	121	this work
<i>rpoDF1<sup>a,b</sup></i> <i>rpoDR1</i>	5' CAGAGGGTTGCAATTTCTTGATT 3' 5' TCTGACCCACCATGTTGCAT 3'	<i>rpoD</i>	121	this work
<i>16SFla<sup>b</sup></i> <i>16S-R1</i>	5' AGCGCAACCCTCATCCTTAG 3' 5' AGCCACCCCTATAAGGGCC 3'	<i>rRNA 16S</i>	121	this work
<i>mapIgardP<sup>a</sup></i> <i>mapIgardR</i>	5' CACTTGAAGGAATGCCAGTTTCTC 3' 5' CTTAGGATTTGTAGCATTGATTACTGACACT 3'	<i>mapI</i>	85	this work
<i>AB128</i> <i>AB129</i>	5' ACTAGTAGAAATTGCACAATCTAT 3' 5' TGATAACTTGGTGCGGGAAATCCTT 3'	<i>pCS20</i>	278	Martinez et al., 2004
<i>NKpnI-pdN9</i>	5' GTGGTACCGCTCTCCGTCGANNNNNNNNN 3'	<i>KpnI</i>	/	Daigle et al., 2001
<i>NKpnI</i>	5' GTGGTACGGCTCTCCGTCGA 3'	<i>NKpnI tag</i>	200-400	
<i>ARN16SGarF</i> <i>ARN16SGarR</i>	5' AACTTGAGAGTTTGATCCTGGCT 3' 5' AGGAGGTAATCCAGCCGCGAGTT 3'	<i>rRNA 16S</i>	1503	this work this work
<i>ARN5SGarR</i> <i>ARN23SGarF</i>	5' TCTCCCGTGCCTTAAGACAAA 3' 5' TTGATGGATGCCTTGGCGTTAA 3'	<i>rRNA 23S</i> <i>rRNA 5S</i>	2935	this work this work

<sup>a</sup>: pair of primers use for RT-PCR<sup>b</sup>: pair of primers use for qRT-PCR

lowed by 25 cycles at 94°C for 45 s, at 60°C for 45 s, at 72°C for 60 s and a final extension of 10 min at 72°C.

Selective capture of bacterial cDNA was done as previously described by [4] and [28]. ER gDNA (0.3 µg) was photobiotinylated and then mixed with the rDNA 16S and 23S+5S (0.5 µg of each plasmid pGEMT) in order to block the rRNAs region sites on the gDNA in TE buffer pH 9. The mixture was then sonicated and precipitated in ethanol 100% (2.5 v/v), NaOAc 3 M (0.1 v/v) and 1 µl Glycogen (1 µg/ml). The gDNA-rDNA mixture and 5 µg cDNA were denatured separately 3 min at 99°C in 4 µl of hybridization buffer (10 mM EPPS [N-(2-hydroxyethyl) piperazine-N'-3-propanesulfonic acid]/1 mM EDTA) and pre-hybridized at 50°C for 30 min. The temperature of hybridization was evaluated from the percentage of GC of ER genome (27%). The prehybridization step allows the hybridization of ER rDNA to the gDNA, as it also allows the normalization of bacterial and eukaryotic cDNA by self-hybridization of highly present cDNA [29,30]. Immediately after adding 1 µl of NaCl 1.5 M, the cDNA and bacterial gDNA pre-blocked with rDNA (gDNA-rDNA) were mixed and hybridized for 18 h at 50°C (hybridization step). Hybrids, representing ER cDNA fixed to ER gDNA, were captured with streptavidin-coated magnetic beads

(Dynal 280) according to the manufacturer's protocol. Selective cDNA was then eluted in 100 µl of NaOH 0.4 N, precipitated and amplified by *NKpnI* specific PCR as described. The PCR products were then visualized by ethidium bromide staining in 1.2% agarose gels (Seakem) in 1 × TAE buffer (40 mM Tris-HCl, 6% acetic acid, 1 mM EDTA, pH8) and purified with PCR purification kit (Qiagen). The initial amount of total RNA before SCOTS method for each time of infection was between 10 to 20 µg and was divided in 2 to 5 tubes with 0.5 µg/tube. After this first round (SCOTS 1×), the tubes corresponding to one condition were pooled and 10 µl to 20 µl of the pooled sample were precipitated depending on the intensity of purified cDNAs from the previous round. Two additional rounds of capture (SCOTS 2× and 3×) were performed for each sample at each time point of infection (24, 48, 72, 96 and 120 hpi) for further microarrays and qRT-PCR analyzes.

#### Microarrays experiments

##### a. ER Microarrays design

ER microarrays (8 × 15 k) used in this study were developed based on long oligo arrays generated by Agilent technology. These arrays contain 60-mer probes corresponding to 936 coding sequences (CDS) of Gardel

strain and 909 sequences of Welgevonden strain, including the specific CDSs for each strain determined by the previous annotation of the two genomes [17]. The probes were designed using the following procedure: first, all 60-mer probes were generated from CDS of Gardel and Welgevonden strains using standard thermodynamical constraints (TM  $\leq$  [78, 83] and GC  $\leq$  [20,36]) and a modified version of the OligoArray program [31]. For non specific CDSs, we chose the same probe for the same pair of orthologs between Gardel and Welgevonden strains. Then, we selected 1 or 2 of the most specific probes per gene by minimizing the number of matches to human and bovine mRNAs, extracted from the Ensembl database [32]. The microarrays contained a total of 1800 probes, with 5 replicates per probe, including 28 bovine genes as negative controls. Experimental data and associated microarray designs have been deposited in the NCBI Gene Expression Omnibus (GEO) <http://www.ncbi.nlm.nih.gov/geo/> under platforms GPL9697, GPL 9698 and serie GSE19208. Two different labeling procedures have been performed depending on whether the sample was treated or not following the SCOTS procedure.

#### b. cDNA labeling

Five hundred nanograms of total cDNA (from Gardel strain passage 39 at 24 and 96 hpi) from SCOTS procedure were randomly amplified and fluorescently labeled with the BioPrime array CGH Genomic labeling System kit (Invitrogen), by the incorporation of Cy3-dCTP (Amersham Biosciences) and then purified on a MinElute cleanup column (Qiagen). Before hybridization, quantification of Cy3-dCTP incorporation was performed by absorbance measurement at 550 nm. The yield of cDNA labeling and the specific activity always exceeded to 1.65  $\mu$ g and 9 pmol of Cy3 per cDNA, respectively, according to the manufacturer's recommendations.

#### c. In vitro transcription (IVT)

Total RNA was extracted as described above from Gardel strain samples passage 47, at 24 and 96 hpi. Eukaryotic ribosomal RNA was removed from the mixture by using the RiboMinus Transcriptome isolation kit (Invitrogen), according to the manufacturer's protocol. The amount and quality of purified RNA were monitored at various points throughout the purification process.

Three hundred nanograms of RiboMinus RNA fraction (Total RNA without eukaryotic rRNA) were then amplified and labeled using the Quick Amp Labeling kit of Agilent. The method consists of converting mRNA primed with a random primer containing a T7 promoter into double strand cDNA with MMLV-RT and then amplifying samples using a T7 RNA polymerase, which generates Cy3-labeled complementary (anti-sense) RNA (cRNA).

#### d. Microarrays hybridization

Cy3-labeled cRNAs and cDNAs obtained with both methods were used for hybridizations with Agilent Gene Expression Hybridization Kit (Agilent Technologies). Arrays were incubated at 65°C for 20 h in the hybridization chamber. After hybridization, arrays were washed according to the Agilent protocol. Genomic DNAs of ER Welgevonden and Gardel strains were labeled using BioPrime array CGH (Invitrogen) and then hybridized as positive control. cDNA samples (SCOTS 0 $\times$ , 1 $\times$ , 2 $\times$ , 3 $\times$ ) from Gardel strain passage 39 and samples generated by IVT from Gardel strain passage 47 at 24 hpi and 96 hpi were hybridized on ER microarrays.

#### e. Microarrays analysis

Arrays were scanned and images were saved in TIFF format. The signal intensities of all spots on each image were quantified by Genepix pro 6.0 software (Molecular Devices Corporation, Downingtown, PA), and data were saved as ".txt" files for further analysis.

Data were log-transformed, mean-centered and reduced for an equal standard deviation between each slide using the GeneANOVA software [33]. The median value for each gene was calculated and correlation coefficients ( $R^2$ ) between the several conditions were calculated. Genes were considered detected when their intensity of fluorescence was superior or equal to 3 fold the mean of background intensity.

#### ER Southern blots

Southern blots were carried out as described previously using the hybridization conditions explained by [34]. cDNAs produced at different times of infection (before SCOTS and after selective captures with 1 $\times$ , 2 $\times$  and 3 $\times$  rounds of SCOTS) with rDNA16S and 23+5S were amplified by PCR digoxigenin (DIG)-labeling mix (Roche diagnostics, Meylan, France) according to manufacturer's instructions. Nylon membranes containing ER gDNA digested by *Hind*III were pre-hybridized at 50°C for 2 h in hybridization buffer [SSPE 6 $\times$  (1 $\times$  SSPE is 0.18 M NaCl, 10 mM NaH<sub>2</sub>PO<sub>4</sub> and 1 mM EDTA {pH 7.7}), 0.5% sodium dodecyl sulfate [SDS], 2% (w/v) blocking reagent (Roche diagnostics). Probes were added to the hybridization buffer and incubated with the membranes for 16 h at 50°C. Blots were then washed twice in buffer containing 2 $\times$  SSPE, 0.1% SDS (w/v). Colorimetric detection was performed with anti-DIG antibody conjugated to alkaline phosphatase (Roche) and BCIP/NBT Buffered Substrate (Sigma Aldrich, France).

#### Real-time quantitative RT-PCR targeting ribosomal 16S genes

In order to evaluate the contaminant of ER rRNA16S transcripts in cDNA samples produced before and after SCOTS, qRT-PCR targeting ER 16S gene was performed on

each sample of Gardel strain passage 39. Sybergreen fluorescent master mix reagent (Applied Biosystem) was used for qRT-PCR. Primers 16S-F1 and 16S-R1 used were described in table 1 and the size of the amplicon was 121 pb. qRT-PCR was performed systematically on cDNA before and after each round of SCOTS (1 $\times$ , 2 $\times$ , 3 $\times$ ), using the following program: initial denaturation of 10 min at 95 $^{\circ}$ C, followed by 35 cycles of denaturation at 95 $^{\circ}$ C for 30 s, hybridization step at 60 $^{\circ}$ C for 30 s and extension at 72 $^{\circ}$ C for 60 s. In order to quantify the number of copies per sample, a standard curve was established using gDNA of Gardel strain passage 27 serially diluted (from  $2.9 \times 10^5$  copies to  $2.9 \times 10^1$ ). Each sample was analyzed in duplicate.

#### ER gene detection by RT-PCR and qRT-PCR in cDNA samples after selective capture

The presence of 5 genes in cDNA samples and their enrichment by SCOTS method were evaluated using RT-PCR and qRT-PCR amplifying small fragments of target genes ( $\sim$ 120-300 pb) (table 1) that are *recA*, *ffh*, *rpoD*, *map1* and *pCS20* [18]. *map1* gene of *ER* encodes an outer membrane protein (*major antigenic protein*) [35,36]. *ER* specific gene *pCS20* is highly conserved and used as target for molecular *ER* detection [24]. *rpoD* gene encoding a polymerase sigma factor [37], *ffh* gene encoding a signal recognition particle protein and *recA* gene (recombinase A) [38] are three housekeeping genes involved in *ER* metabolism. The same primers and PCR conditions were used for RT and qRT-PCR.

For *ER* gene detection by RT-PCR, cDNA samples were diluted in order to assess optimal cDNA concentrations and then to observe the gradual increase of cDNA detection after the different rounds of SCOTS (from 1 $\times$  to 3 $\times$ ). The dilution of cDNA samples depended on the time of infection:  $10^3$  fold for 24 hpi,  $10^4$  fold for 72 hpi,  $10^5$  for 96 and 120 hpi. These dilutions were used for all the target genes except *pCS20*. For *pCS20*, samples were not diluted for 24 hpi, and diluted 10,  $10^3$  and  $10^2$  fold for 72, 96 and 120 hpi. RT-PCR was performed systematically on cDNA before and after each round of SCOTS (1 $\times$ , 2 $\times$ , 3 $\times$ ) using the following program: initial denaturation of 3 min at 94 $^{\circ}$ C followed by 40 cycles of denaturation at 94 $^{\circ}$ C for 50 s, hybridization step at 60 $^{\circ}$ C for 50 s and extension at 72 $^{\circ}$ C for 50 s and a final extension of 7 min at 72 $^{\circ}$ C. DNA of Gardel strain passage 18 was used as a positive control. PCR products were visualized by ethidium-bromide-stained revelation in agarose gels. qRT-PCRs targeting *map1*, *ffh* and *recA*, were performed on the ABI Prism 7000 (Applied Biosystems) in a total reaction volume of 25  $\mu$ l. This reaction contained 2  $\mu$ l of undiluted cDNA template (from the initial cDNA reverse transcribed (0.5  $\mu$ g/ $\mu$ l) before capture and after SCOTS 3 $\times$ ). In order to quantify the number of copies per sample, a standard curve was made with the gDNA of *ER* (from  $2.9 \times 10^5$  to  $2.9 \times 10^1$

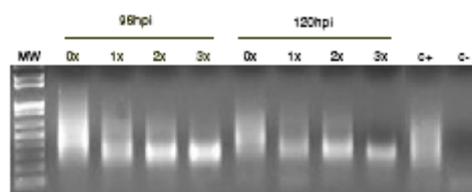
copies) as a template. Each sample was done in duplicate. A dissociation curve was produced in order to verify the presence of a single amplicon.

## Results

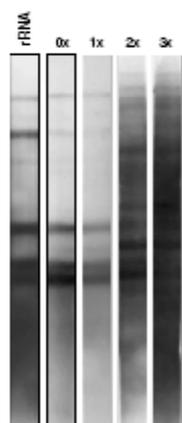
### Enrichment of bacterial cDNA after SCOTS

*ER* cDNAs were produced after 1 $\times$ , 2 $\times$ , and 3 $\times$  SCOTS captures for each time point of infection. A PCR using *KpnI* primer was performed on cDNAs after each capture. Figure 1 shows amplicons obtained after *KpnI* PCR on Gardel strain passage 39, at 96 hpi and 120 hpi. The signal detected after capture confirmed the efficiency of the selective capture by high affinity hybridization of cDNA to gDNA. Moreover, we showed a progressive diminution of the size of the amplified transcribed sequences following successive captures. Similar results were obtained for other post infection time points (data not shown). These results indicated that same capture phenomena were observed independently from the time of infection.

Southern blots were done on cDNAs before and after each capture for each time point of infection. Southern blots obtained using cDNA at 96 hpi as probes on *HindIII*-digested *ER* gDNA are shown in figure 2. DIG-rDNA 23S+5S and 16S were used to reveal the bands corresponding to *ER* rDNA (lane 1). Five distinct bands corresponding to rDNA were observed both before capture and after the first capture (figure 2). After successive capture, the results showed a significant increase of the colorimetric signal that traduces a larger recognition of *ER* genes by DIG-cDNA at 96 hpi. These results demonstrated a progressive increase of bacterial cDNA complexity and amount following the successive rounds of captures. For other time points of infection, southern blots were done systematically and similar results were observed with a diminution of the ribosomal cDNA and an increase of the diversity of *ER* transcripts after the third capture (data not shown).



**Figure 1**  
NKpnI PCR amplicons of cDNAs of Gardel strain passage 39 after SCOTS. 0 $\times$ , 1 $\times$ , 2 $\times$  and 3 $\times$ : cDNA amplicons of NKpnI PCR after 0, 1, 2 or 3 rounds of capture. C+: positive control cDNA with NKpnI tag. C-: Negative control (Water). MW: Molecular weight 100 pb DNA ladder. hpi: hours post Infection



**Figure 2**  
Southern blot using DIG labeling cDNAs of Gardel strain at 96 hpi. 0x, 1x, 2x and 3x: Southern blots showing cDNAs of Gardel strain passage 39 before and after one, two and three rounds of SCOTS. rRNA: Southern blot using DIG-rDNA 23S+5S and 16S.

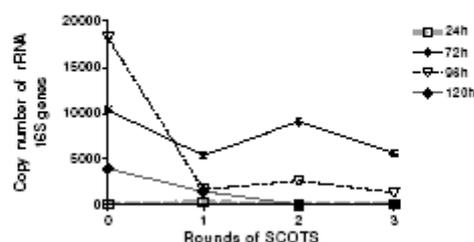
The diminution of the amplicon sizes after *KpnI* PCR and results of Southern blot validates the enrichment of ER cDNA and diminution of ribosomal cDNA and eukaryotic contaminant.

#### Quantification of ribosomal RNA 16S contaminant after SCOTS by qRT-PCR

The quantification of cDNA corresponding to 16S cDNA contaminant was carried out before and after each round of SCOTS at the different time post infection by qRT-PCR (figure 3). Before capture, the amount of 16S cDNA contaminant varied depending on time points post infection and on the amount of bacteria. At 24 hpi, there was a few 16S cDNA contaminant (89 copies per sample). The range of contamination was between  $4 \times 10^3$  to  $18.2 \times 10^3$  copies for other time points post infection. We observed a decrease of 46%, 92% and 99% in ribosomal content after the third capture at 72, 96 and 120 hpi, with a final number of copies of  $5.6 \times 10^3$ ,  $1.4 \times 10^3$  and 43 copies (figure 3). The main decrease was observed after the first capture. For 24 hpi, the initial number of 16S cDNA copies was already low (<90) and remained low after different captures.

#### Detection of different ER transcribed sequences by RT-PCR and qRT-PCR before and after capture

Amplicons corresponding to RT-PCR targeting *map1*, *recA*, *rpoD*, *ffh* and *pCS20* at different time points and following successive captures were shown in table 2. Before capture,



**Figure 3**  
Quantitative RT-PCR targeting ER rRNA 16S on cDNAs of Gardel strain after SCOTS at different time point post infection: 24, 72, 96 and 120 hours post infection.

the cDNAs of these genes were not detected. For *map1* transcripts, one capture was sufficient to obtain a *map1* amplicon at any time post infection. There was a significant and progressive increase of the amplification intensity after the second and third rounds, which demonstrated the enrichment of *map1* transcripts (table 2). Depending on the gene and time point post infection considered, a positive signal was detected after the first (i.e. *ffh* at 24 hpi and *recA*, *ffh* at 72 hpi), second (i.e. *recA*, *ffh* at 96 hpi and *ffh* at 120 hpi) or third capture (i.e. *recA* at 24 hpi and 120 hpi). The increase of signal intensity was observed for all genes at any time point post infection after successive captures, except for *pCS20* at 24 and 72 hpi. At 24 hpi, undiluted samples were used to detect *pCS20* transcripts and there was no signal before capture. At 120 hpi, *pCS20* amplification samples were diluted only at  $10^{-2}$  to observe enrichment, compared to the  $10^{-5}$  dilution used for the other genes. Globally, 3 rounds of capture were necessary to efficiently enrich the genes poorly expressed.

The qRT-PCR targeting *map1*, *recA* and *rpoD* transcripts before and after 3 rounds of capture allowed to quantify the enrichment due to SCOTS method. Results expressed as the number of transcripts for each gene obtained by qRT-PCR are presented in figure 4. Before capture, the number of transcripts was different depending on the target gene. For example, at 24 hpi there were  $1.58 \times 10^3$  copies of *map1* transcripts and only 70 and 10 copies of *recA* and *rpoD*. For all time points post infection, *map1* was highly expressed compared to the 2 other genes (from 1 to 3 log<sub>10</sub> higher than *recA* and *rpoD*) (figure 4).

After 3 captures at 24 hpi, there was enrichment around 1000 fold of the amount of *recA* transcripts (from 70 to  $66 \times 10^3$  copies). For *map1* transcripts, the number of copies increased from  $1.58 \times 10^3$  to  $1 \times 10^8$  copies after capture. Even when there was a single transcript per sample, for

**Table 2: Detection of ER specific genes by RT-PCR on total cDNA before and after SCOTS**

Gene	24 hpi				72 hpi				96 hpi				120 hpi			
	0×	1×	2×	3×	0×	1×	2×	3×	0×	1×	2×	3×	0×	1×	2×	3×
<i>map1</i>	-	+	++	+++	-	+	++	+++	-	+	++	+++	-	+	++	+++
<i>recA1</i>	-	-	-	+	-	+	++	+++	-	-	++	+++	-	-	-	+++
<i>rpoD</i>	-	-	+	++	-	+	++	+++	-	-	++	+++	-	-	-	+++
<i>flh</i>	-	+	++	+++	-	+	++	+++	-	-	++	+++	-	-	++	+++
<i>pCS20</i>	-	+	+	+	-	+	+	+	-	+	++	+++	-	+	++	+++

-: absence of amplicon after RT-PCR, +: Presence of amplicon after RT-PCR.

The number of (+) corresponded to the intensity of amplification signal. 0×, 1×, 2× and 3× corresponded to results obtained on cDNA before and after one, two and three rounds of capture respectively. hpi: hours post infection

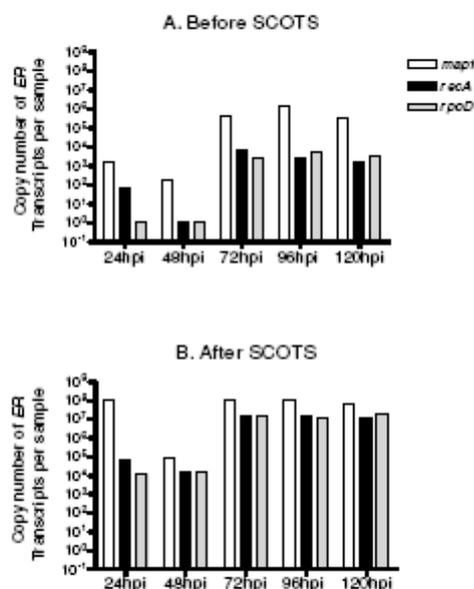
*rpoD* at 24 and 48 hpi and *recA* at 48 hpi, there was  $11 \times 10^3$ ,  $14 \times 10^3$  copies and  $15.4 \times 10^3$  copies after SCOTS. Before capture, there was an approximate 2 to 3 log<sub>10</sub> difference of expression between *map1* and *recA* or *rpoD*, whereas an approximate 1 log<sub>10</sub> difference was measured after capture. The difference of expression between *map1* and *recA* or *rpoD* was still observed after capture, whatever the culture time considered.

#### Validation of selective captures using ER Microarrays

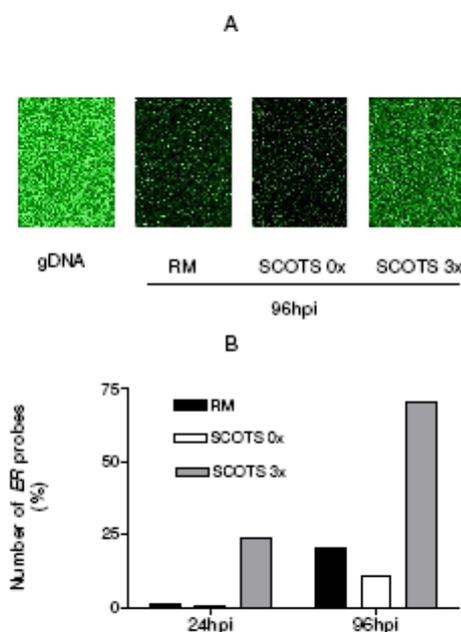
To validate ER selective captures, whole genome microarrays of ER were used. Firstly, ER probes specificity was assessed by hybridizations with gDNA of Gardel and Welgevonden strains. Of the 1800 probes represented on our microarrays, 99.2 (1758/1772) and 99.1% (1757/1772) of probes were detected for Gardel and Welgevonden strains gDNA respectively (figure 5a). There was no detection of bovine probes corresponding to contaminants. When comparing gDNA hybridizations obtained for 5 replicates per probe on two microarrays slides, there was a high correlation coefficient ( $R^2 = 0.97$ ).

Using ER microarrays, two methods for generating ER transcripts probes were evaluated: direct IVT from 'RiboMinus' RNA fraction and cDNA classical random priming for SCOTS samples (figure 5). The percentages of genes detected using samples generated by these methods are presented in figure 5b. Before any treatment (SCOTS 0×), only a small fraction of transcripts could be detected (from 0.05 to 10.4%) for samples collected at 24 and 96 hpi. For early time post infection, there were only 19 genes (1%) detected by microarrays using the IVT method on 'RiboMinus' RNA fraction, whereas 3 rounds of SCOTS allowed the detection of 24% of transcripts (figure 5b). Even at 96 hpi, the percentage of transcripts detected was low using the IVT method. SCOTS method allowed the detection of 7 and 3.5 fold more gene transcripts compared to untreated and RiboMinus samples. Thus, these results confirmed the efficiency of SCOTS method for transcriptomic analysis.

To evaluate a potential bias of the SCOTS method for differential gene expression analysis, we calculated the coefficients of correlation between successive captures on Gardel strain cDNAs at 96 hpi (table 3). The coefficients of correlation between each round of capture were 0.84, 0.98 and 0.98 respectively (table 3). The comparison of an untreated and a three SCOTS capture sample ( $R^2 = 0.7$ ) indicated that SCOTS-mediated amplification was roughly linear.



**Figure 4**  
Quantitative RT-PCR of *map1*, *recA* and *rpoD* genes on cDNA samples of Gardel strain before and after 3 rounds of capture. hpi: hours post infection.



**Figure 5**  
**Detection of probes using Gardel strain cDNAs generated after Ribominus (RM) and SCOTS treatment.**  
 A: ER Microarray Hybridization with gDNA of Gardel strain passage 40, with cRNAs from Gardel strain passage 47 using Ribominus purification (RM), with cDNAs from Gardel strain passage 39 before SCOTS (0x) and after 3 rounds of capture (3x). B: Percentage of ER detected probes at 24 and 96 hpi for cRNAs from Gardel strain passage 47 using Ribominus purification (RM) with cDNAs from Gardel strain passage 39 before SCOTS (0x) and after 3 rounds of capture (3x).

## Discussion

Obligate intracellular pathogens are a challenge for functional genomic studies to identify genes involved in bacterial pathogenesis, especially at different stages of development. Until now, only a few studies have been performed on whole Rickettsiales transcriptomes, thus providing a good illustration of this constraint [39]. Glo-

**Table 3: Correlation coefficient ( $R^2$ ) for gene detection by ER microarrays between the different rounds of SCOTS at 96 hpi**

SCOTS	0x/1x	1x/2x	2x/3x	0x/3x
$R^2$	0.84	0.98	0.98	0.7

bal proteomic expression studies were preferentially used in order to enhance our knowledge on pathogenesis of obligate intracellular pathogens, such as *Ehrlichia*, *Rickettsia* and *Anaplasma* [40-44]. Proteomic studies also avoided the inconvenience of working with ribosomal prokaryotic and host cell contaminants.

In our Rickettsiales model, we postulated that pathogenicity determinants should be differentially expressed in the virulent strains of ER when compared to the same strain attenuated *in vitro* [45,46]. However, any future whole transcriptomic analysis of Rickettsiales bacteria will need to use efficient tools to eliminate both host cells contaminant and prokaryotic ribosomal transcripts.

A method of selection of prokaryotic transcripts, the Ribominus method, based on the removal of eukaryotic ribosomal RNA, was tested for our model. No or few genes were detected on microarrays at any time post infection when using Ribominus cRNA samples. The failure of the method could be due to the interference of important amount of eukaryotic transcripts. Another strategy, combining removal of eukaryotic contaminants with subsequent random amplification of prokaryotic cDNA, was used previously for *Rickettsia conorii* and gave convincing results for microarrays analysis [39,47]. However, SCOTS method for microarrays analysis uses small amounts of initial cDNA (3  $\mu$ g of total RNA) compared to this selective method (50  $\mu$ g for MicroBENRICH) [47]. Thus, using SCOTS method, there was no limitation to produce biological samples and several different conditions (strains and time-points post infection) could be studied.

The innovative method of SCOTS, which has been used previously for non obligatory intracellular parasites [6], was adapted to our ER model to obtain adequate samples for further whole-genome transcripts profile analysis.

The size diminution of ER cDNAs following successive captures illustrates a mechanism of generation of small cDNAs due to the nature of the Taq polymerase. Moreover, it shows the capture of smaller transcribed sequences (around 400 to 200 bp) after each successive round of SCOTS. This phenomenon was observed previously by Graham and Clark-Curtiss for *Mycobacterium tuberculosis* [4]. In their study, they showed that although there were potential biases in representing total mRNA of *Mycobacterium* due to SCOTS method (normalization and enrichment of cDNAs), the use of random priming to create cDNA provided a variety of different transcripts, thereby decreasing potential losses during amplification, normalization and enrichment. A progressive enrichment of ER cDNA between the first and the third capture was observed by Southern blotting as observed with *Mycobacterium tuberculosis* [4].

Interference due to ribosomal contaminants in transcriptomic analysis was shown previously in a study comparing subtractive hybridization and SCOTS methods for *Mycobacterium avium* [48]. For initial samples, few signals were observed corresponding to rRNAs. After subtractive hybridization and SCOTS, an increased quantity of messenger RNAs was observed. The deficiency of detection before mRNA selection seemed to be due to the high amount of rRNAs. In our model, we also obtained a strong decrease of the quantity of ribosomal contaminant as demonstrated by Southern blots and qRT-PCR targeting 16S gene. In an independent experiment, similar results were observed for attenuated Gardel strain by qRT-PCR (data not shown). Even if there was still detection of 16S transcripts in samples after capture, the level of contamination was negligible compared to total ER cDNAs and should not hinder transcriptomic analysis.

The detection of all the 5 tested genes including bacterial housekeeping genes *rpoD*, *ffh* and *recA* by RT-PCR or qRT-PCR suggests that SCOTS method in ER is efficient enough to enhance gene detection. Depending on the gene studied, variable numbers of captures were necessary to detect the specific transcript, illustrating the differential expression of genes in relation to life cycle. Three captures are required in order to detect poorly/lowest expressed genes and used for further transcriptomic analysis. Results of *recA* and *rpoD* qRT-PCR demonstrated that SCOTS method allowed the detection of transcripts accounted even when present as a single copy for early time-points post infection. Thus, our results demonstrated the efficiency of the SCOTS method for further expression analysis of an intracellular pathogen at early time-points post infection where the amount of eukaryotic contaminants was high. Beside this study, we report that *map1* gene was strongly expressed as demonstrated both by RT-PCR and qRT-PCR. In parallel, we showed that *pCS20* was the lowest expressed gene independently of the time of infection. For example, there was no detection of *pCS20* transcripts after RT-PCR at early time post infection on undiluted samples.

As multiple PCRs were used for SCOTS method, one could suggest that all the transcripts would have the same level after selective capture. In this study, we demonstrated by both RT-PCR and qRT-PCR that the differential levels of expression were still observed after SCOTS. In a previous study, transcriptional analysis of *S. enterica* serovar Typhi within the macrophage revealed approximately 300 genes up-regulated at the defined point post infection compared to the supernatant [6].

In order to finalize the validation of SCOTS method for our model, we used ER microarrays. We first hybridized

Gardel and Welgevonden strains gDNA and observed that more than 99% of probes were detected. These microarrays offered an exciting opportunity to do the genome-wide analysis of ER gene expression.

As previously shown by RT-PCR and qRT-PCR on a limited number of genes, our microarrays results with samples generated by SCOTS confirmed the efficiency of this method for our model. Thus, SCOTS method seems to be ideal for whole genome expression profiling of ER. This method is crucial for the study of early time-points post infection: 24% of ORFs could be detected whereas less than 1% was detected on untreated cDNA samples. For late time-points post infection, up to 70.7% of ORFs were detected after 3 rounds. Considering all the time points of infection, 80% of the annotated ORFs were detected in our model (data not shown), which is similar to what was previously observed with *Salmonella* Typhi within the macrophage [6]. Through the use of SCOTS, comparison of gene expression between ER stages of development and between virulent and attenuated strains could be done on the overall CDS. This allows the targeting of genes involved in the invasion of host cells, in metabolism associated with bacterial growth (cell wall biogenesis, energy production, translation, traduction) and in pathogenesis *sensu stricto*.

Moreover, microarrays results demonstrated for overall ER genes that there was a good correlation between expression of genes comparing any round of SCOTS at 96 hpi. The lowest coefficient of correlation ( $R^2 = 0.7$ ) was between SCOTS 0 $\times$  and 3 $\times$  samples. This was mainly due to the absence of detection of several genes before capture which could be detected after 3 rounds: the percentage of genes detected increased from 10.4 (SCOTS 0) to 70.7% (SCOTS 3). For genes already detected before capture, their level of expression was saturated after capture diminishing the correlation between SCOTS 0 $\times$  and 3 $\times$ . For attenuated Gardel, SCOTS 0 $\times$  and 3 $\times$  samples (96 hpi) were hybridized on ER microarrays and a higher correlation coefficient ( $R^2 = 0.87$ ) was observed (data not shown). This preliminary result on attenuated Gardel strain, confirmed the limited bias due to SCOTS method on the gene expressions. Our microarrays data supported results obtained by qRT-PCR on *map1*, *recA* and *rpoD* (at any time-point post infection). The differential of gene expression diminished but was still detected before and after capture. However, for further transcriptomic analysis, we will focus on genes strongly differentially expressed or presence/absence of genes in order to target genes mainly involved in pathogenesis [39,49]. In our model, difference of gene expression will be assessed by hybridization of SCOTS cDNA on total ER microarrays and then validated by qRT-PCR on untreated cDNA.

## Conclusions

Our study reported herein demonstrated that SCOTS method has proven to be suitable for microarray-based transcriptome analysis of ER and as such can be potentially applicable to other obligate intracellular bacteria. SCOTS method avoids interferences due to host cells and prokaryotic ribosomal contaminants. Moreover, it allows the enhancement of specific transcripts and induced a limited bias in their relative amount. Thus, SCOTS method will offer the opportunity to study molecular mechanisms that take place in early stages of ER infection and to identify genes involved in the pathogenesis of this obligate intracellular bacterium.

## Authors' contributions

Conceived and designed the experiments: LE, FD, DFM, BM, RE, AV, PB, DM, TL, NV. Performed the experiments: LE, VP, CS, VM, NV. Analyzed the data: LE, DFM, BM, NV. Contributed reagents/materials/analysis tools: LE, FD, BM, AV. Wrote the paper: LE, FD, DFM, TL, NV. All authors read and approved the final manuscript.

## Acknowledgements

This work was supported by EPIGENEVAC project under contract number [INCO-CT-2005-003713] financed by the European Union, by "Institut National en Recherche Agronomique" and "Centre de Coopération International en Recherche Agronomique pour le Développement" financial fund and by Franco-Canadian fund: "Fonds Franca-Canada pour la Recherche".

We thank Dr Rupert Pogram for his helpful comments.

## References

- Bolland RJ, Zhong G, Crane DD, Hogan D, Sturdevant D, Sharma J, Beatty WL, Caldwell HD: Genomic transcriptional profiling of the developmental cycle of *Chlamydia trachomatis*. *Proc Natl Acad Sci USA* 2003, 100(14):8478-8483.
- Hinton JC, Hautsfort I, Eriksson S, Thompson A, Rhen M: Benefits and pitfalls of using microarrays to monitor bacterial gene expression during infection. *Curr Opin Microbiol* 2004, 7(3):277-282.
- Alland D, Kramnik I, Weisbrod TR, Otsubo L, Corry R, Miller LP, Jacobs WR Jr, Bloom BR: Identification of differentially expressed mRNA in prokaryotic organisms by customized amplification libraries (DECAL): the effect of isoniazid on gene expression in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 1998, 95(22):13227-13232.
- Graham JE, Clark-Curtis JE: Identification of *Mycobacterium tuberculosis* RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). *Proc Natl Acad Sci USA* 1999, 96(20):11554-11559.
- Daigle F, Graham JE, Curtis R: Identification of *Salmonella typhi* genes expressed within macrophages by selective capture of transcribed sequences (SCOTS). *Mol Microbiol* 2001, 41(5):1211-1222.
- Faucher SP, Porwellik S, Dezobis CM, McClelland M, Daigle F: Transcriptome of *Salmonella enterica* serovar Typhi within macrophages revealed through the selective capture of transcribed sequences. *Proc Natl Acad Sci USA* 2006, 103(6):1906-1911.
- Baltas N, Buetner FF, Gerlach GF: Selective capture of transcribed sequences (SCOTS) of *Actinobacillus pleuropneumoniae* in the chronic stage of disease reveals an HlyX-regulated autotransporter protein. *Vet Microbiol* 2007, 123(1-3):110-121.
- Baltas N, Gerlach GF: Identification of genes transcribed by *Actinobacillus pleuropneumoniae* in necrotic porcine lung tissue by using selective capture of transcribed sequences. *Infect Immun* 2004, 72(11):6711-6716.
- Procot A, Bezuidenhout J: The historical background and global importance of heartwater. *Onderstepoort J Vet Res* 1987, 54(3):165-169.
- Barre N, Uilenberg G, Moral PC, Camus E: Danger of introducing heartwater onto the American mainland: potential role of indigenous and exotic *Amblyomma* ticks. *Onderstepoort J Vet Res* 1987, 54(3):405-417.
- Uilenberg G: Experimental transmission of *Cowdria ruminantium* by the Gulf coast tick *Amblyomma maculatum*: danger of introducing heartwater and benign African theileriosis onto the American mainland. *Am J Vet Res* 1982, 43(7):1279-1282.
- Jongejan F, Wassink LA: Lack of cross-protection between *Cowdria ruminantium* and *Ehrlichia phagocytophila*. *Rev Elev Med Vet Pays Trop* 1991, 44(4):425-428.
- Mahan SM, Allsopp B, Kocan KM, Palmer GH, Jongejan F: Vaccine strategies for *Cowdria ruminantium* infections and their application to other ehrlichial infections. *Parasitol Today* 1999, 15(7):290-294.
- Raddy GR, Sulsona CR, Harrison RH, Mahan SM, Burridge MJ, Barbet AF: Sequence heterogeneity of the major antigenic protein 1 genes from *Cowdria ruminantium* isolates from different geographical areas. *Clin Diagn Lab Immunol* 1996, 3(4):417-422.
- Zweygath E, Josemans AJ, Van Strijp MF, Lopez-Rebolor L, Van Kleeff M, Allsopp BA: An attenuated *Ehrlichia ruminantium* (Wolgewonden stock) vaccine protects small ruminants against virulent heartwater challenge. *Vaccine* 2005, 23(14):1695-1702.
- Jongejan F, Zandbergen TA, Wiel PA van de, de Groot M, Uilenberg G: The tick-borne rickettsia *Cowdria ruminantium* has a *Chlamydia*-like developmental cycle. *Onderstepoort J Vet Res* 1991, 58(4):227-237.
- Frucos R, Viari A, Ferraz C, Bensaid A, Morgat A, Boyer F, Colucci E, Vachery N, Demalle J, Martinez D: Comparative genomics of three strains of *Ehrlichia ruminantium*: a review. *Ann NY Acad Sci* 2006, 1081:417-433.
- Frucos R, Viari A, Ferraz C, Morgat A, Eychenia S, Kandassamy Y, Charcal I, Bensaid A, Colucci E, Vachery N, Demalle J, Martinez D: Comparative genomic analysis of three strains of *Ehrlichia ruminantium* reveals an active process of genome size plasticity. *J Bacteriol* 2006, 188(7):2533-2542.
- Frucos R, Viari A, Vachery N, Boyer F, Martinez D: *Ehrlichia ruminantium*: genomic and evolutionary features. *Trends Parasitol* 2007, 23(9):414-419.
- Uilenberg G, Camus E, Barre N: [A strain of *Cowdria ruminantium* isolated in Guadeloupe (French West Indies)]. *Rev Elev Med Vet Pays Trop* 1985, 38(1):34-42.
- Bezuidenhout J, Paterson CL, Barnard B: In vitro cultivation of *Cowdria ruminantium*. *Onderstepoort J Vet Res* 1985, 52(2):113-120.
- Martinez D, Swinkels J, Camus E, Jongejan F: [Comparison between 3 antigens for the serodiagnosis of heartwater disease by indirect immunofluorescence]. *Rev Elev Med Vet Pays Trop* 1990, 43(2):159-166.
- Vachery N, Lefrancois T, Estèves I, Mollis S, Sheikboudou C, Kandassamy Y, Martinez D: Optimisation of the inactivated vaccine dose against heartwater and in vitro quantification of *Ehrlichia ruminantium* challenge material. *Vaccine* 2006, 24(22):4747-4756.
- Martinez D, Vachery N, Stachurski F, Kandassamy Y, Ralainina M, Aprelon R, Gueye A: Nested PCR for detection and genotyping of *Ehrlichia ruminantium*: use in genetic diversity analysis. *Ann NY Acad Sci* 2004, 1026:106-113.
- Martinez D, Mallard JC, Coisne S, Sheikboudou C, Bensaid A: Protection of goats against heartwater acquired by immunisation with inactivated elementary bodies of *Cowdria ruminantium*. *Vet Immunol Immunopathol* 1994, 41(1-2):153-163.
- Perez JM, Martinez D, Dobus A, Sheikboudou C, Bensaid A: Development of an in vitro cloning method for *Cowdria ruminantium*. *Clin Diagn Lab Immunol* 1997, 4(5):620-623.

27. Froussard P: A random-PCR method (rPCR) to construct whole cDNA library from low amounts of RNA. *Nucleic Acids Res* 1992, 20(11):2900.
28. Daigle F, Hou JY, Clark-Curtiss JE: Microbial gene expression elucidated by selective capture of transcribed sequences (SCOTS). *Methods Enzymol* 2002, 358:108-122.
29. Hahn WE, Pettijohn DE, Van Ness J: One strand equivalent of the *Escherichia coli* genome is transcribed: complexity and abundance classes of mRNA. *Science* 1977, 197(4303):582-585.
30. Ko MS, Ko SB, Takahashi N, Nishiguchi K, Abe K: Unbiased amplification of a highly complex mixture of DNA fragments by 'lone linker'-tagged PCR. *Nucleic Acids Res* 1990, 18(14):4293-4294.
31. Rouillard JM, Zukar M, Gubari E: OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res* 2003, 31(12):3057-3062.
32. Hubbard TJ, Aken BL, Ayling S, Ballestar B, Beal K, Bragin E, Brent S, Chen Y, Clapham P, Clark L, Coates G, Fairley S, Fitzgerald S, Fernandez-Banet J, Gordon L, Graf S, Haider S, Hammond M, Holland R, Howe K, Jenkinson A, Johnson N, Kahari A, Keefe D, Keenan S, Kinsella R, Kokodinski F, Kulaha E, Lawson D, Longden I, Megy K, Meidl P, Overduin B, Parker A, Pritchard B, Rios D, Schuster M, Slater G, Smedley D, Spooner W, Spudis G, Travanion S, Vilella A, Vogel J, White S, Wilder S, Zadissa A, Birney E, Cunningham F, Curwen V, Durbin R, Fernandez-Suarez XM, Harris J, Kasprzyk A, Proctor G, Smith J, Searle S, Flicek P: Ensembl 2009. *Nucleic Acids Res* 2009, D690-697.
33. Dridar G, Brezellec P, Remy E, Henaut A: GenoANOVA--gene expression analysis of variance. *Bioinformatics* 2002, 18(3):490-491.
34. Bekkar CP, Postigo M, Taoufik A, Bell-Salyt L, Ferraz C, Martinez D, Jongejans F: Transcription analysis of the major antigenic protein 1 multigene family of three in vitro-cultured *Ehrlichia ruminantium* isolates. *J Bacteriol* 2005, 187(14):4782-4791.
35. Alsbopp MT, Dorfling CM, Mallard JC, Bensaid A, Hayden DT, van Haerden H, Alsbopp BA: *Ehrlichia ruminantium* major antigenic protein gene (map1) variants are not geographically constrained and show no evidence of having evolved under positive selection pressure. *J Clin Microbiol* 2001, 39(11):4200-4203.
36. van Vliet AH, Jongejans F, van Kleef M, Zijlstra BA van der: Molecular cloning, sequence analysis, and expression of the gene encoding the immunodominant 32-kilodalton protein of *Cowdria ruminantium*. *Infect Immun* 1994, 62(4):1451-1456.
37. Savli H, Karadenizli A, Kobay F, Gundas S, Ozbek U, Vahaboglu H: Expression stability of six housekeeping genes: A proposal for resistance gene quantification studies of *Pseudomonas aeruginosa* by real-time quantitative RT-PCR. *J Med Microbiol* 2003, 52(Pt 5):403-408.
38. Takle GV, Toth IK, Brurberg MB: Evaluation of reference genes for real-time RT-PCR expression studies in the plant pathogen *Pectobacterium atrosepticum*. *BMC Plant Biol* 2007, 7:50.
39. Ranesto P, Rovery C, Schrenzel J, Leroy Q, Huyghe A, Li W, Lepidi H, Francois P, Raoult D: *Rickettsia conorii* transcriptional response within inoculation eschar. *PLoS ONE* 2008, 3(11):e3681.
40. Ge Y, Rikhsa Y: Identification of novel surface proteins of *Anaplasma phagocytophilum* by affinity purification and proteomics. *J Bacteriol* 2007, 189(21):7819-7828.
41. Wang X, Kiluchi T, Rikhsa Y: Proteomic identification of a novel *Anaplasma phagocytophilum* DNA binding protein that regulates a putative transcription factor. *J Bacteriol* 2007, 189(13):4880-4886.
42. Oganwa M, Ranesto P, Azza S, Molinar D, Fourquet P, Gorval JP, Raoult D: Proteome analysis of *Rickettsia felis* highlights the expression profile of intracellular bacteria. *Proteomics* 2007, 7(8):1232-1248.
43. Ranesto P, Azza S, Dolla A, Fourquet P, Vezris G, Gorval JP, Raoult D: Proteome analysis of *Rickettsia conorii* by two-dimensional gel electrophoresis coupled with mass spectrometry. *FEMS Microbiol Lett* 2005, 245(2):231-238.
44. Huang H, Lin M, Wang X, Kiluchi T, Mottaz H, Norbeck A, Rikhsa Y: Proteomic analysis of and immune responses to *Ehrlichia chaffeensis* lipoproteins. *Infect Immun* 2008, 76(8):3405-3414.
45. Jongejans F: Protective immunity to heartwater (*Cowdria ruminantium* infection) is acquired after vaccination with in vitro-attenuated rickettsiae. *Infect Immun* 1991, 59(2):729-731.
46. Martinez D: Analysis of the immune response of ruminants to *Cowdria ruminantium* infection. Utrecht: Utrecht University; 1997.
47. La MV, Francois P, Rovery C, Robineau S, Barbry P, Schrenzel J, Raoult D, Ranesto P: Development of a method for recovering rickettsial RNA from infected cells to analyze gene expression profiling of obligate intracellular bacteria. *J Microbiol Methods* 2007, 71(3):292-297.
48. Hou JY, Graham JE, Clark-Curtiss JE: *Mycobacterium avium* genes expressed during growth in human macrophages detected by selective capture of transcribed sequences (SCOTS). *Infect Immun* 2002, 70(7):3714-3726.
49. Faucher SP, Curtis R, Daigle F: Selective capture of *Salmonella enterica* serovar typhi genes expressed in macrophages that are absent from the *Salmonella enterica* serovar Typhimurium genome. *Infect Immun* 2005, 73(8):5217-5221.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)



#### 4.3. Résultats complémentaires : Hybridation de l'ADN génomique des souches Gardel et Sénégal (virulente et atténuée) sur les puces à ADN

Parallèlement à l'hybridation des génomes de la souche Gardel virulente et Welgevonden virulent sur les puces à ADN spécifique de ces deux souches, les génomes des souches Gardel atténuée, Sénégal virulent et atténuée ont été hybridées dans les mêmes conditions. Les résultats de l'analyse montrent une très forte colinéarité génomique entre Gardel virulente et atténuée ( $R^2 = 0,98$ ) et une bonne colinéarité entre Sénégal virulente et atténuée ( $R^2 = 0,91$ ) (Figure 16 et 17). L'hybridation des génomes de Gardel et de Sénégal qu'elles soient virulente ou atténuée, donne des coefficients de corrélation extrêmement faible avec des valeurs de  $R^2 = 0,47$  et  $R^2 = 0,50$  (Figure 18 et 19). La faible corrélation obtenue entre les deux souches Sénégal par rapport à la souche Gardel doit être minimisée car les lames ne sont pas spécifiques de la souche Sénégal. Par contre ces résultats montrent bien que la souche Sénégal présente bien des spécificités génomiques très différentes de Gardel. Une représentation des plus fortes intensités (heat map, Figure 20) chez les souches virulentes Gardel, Welgevonden et Sénégal montre les différences significatives existant entre ces génomes. Les résultats montrent la spécificité des lames pour chacun des CDS des souches Gardel et Welgevonden puisque les CDS spécifiques de chaque souche ne sont pas détectés par l'autre souche (Figure 20).

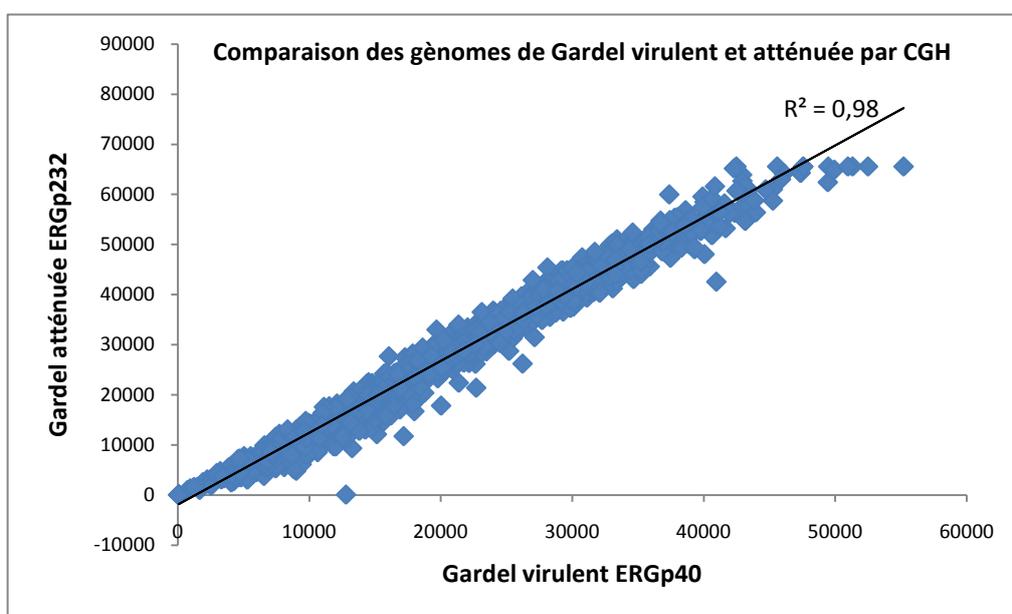


Figure 16 : Coefficient de corrélation entre les deux génomes Gardel virulente et atténuée

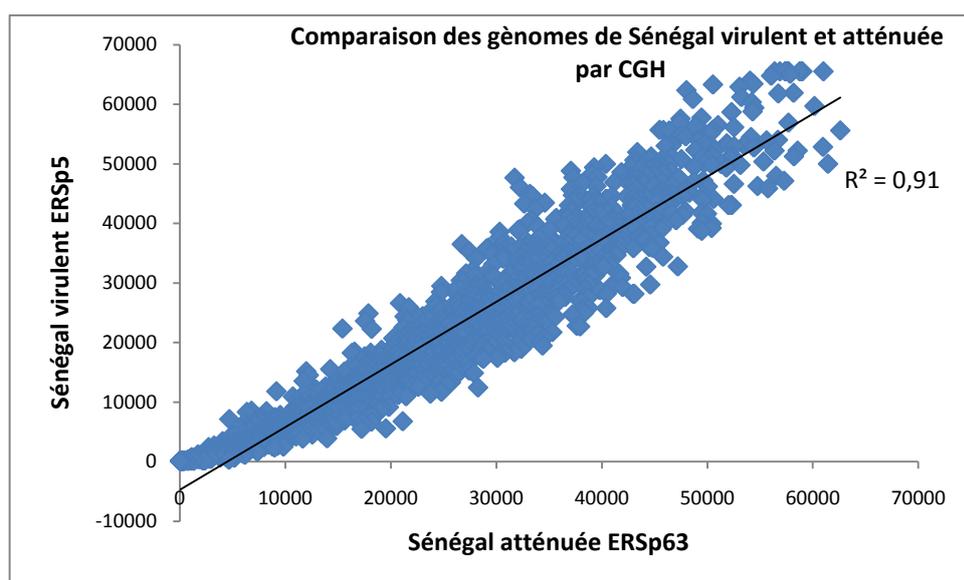


Figure 17 : Coefficient de corrélation entre l'hybridation génomes de Sénégal virulent et atténuée sur les puces pangénomiques d'*E. ruminantium* Gardel virulent

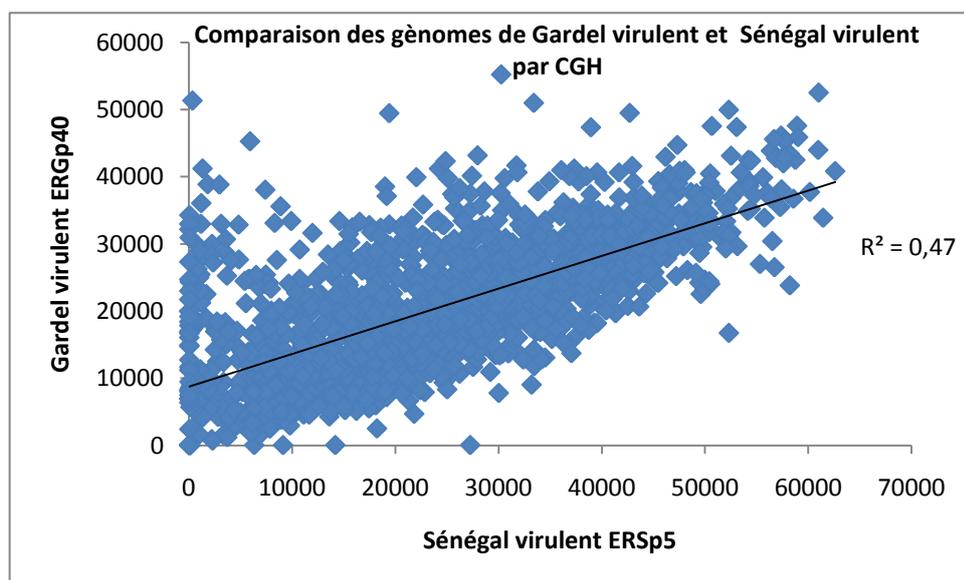


Figure 18 : Coefficient de corrélation entre les deux génomes Gardel virulent et Sénégal virulent

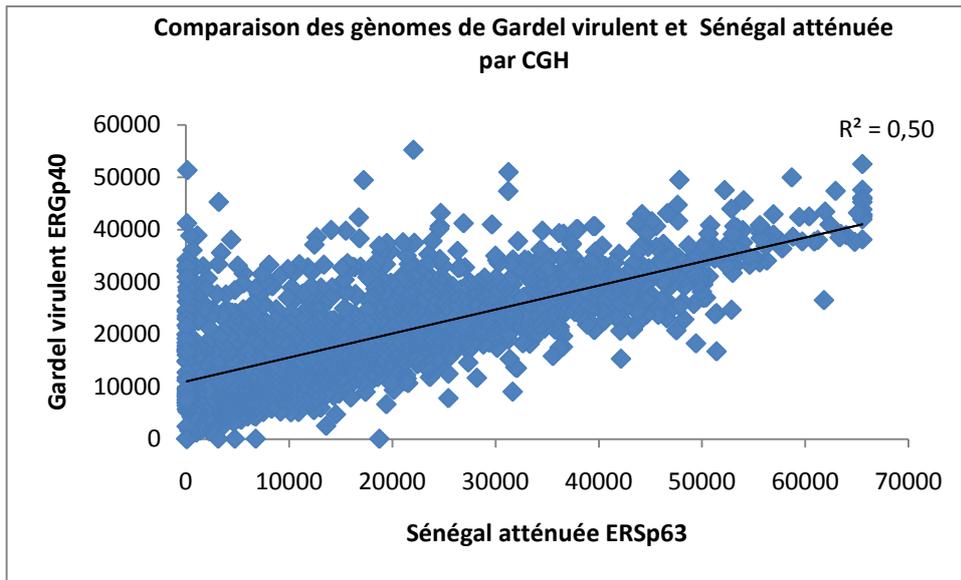


Figure 19 : Coefficient de corrélation entre les deux génomes Gardel virulent et Sénégal atténuée

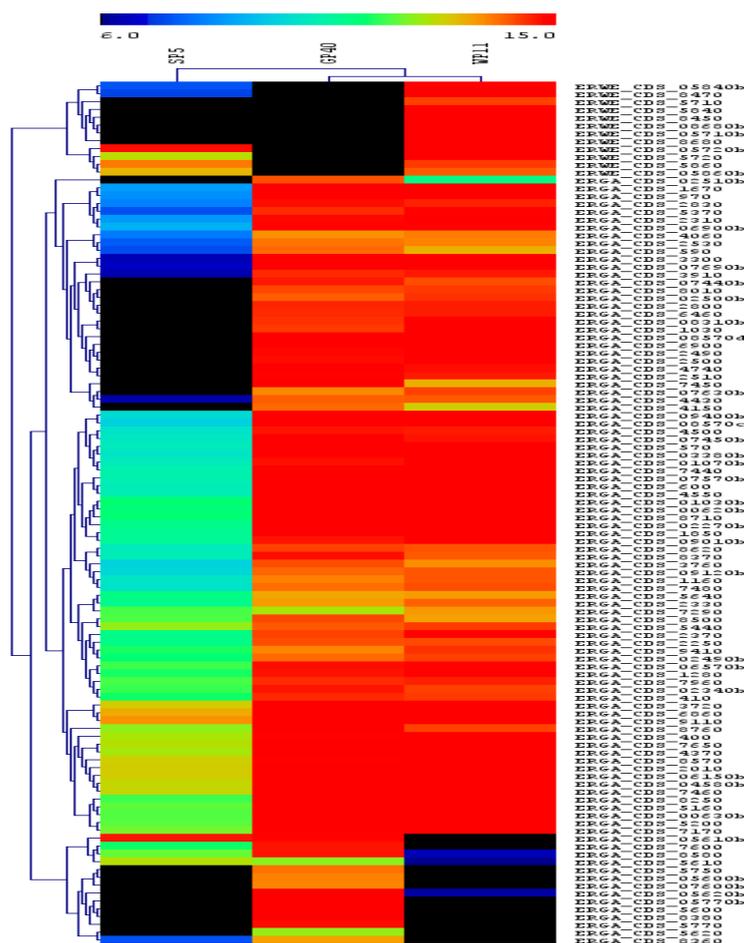


Figure 20 : Heat map : Les gènes les plus marquants permettant la différenciation des souches Sénégal (SP5), Gardel (GP40) et Welgevonden (WP11)

## 5. Conclusion

Les résultats présentés dans cet article ont permis de valider l'adaptation de la méthode SCOTS à *E. ruminantium* en utilisant *in fine* les puces à ADN spécifiques de la bactérie. Même si la méthode RiboMinus peut être largement améliorée elle semble limitée par la faible quantité d'ARNm bactérien récolté aux cours de chaque extraction pour les stades précoces particulièrement à 24 h post-infection. En effet la méthode SCOTS semble être indispensable pour analyser les gènes exprimés aux temps précoces si l'on veut comme dans notre cas obtenir un profil d'expression globale des souches à l'étude, c'est-à-dire à tous les temps de la cinétique. Cependant les phases tardives (de 72 hpi à 120 hpi) représentent un enjeu capital puisqu'il y a plusieurs évènements qui s'y produisent, notamment la formation

des *morulas*, la lyse cellulaire et la réinfection des cellules endothéliales. Par rapport à l'ensemble des méthodes disponibles pour analyser le transcriptome de bactérie pathogène, en particulier les intracellulaires, la méthode SCOTS semble être la plus appropriée. En effet la méthode de transcription *in vitro* couplée au RiboMinus présenté dans l'article donne des résultats encourageant particulièrement aux phases tardives (96 hpi) avec 20 % des CDS détectés ils ne sont pas suffisant par rapport à SCOTS qui est 3, 5 fois plus sensible. SCOTS est incontournable en particulier à la phase précoce du développement à 24 hpi où la détection est quasi nulle en utilisant la méthode IVT (RiboMinus). La plupart des techniques actuellement disponibles pour l'étude du transcriptome présente des désavantages notables et ne peuvent pas être efficacement adaptées à notre modèle expérimental. Une des méthodes présentée précédemment et qui serait applicable à la bactérie est la SSH. Cependant le séquençage et l'hybridation comparative des génomes des souches Gardel virulente et atténuée ont montré que les différences géniques significatives entre les deux génomes sont en faible nombre. De plus, la technique SSH n'est pas appropriée à la détection de petits fragments génétiquement variables, comme des insertions ou des variabilités nucléiques. La technique ne donne également aucune indication sur la localisation de ces séquences uniques sur le génome bactérien. Donc au cours d'une telle analyse transcriptomique nous nous attendons plus à voir un différentiel d'expression entre souche que la présence ou absence d'expression de gène. L'inconvénient majeur des deux méthodes (IVET et STM) est qu'elles ne sont pas réellement applicables aux bactéries non cultivables sur milieux conventionnels car elle nécessite une modification génétique de la bactérie étudiée. Or les *Rickettsiales* et d'autres bactéries intracellulaires sont très difficilement cultivables hors de la cellule hôte ce qui complique le développement de la mutagenèse dirigée chez ces bactéries intracellulaires obligatoires et explique le très faible nombre de travaux sur ces méthodes (Binet and Maurelli, 2009; Driskell *et al.*, 2009). La plupart des travaux rapportés sur les *Rickettsiales*, concernent l'utilisation de transposons allant s'insérer de façon aléatoire dans le génome, ce qui ne permet pas de cibler et d'étudier spécifiquement le rôle de gènes d'intérêt chez la bactérie (Liu *et al.*, 2007; Qin *et al.*, 2004).

Bien que la corrélation entre les rondes successives de SCOTS soit bonne, les analyses des sondes ont révélé un biais, matérialisé par la très faible colinéarité et corrélation des sondes de

chaque CDS à 3x (en moyenne 2 sondes par gènes) comparées aux autres rondes (1x et 2x) et même à 0x. Dans notre cas, les échantillons SCOTS 3x ne peuvent pas servir pour des mesures de niveau d'expression de gènes entre souches pour différentes conditions, si les 2 sondes pour un même CDS ne sont pas similaires. Les échantillons SCOTS 2x hybridés pour les deux souches Gardel virulente et atténuée sont meilleurs, puisque les réplicats de chaque sonde semblent être de meilleures qualités avec un niveau de saturation moins important. Le biais observé provient plus de la reproductibilité de l'allumage des sondes que de l'enrichissement par SCOTS, puisque nos travaux ont montré que SCOTS n'induisait pas de biais important dans le niveau d'expression des gènes. Ainsi ce travail réalisé par l'intermédiaire de la méthode SCOTS permettra l'analyse du transcriptome d'*E. ruminantium* mais aussi d'autres bactéries intracellulaires obligatoires.

## **CHAPITRE III.**

**Analyse du transcriptome de la bactérie *Ehrlichia ruminantium***

## 1. Contexte

Les mécanismes moléculaires qui régissent l'interaction spécifique existant entre la bactérie et leurs cellules cibles chez l'hôte sont encore mal connus. De plus, les mécanismes liés à la pathogenèse et l'atténuation des souches d'*E. ruminantium* n'ont pas été élucidés jusqu'à présent. Ces limites sont en partie dues au nombre restreint de méthodes efficaces disponibles pour étudier l'ensemble des transcrits exprimés par ces bactéries intracellulaires obligatoires. Afin de mieux comprendre les événements moléculaires qui régulent l'infection des cellules endothéliales hôtes par la bactérie *E. ruminantium*, la méthode SCOTS permettant d'obtenir des transcrits bactériens débarrassés entre autres de leurs contaminants eucaryotes a été combinée aux microarrays pour analyser et comparer l'ensemble des transcrits générés par les souches Gardel virulente et atténuée à plusieurs moments du cycle d'infection (entre la phase précoce de l'infection et la phase de lyse cellulaire). Les gènes différentiellement exprimés par ces deux souches (induits ou réprimés) ont été identifiés. La fonction de ces gènes a ensuite été déterminée par la génomique comparée lorsque la fonction était identifiée pour d'autres pathogènes, ou par la recherche de motifs protéiques connus, pour les gènes de fonctions inconnues. A partir des différentiels observés, nous avons obtenu un véritable profil d'expression global pour la bactérie *E. ruminantium*, avec l'identification de nombreux gènes ou systèmes de régulation impliqués dans la pathogenèse de la bactérie. Les gènes les plus marquants seront présentés en fonction de leurs catégories et leurs fonctions. Une validation du différentiel d'expression obtenu par microarrays a été initiée par qRT-PCR.

## 2. Matériels et méthodes

### 2.1. Production des échantillons biologiques *in vitro*: souches Gardel virulente et Gardel atténuée

Le protocole de production des échantillons biologiques et les conditions de culture sont détaillés dans le chapitre II. A chaque collecte, 1/10 du surnageant a été conservé pour une extraction d'ADN total suivi d'une quantification du nombre de copie d'*E. ruminantium* par

qPCR, utilisant une sonde taqman spécifique du gène *map1* de la bactérie comme décrit dans le chapitre II.

#### 2.2. Capture sélective des transcrits d'*Ehrlichia ruminantium* par SCOTS

L'ARN des échantillons collectés aux différents temps a été extrait selon le protocole d'utilisation du TRIzol (Invitrogen). Les échantillons d'ARN ont été traités avec la DNase sans RNase- (Ambion). Chaque échantillon d'ARN a été converti en ADNc et la méthode SCOTS a été appliquée comme décrit précédemment dans le chapitre II (Emboule *et al.*, 2009). Deux captures successives (rondes) de SCOTS ont été générées pour obtenir les échantillons SCOTS 2x utilisés pour les analyses microarrays.

#### 2.3. Puces à ADN pangénomique d'*Ehrlichia ruminantium*: description et analyse statistique

La description des puces à ADN couvrant l'ensemble du génome de la bactérie *E. ruminantium* pour les souches Gardel et Welgevonden, les conditions de marquage et d'hybridation des ADNc, l'acquisition des données ainsi que la normalisation des résultats sont décrits précédemment dans l'article présentant l'adaptation de la méthode SCOTS à notre modèle expérimental (Emboule *et al.*, 2009). En plus de l'ADNc correspondant aux échantillons sans capture et après deux captures (SCOTS 0x ou 2x) de chaque condition testée, l'ADN génomique des souches Gardel et Welgevonden a été hybridé comme contrôle positif d'hybridation.

#### 2.4. Acquisition et analyse de l'image

Brièvement, la lecture des fluorescences est réalisée avec une résolution de 10 µm à l'aide du scanner GenePix 4000B (Axon Instruments). Pour chaque lame, le gain optimal d'acquisition des images est déterminé en fonction de l'intensité du signal et du bruit de fond ambiant, afin d'obtenir une dynamique maximale du signal. Le logiciel GenePix Pro (Axon Instruments) permet l'acquisition des images puis leur analyse par un processus qui s'effectue en trois étapes : la localisation des spots par application sur l'image d'une grille, la segmentation et l'extraction des données. Les images en format .tif ainsi que les fichiers de quantifications sont ensuite stockés sur le système d'information du laboratoire Mediante (Le Brigand and Barbry, 2007) permettant de récupérer facilement les données pour l'analyse statistique.

## 2.5. Analyse statistique des résultats

Deux approches statistiques ont été réalisées : i) la première à l'aide du package d'analyse limmaGUI disponible sous le logiciel "R" (<http://cran.rproject.org/>); une normalisation inter-lame a été effectuée par la méthode quantile qui permet d'homogénéiser la distribution des intensités entre les lames, l'analyse statistique des données est ensuite effectuée et permet d'obtenir une liste de gènes différentiellement exprimés à l'aide d'un score statistique dérivé d'un modèle linéaire utilisant une approche empirique bayésienne ; ii) une analyse de variance effectuée sur les valeurs d'intensité normalisées (quantile) puis transformées en *logarithme de base 2* en utilisant le logiciel GeneANOVA (Didier *et al.*, 2002).

## 2.6. Annotation des CDS différentiellement exprimés

La fonction des gènes identifiés comme exprimés de façon différentielle entre la souche virulente et atténuée a été recherchée à l'aide de la banque de données NCBI (Blastp). Pour les gènes de fonctions inconnues, la banque de données Swiss-Prot (ExPASy Proteomics server) a été utilisée principalement pour identifier des domaines protéiques connus. La séquence nucléotidique des gènes contigus ou avoisinants *recB* sur le génome de la souche virulente et atténuée a été particulièrement étudiée pour déterminer s'il y avait des événements de mutation (insertion/délétion) sur ces gènes, permettant de différencier la souche virulente et la souche atténuée de Gardel.

## 2.7. Validation des gènes d'intérêt par qRT-PCR

Des CDS (Coding Sequences ou séquences codantes) différentiellement surexprimés et sous-exprimés par la souche virulente aux différents temps mis en évidence par microarrays sur les échantillons capturés après deux rondes (SCOTS2x) ont été choisis pour la validation par qPCR. De même, des gènes ont été choisis à partir des analyses microarrays sur les échantillons sans capture (SCOTS0x) collectés au temps tardif. La validation du différentiel d'expression entre souches obtenu par microarrays a été effectuée sur les replicats biologiques ERGp232 et ERGp29. La conversion en cDNA pour ces deux souches a été réalisée en utilisant des amorces pdN<sub>6</sub> et une reverse transcriptase suivant les recommandations du

fabricant (Kit First-Strand cDNA Synthesis Kit Kit Amersham Biosciences). Les échantillons de cDNA de la souche virulente ERGp39 et de la souche atténuée ERGp243 convertis en utilisant l'étiquette NkpNI ont aussi été testés en qRT-PCR pour la validation des gènes. L'amplification des gènes d'intérêt a été réalisée par qPCR en utilisant des amorces spécifiques et du Sybergreen (Applied Biosystems). La liste des gènes ainsi que toutes les amorces dérivées se trouvent dans le tableau en annexe (1 et 2). La méthode d'analyse relative d'expression utilisant la méthode des  $\Delta\Delta Ct$  a été utilisée.  $\Delta Ct = Ct$  gène cible -  $Ct$  gène endogène.  $\Delta\Delta Ct = \Delta Ct_{\text{virulent}} - \Delta Ct_{\text{atténué}}$  avec  $RQ = 2^{-\Delta\Delta Ct}$ . RQ correspond au nombre de fois que le gène est exprimé pour la souche virulente par rapport à la souche atténuée. Le calibrateur choisi est la souche atténuée ainsi les ratios positifs sont représentatifs de la surexpression d'un transcrite de la souche virulente par rapport à celui de la souche atténuée et les ratios négatifs de la sous expression de cette même souche.

La normalisation se fait en utilisant 3 gènes endogènes différents, ARN *16S*, *recA* et *pCS20* pour lesquels la stabilité a été évaluée dans cette étude pour la souche Gardel (Chapitre II).

### 3. Résultats

#### 3.1. Comparaison des cinétiques de croissance bactérienne pour l'analyse du transcriptome

La quantification du nombre de bactéries aux cours du temps et l'observation des différentes phases évolutives au microscope tout au long de l'infection ont permis de choisir les temps à comparer entre la souche virulente et atténuée (Chapitre II). Quatre points de cinétique ont été étudiés pour la souche virulente et atténuée 24 hpi, 72 hpi, 96 hpi et 120 hpi et 24 hpi, 48 hpi, 72 hpi et 96 hpi respectivement (Tableau 10). D'après les résultats de la quantification bactérienne à 24 hpi (Figure 12), les deux souches se situent pendant ou à la fin de la phase de latence, ce que l'on appellera le temps de comparaison 1 (T1) (Tableau 10). A 72 h post-infection pour la souche virulente et à 48 hpi pour la souche atténuée, les bactéries se situent dans la phase exponentielle de croissance, ce qu'on appellera le temps de comparaison 2 (T2) (Figure 12). Les échantillons correspondant à 96 hpi et 72 hpi pour la souche virulente et atténuée ont également été hybridés sur lame mais nous n'aborderons pas les résultats issus de cette comparaison dans cette partie. En effet, il semble qu'à ce stade de développement, il y ait présence de différents stades morphologiques d'*E. ruminantium* ce qui pourrait compliquer l'interprétation des résultats. Enfin les temps 120 hpi et 96 hpi correspondent à la lyse des

## Analyse du transcriptome de la bactérie *Ehrlichia ruminantium*

cellules endothéliales et à l'obtention de corps élémentaires pour la souche virulente et la souche atténuée respectivement. A ce stade, les bactéries se situent également toutes les deux dans la phase stationnaire du cycle de croissance (T3) (Figure 12). De plus, ce stade correspond à la lyse cellulaire pour les deux souches donc on peut supposer que l'état physiologique des bactéries est le même et qu'elles sont présentes uniquement sous forme de corps élémentaires.

Les analyses croisées entre différentes phases de développements différentes restent toujours possibles, les échantillons de chaque temps étant hybridés séparément.

Tableau 10 : Cinétique et comparaison des différents stades de développement

	Gardel virulente	Gardel atténuée	Stade	Observations microscope
T1	24 hpi	24 hpi	Phase de latence	Condensation
T2	72 hpi	48 hpi	Croissance exponentielle	Petites <i>morulas</i>
T3'	96 hpi	72 hpi	Phase plateau	Grosses <i>morulas</i> & CE
T3	120 hpi	96 hpi	Phase stationnaire	CE

CE=corps élémentaires

### 3.2. Analyse du transcriptome d'*E. ruminantium* par microarrays

#### 3.2.1. Analyse globale et répartition des catégories de gènes

Globalement à partir de l'analyse des puces à ADN hybridées avec les transcrits SCOTS 2x, la comparaison du niveau d'expression des gènes entre Gardel virulente et atténuée d'*E. ruminantium* a révélé 239 transcrits, soit 25% (239/950) des CDS différentiellement exprimés à tous les temps analysés (T1, T2 et T3) ( $p < 0,02$ ). Ces gènes se répartissent selon deux groupes distincts: les gènes surexprimés et les gènes sous-exprimés chez la souche virulente ERGp39 par rapport à la souche atténuée ERGp243, pour chaque point de la cinétique d'infection. Parmi ces gènes, 22% (53/239) sont différentiellement exprimés à T1 (24 hpi), 29% (70/239) à T2 pendant la phase de croissance exponentielle et 49% (116/239) à T3 pendant la phase de lyse complète des cellules. La majorité des gènes différentiellement exprimés le sont à T3. A T1, 55% (29/53) sont surexprimés par la souche virulente et 45%

### Analyse du transcriptome de la bactérie *Ehrlichia ruminantium*

(24/53) sont sous exprimés par rapport à la souche atténuée (Tableau 11). Aux temps T2 et T3, la proportion de gènes surexprimés et sous exprimés est similaire au temps T1 : 54% vs 46% et 56% vs 44% respectivement. De plus l'analyse statistique croisée des données (R et Excel) a permis d'identifier 3 gènes différentiellement exprimés et communs entre les phases T1 et T2, 1 entre T1 et T3 et 15 entre T2 et T3 (soit 13% des gènes différentiellement exprimés à T3 au moment de la lyse). Les autres semblent être spécifiquement surexprimés ou sous exprimés par l'une des deux souches. En conséquence le nombre de gènes spécifiquement exprimés pour chaque phase étaient de 49 à T1 (24 hpi), 52 à T2 (48 hpi-72 hpi) et 100 à T3 (96 hpi-120 hpi).

Plus de 55% (132/239) des gènes différentiellement exprimés détectés entre les deux souches sont des éléments surexprimés par la souche virulente ERGp39 par rapport à la souche atténuée ERGp243 (45% sous-exprimés (107/239)) (Tableau 11).

Tableau 11 : Répartition des gènes différentiellement exprimés en fonction des stades de développement entre ERGp39 et ERGp243

	<b>T1</b>	<b>T2</b>	<b>T3</b>	<b>Global</b>
<b>Gènes différentiellement exprimés</b>	22 % (53/239)	29 % (70/239)	49 % (116/239)	100%
<b>Gènes surexprimés pour Gardel virulente</b>	55 % (29/53)	54 % (38/70)	56 % (65/116)	55% 132/239
<b>Gènes sous exprimés pour Gardel virulente</b>	45 % (24/53)	46 % (32/70)	44 % (51/116)	45% 107/239

En utilisant les données obtenues suite à l'annotation du génome d'*E. ruminantium* souche Gardel NC\_006831, nous avons pu identifier un grand nombre de gènes à chaque stade de développement, dont la fonction avait déjà été décrite chez d'autres espèces bactériennes. Parmi les 239 CDS différentiellement exprimés, 68% (163/239) d'entre eux présentent des annotations fonctionnelles (Figure 21A). Les autres CDS sont de fonction inconnue ou hypothétique. Les gènes servant à la réplication et à la réparation de l'ADN, la biogénèse des

ribosomes et la traduction, la modification post-transductionnelle et la production d'énergie sont aussi largement représentés (Figure 21A). Chacune de ces catégories de gènes contient une nette majorité de gènes surexprimés par la souche virulente par rapport à ceux surexprimés par la souche atténuée (sous-expression pour la virulente), avec un net différentiel dans l'expression des gènes intervenants dans la production d'énergie et la modification post-transductionnelle, la réplication et la réparation de l'ADN (figure 21B). Parmi les gènes surexprimés pour la souche virulente par rapport à la souche atténuée, certains sont déjà connus chez d'autres pathogènes intracellulaires (*Rickettsia conorii*, *Coxiella burnetii*) comme étant impliqués dans l'échappement à la réponse de l'hôte : protéines impliquées dans la réparation de l'ADN (helicase, ligase), dans l'échappement au stress oxydatif (thioredoxine), au choc osmotique ainsi que dans l'inhibition de la fusion des endosomes et dans le ralentissement de l'apoptose (Ankyrine A). Parmi les gènes sous exprimés chez la souche virulente par rapport à la souche atténuée (surexprimés pour la souche atténuée) on identifie beaucoup de gènes intervenant dans le transport et le métabolisme de nombreux éléments (phosphate, lipide, nucléotide, carbohydrate, acide aminés, coenzyme). On retrouve également plusieurs gènes appartenant à la famille multigénique *map* soit 8 CDS sur 16, codant pour des protéines de la membrane externe d'*E. ruminantium* parmi lesquels plusieurs sont surexprimés deux fois exclusivement entre les temps T2 et T3.

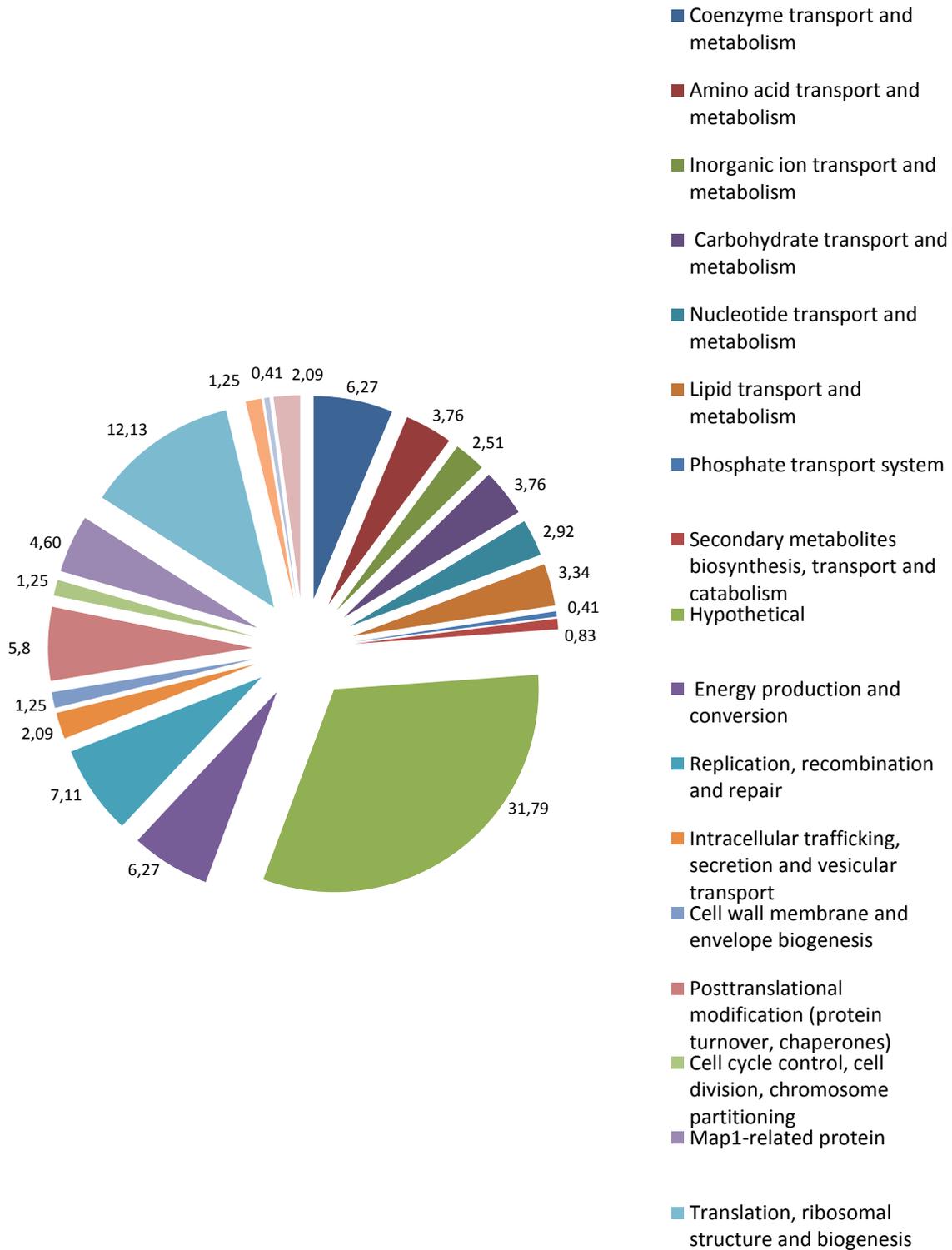


Figure 21 A : Pourcentages des catégories des 239 gènes différentiellement exprimés de T1 à T3 entre la souche Gardel virulente (ERGp39) et la souche Gardel atténuée (ERGp243)

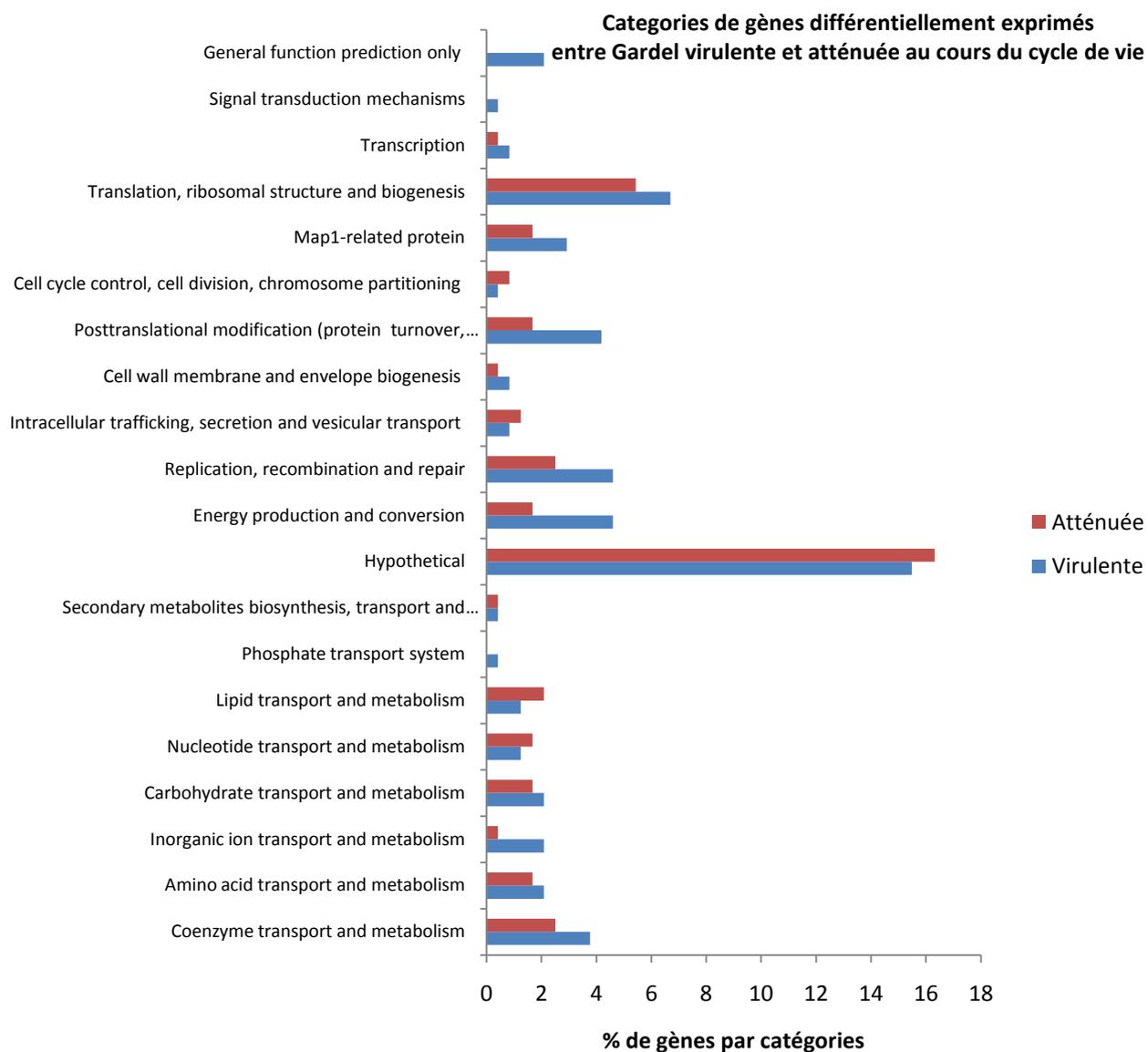


Figure 21 B : Gènes différentiellement exprimés entre Gardel atténuée et virulente : répartition selon leur fonction connue ou supposée

3.2.2. Identification des gènes différentiellement exprimés par *Ehrlichia ruminantium* au cours des différentes phases du cycle de développement

3.2.2.1. Gènes surexprimés par la souche Gardel virulente à T1 (24 hpi vs 24 hpi)

Au temps T1, les deux catégories de gènes les plus représentés sont les gènes intervenants dans (i) la traduction, la structure et la biogénèse des ribosomes (>28 %), et (ii) les gènes de fonctions inconnues (21%). On retrouve également des gènes impliqués dans la production d'énergie (6%), la réplication (>7%), la transcription (2%), régulation post-traductionnelle (>7%), métabolisme et le transport (23%) avec plusieurs catégories, le trafic intracellulaire et sécrétion (4%) et la division cellulaire (2%) (Figure 22).

Le gène le plus différentiellement exprimé à T1 entre la souche virulente et atténuée est le gène codant pour l'enzyme peptide deformylase (*def* ou PDF) (Tableau 12). Il est 15 fois plus exprimé chez Gardel virulente (FC=7.39 normalisé log2). L'enzyme PDF est essentielle chez les bactéries car elle fait partie de la voie de l'excision de la méthionine Nterminale qui touche la plupart des protéines nouvellement synthétisées. La PDF est responsable du clivage du groupement formyl porté par cette première méthionine (Chan *et al.*, 2003; Mazel *et al.*, 1994). Aucune mutation n'est détectée dans ce gène chez la souche atténuée.

A T1, parmi les gènes différentiellement exprimés appartenant à la catégorie de gènes impliqués dans la traduction, la structure et la biogénèse des ribosomes on identifie une large majorité de gènes surexprimés par la souche virulente. Le gène *infC* codant pour le facteur de démarrage de la traduction IF3 fait partie des 5 gènes les plus surexprimés par la souche virulente (Fold change FC= 5,52). Des gènes codants pour des protéines constitutives de la petite sous-unité du ribosome (30S), impliqué dans la lecture de l'ARNm font également partis des gènes les plus surexprimés par la souche virulente notamment *rpsA* (FC=5,62), les gènes *rpsB* et *rpsG* étant sous exprimés (Tableau 13). Sept gènes codant pour des protéines ou sous unités constitutives de la grande sous unité du ribosome (50S) qui est le catalyseur direct dans la synthèse protéique sont surexprimés par la souche virulente (FC= 3 à 7) (Tableau 12). Cette grande sous unité (50S) chez les procaryotes comporte plus de 30 protéines différentes. Les gènes *rpmF*, *rplR*, *rplM*, *rplT*, *rpmA*, *rplU*, *rplQ* codent chacun pour une sous unité (Tableau 12). En tout 8 gènes codant pour des protéines ribosomiques sont surexprimés par la souche virulente alors que seuls 3 sont surexprimés par la souche atténuée (Tableau 13).

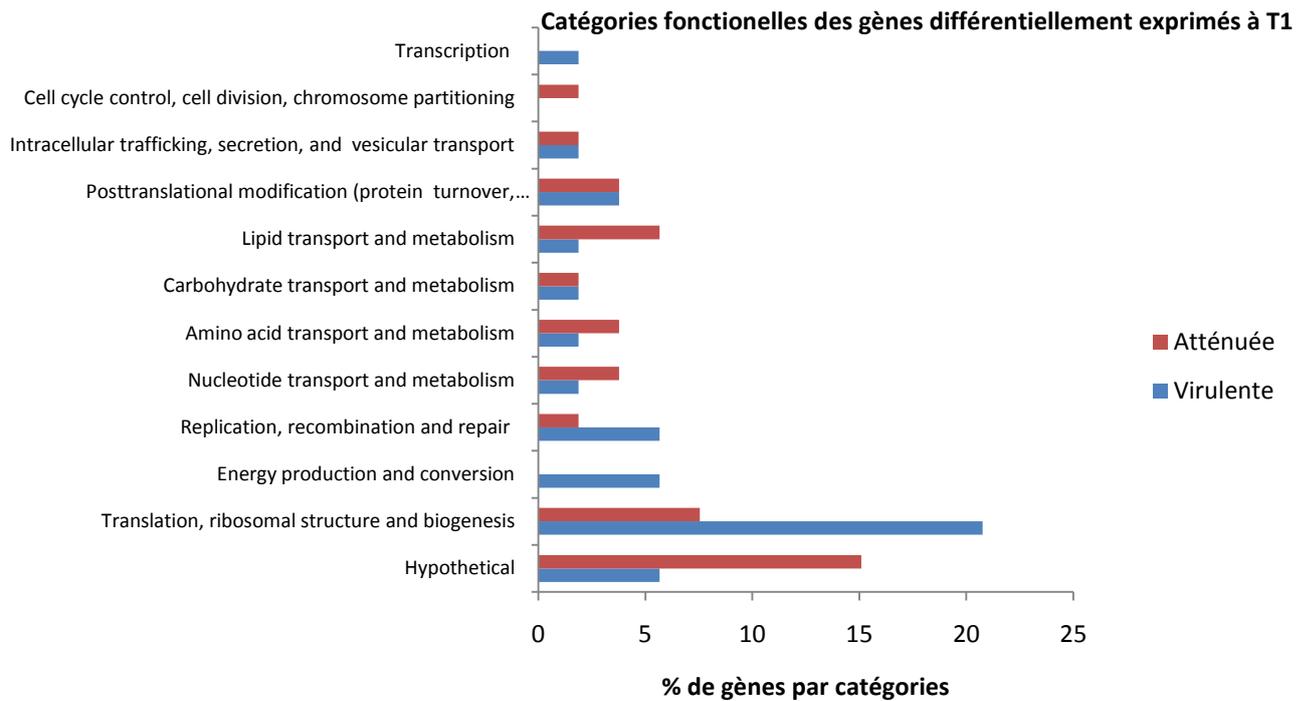


Figure 22 : Répartition et fonction des gènes différentiellement exprimés entre Gardel atténuée et virulente à 24 hpi

Tous les gènes différentiellement exprimés qui sont impliqués dans la production d'énergie à 24 hpi sont surexprimés par Gardel virulent (Figure 22). Parmi ces gènes, on retrouve le gène *sucC* codant pour la chaîne beta de la succinyl-CoA synthetase (Tableau 12). Chez les bactéries Gram négatif, l'holoenzyme, SucCD est composé de *sucC* ( $\beta$ -chaîne) et *sucD* ( $\alpha$ -chaîne). Cette enzyme bactérienne est une transférase qui catalyse la phosphorylation de l'ADP en ATP durant le cycle de Krebs (cycle de l'acide citrique), réaction qui s'accompagne du clivage de la succinyl-CoA en succinate (Mercado-Lubo *et al.*, 2008). On retrouve également le gène *sdhC* codant pour la succinate déshydrogénase. Chez la bactérie *E. coli*, la succinate déshydrogénase (SdhCDAB) catalyse l'oxydation du succinate en fumarate dans le cycle de Krebs et produit des radicaux superoxydes. Cette protéine est également différentiellement exprimée à T2 mais elle est sous-exprimée par la souche virulente à ce stade.

*Analyse du transcriptome de la bactérie Ehrlichia ruminantium*

Tableau 12 : Gènes surexprimés par la souche virulente ERGp39 à 24 hpi

<b>CDS</b>	<b>Fold ch</b>	<b>P-value</b>	<b>Genes</b>	<b>Descriptions</b>
ERGA_CDS_00440	7,39	8,38E-10	def	Peptide deformylase
ERGA_CDS_07790	7,23	1,28E-07	rpmE	50S ribosomal protein L31
ERGA_CDS_05940	6,98	9,44E-08	rpmF	50S ribosomal protein L32
ERGA_CDS_05660	6,91	6,30E-09	unknown	hypothetical polyketide cylcases/dehydrases
ERGA_CDS_07930	6,73	2,56E-09	gltX1	Glutamyl-tRNA synthetase 1
ERGA_CDS_07980	6,38	2,52E-08	unknown	hypothetical protein/ NifU-like domain
ERGA_CDS_01010	6,14	2,50E-08	ihfA	Integration host factor alpha-subunit
ERGA_CDS_02990	6,10	1,44E-06	rpoZ	DNA-directed RNA polymerase omega chain
ERGA_CDS_01840	5,72	2,98E-09	sdhC	Succinate dehydrogenase cytochrome B-556
ERGA_CDS_06340	5,63	9,08E-07	rpsA	30S ribosomal protein S1
ERGA_CDS_09330	5,53	3,24E-05	infC	Translation initiation factor IF-3
ERGA_CDS_08050	5,47	4,46E-05	metG	Methionyl-tRNA synthetase
ERGA_CDS_08180	5,38	1,16E-07	argC	N-acetyl-gamma-glutamyl-phosphate reductase
ERGA_CDS_00990	5,07	1,52E-03	purE	Phosphoribosylaminoimidazole carboxylase catalytic subunit
ERGA_CDS_04960	4,91	4,18E-08	eno	Enolase (2-phosphoglycerate dehydratase)
ERGA_CDS_04380	4,89	2,26E-06	gyrB	DNA gyrase subunit B
ERGA_CDS_06130	4,44	1,85E-05	rplR	50S ribosomal protein L18
ERGA_CDS_08160	4,08	8,48E-03	rplM	50S ribosomal protein L13
ERGA_CDS_01290	3,93	8,43E-05	rplT	50S ribosomal protein L20
ERGA_CDS_04940	3,53	1,93E-03	rpmA	50S ribosomal protein L27
ERGA_CDS_02840	3,46	1,68E-04	ssb	Single-strand binding protein
ERGA_CDS_04950	3,29	5,77E-04	rplU	50S ribosomal protein L21
ERGA_CDS_01440	3,25	3,36E-04	sucC	Succinyl-CoA synthetase beta chain
ERGA_CDS_05770	3,15	6,68E-07	unknown	Hypothetical protein
ERGA_CDS_06540	3,01	1,07E-02	unknown	Hypothetical protein
ERGA_CDS_07850	2,99	4,22E-03	trx	Thioredoxin
ERGA_CDS_05400	2,83	1,73E-04	virB4	VIRB4 protein precursor
ERGA_CDS_08900	2,57	1,20E-02	pyrE	Orotate phosphoribosyltransferase
ERGA_CDS_08000	1,83	2,74E-03	rho	Transcription termination factor rho

Tableau 13 : Gènes sous-exprimés par la souche virulente ERGp39 à 24 hpi

CDS	Fold ch	P-value	Genes	Descriptions
ERGA_CDS_00640	-7,31	5,57E-09	unknown	Hypothetical protein
ERGA_CDS_05930	-7,04	2,97E-09	plsX	phospholipid synthesis protein
ERGA_CDS_06620	-6,31	7,16E-07	clpB	ClpB protein
ERGA_CDS_02060	-6,26	5,45E-08	ftsK	DNA translocase
ERGA_CDS_02870	-6,15	1,88E-07	fabI	Putative Enoyl-[acyl-carrier-protein] reductase [NADH]
ERGA_CDS_03040	-5,87	9,40E-05	pyrF	Orotidine 5'-phosphate decarboxylase
ERGA_CDS_08410	-5,29	2,14E-06	hflK	Protease activity modulator hflK
ERGA_CDS_02560	-5,25	7,86E-07	recR	Recombination protein RecR
ERGA_CDS_04240	-4,80	8,52E-06	iscS	Cysteine desulfurase (NifS protein homolog)
ERGA_CDS_05390	-4,12	5,30E-06	unknown	Conserved hypothetical protein
ERGA_CDS_03830	-3,94	2,31E-06	unknown	Conserved hypothetical protein
ERGA_CDS_06050	-3,77	1,31E-05	rplQ	50S ribosomal protein L17
ERGA_CDS_07740	-3,44	3,74E-04	secB	Protein-export protein secB
ERGA_CDS_05240	-3,19	2,04E-03	rpsB	30S ribosomal protein S2
ERGA_CDS_06880	-3,10	1,47E-02	unknown	Conserved hypothetical protein
ERGA_CDS_09480	-3,05	7,73E-03	glyQ	Glycyl-tRNA synthetase alpha chain
ERGA_CDS_01080	-2,96	3,88E-04	unknown	Hypothetical protein
ERGA_CDS_04570	-2,88	2,99E-03	pccB	Similar to human Propionyl-CoA carboxylase beta chain
ERGA_CDS_04230	-2,80	1,02E-02	unknown	Conserved hypothetical protein
ERGA_CDS_04120	-2,31	1,32E-03	tpiA	Triosephosphate isomerase
ERGA_CDS_01560	-1,80	2,22E-03	rpsG	30S ribosomal protein S7
ERGA_CDS_07350	-1,63	5,32E-04	unknown	Hypothetical protein
ERGA_CDS_06600	-1,54	8,68E-03	pepA	Probable cytosol aminopeptidase A
ERGA_CDS_04310	-1,05	3,82E-03	unknown	Hypothetical protein

Quatre gènes présents dans le génome d'*E. ruminantium* et ayant un rôle primordial dans la réplication ou la réparation de l'ADN ont été identifiés comme différentiellement exprimés. Parmi ces gènes, 3 sont surexprimés par la souche virulente à 24 hpi : *ihfA* (FC=6,14), *gyrB* (FC=4,88) et *ssB* (FC=3,45) (Tableau 12). *IhfA* qui code pour une des sous unités alpha de la protéine hétérodimérique IHF (Integration Host Factor) fait partie de la famille des protéines de type histone et aurait la capacité d'induire une courbure de l'ADN qui faciliterait les interactions protéine-protéine entre l'ARN polymérase et un activateur proximal. Elle régule de nombreux gènes en réponse à un stress. D'autre part, l'influence d'IHF sur la structure de l'ADN contribue notamment au contrôle de la transcription (Goosen and van de Putte, 1995). Le gène *gyrB* code pour la sous unité beta de l'ADN gyrase, topoisomérase qui sert à

désenrouler l'ADN durant la réplication. Enfin *ssb* (single stranded binding) code pour les protéines SSB. Ces protéines vont se fixer à l'ADN simple brin et vont les maintenir en état monocaténaire.

Le gène *rho* est le seul de la catégorie des gènes codant pour la transcription qui a été identifié comme surexprimé pour la souche virulente (FC=1,83) (Tableau 12). *Rho* code pour une hélicase ATP-dépendante à l'origine de la terminaison de la transcription chez les procaryotes.

On détecte à 24 hpi d'autres catégories de gènes surexprimés par la souche virulente notamment ceux intervenant dans la régulation post-traductionnelle ou encore le trafic intracellulaire et la sécrétion (Figure 22). Le gène *trx* est un régulateur de la réponse au stress oxydatif qui code pour une thiorédoxine (FC=2,98) (Tableau 12) (Zeller and Klug, 2006). Le gène *virB4* qui code pour le précurseur de cette même protéine qui est essentielle pour la virulence de nombreuses bactéries intracellulaires (FC=2,84). Ce gène appartient au système de sécrétion de type IV et est bien décrit chez un très grand nombre de bactéries.

La majorité des gènes de fonction inconnue identifiés à 24 hpi sont des gènes sous-exprimés par la souche virulente, soit 8 gènes sur 11 détectés comme étant différentiellement exprimés (Figure 22). Parmi ces gènes hypothétiques, des domaines de similitude avec de motifs protéiques existants ont été identifiés notamment pour le CDS\_ERGA\_07980 qui possède un domaine NifU (Tableau 12). Bien que ces protéines soient très mal connues, des études récentes indiquent qu'elles pourraient intervenir dans l'assemblage des clusters fer-soufre et dans la régulation du fer. Pour les autres gènes aucun motif protéique similaire à d'autres bactéries ou organismes n'a été trouvé.

#### 3.2.2.2. Gènes surexprimés par la souche Gardel atténuée à T1 (24 hpi vs 24 hpi)

Enfin parmi les gènes sous-exprimés chez la souche virulente, 17/24 interviennent soit dans le métabolisme de base ou alors sont des gènes hypothétiques (Tableau 13). Toutefois des gènes importants dans la pathogenèse sont surexprimés par la souche atténuée tels que les *clpB* (FC=6,30), chaperonne qui sert à désagréger les protéines agrégées, instables ou mal repliées, *secB*, le constituant cytoplasmique essentielle à la translocation de protéine *via* les systèmes de sécrétions sec dépendant. De façon surprenante, on identifie également parmi les gènes les plus fortement différentiellement exprimés le gène *hflK* (FC=5,28) qui code pour une protéine

qui lorsqu'elle est complexée avec *hflC* va moduler *ftsH* qui est une protéase majeure participant à la lyse de la membrane cellulaire.

### 3.2.2.3. Gènes surexprimés par la souche Gardel virulente à T2 (48 hpi vs 72 hpi)

Parmi les gènes les plus fortement surexprimés par la souche virulente *mgtE* qui code pour un transporteur de magnésium ( $Mg^{2+}$ ) a été identifié (FC=5,48) (Tableau 14). Chez *Salmonella* les gènes *mgtBC*, qui codent également pour des systèmes de transport de magnésium, sont requis pour la multiplication intracellulaire et l'établissement d'une infection systémique. La souche virulente surexprime également le gène *fdxB* qui code pour la ferrédoxine, impliquée dans la régulation et l'acquisition du fer (FC=4,26). Plusieurs gènes codant pour la réparation et la réplication de l'ADN sont surexprimés notamment les hélicases *recB* (FC=5,43) et *dnaB* (FC=5,13) qui sont parmi les 3 premiers gènes les plus différentiellement exprimés. Parmi toutes les hélicases, qui sont responsables du déroulement de l'ADN lors de la réplication, DnaB est de loin la protéine la plus efficace. Quant à *recB*, il code pour une protéine majeure impliquée dans la virulence par ses fonctions de réparation de l'ADN et de colonisation de l'hôte notamment chez *Helicobacter pylori* (Wang and Maier, 2009). On note également une surexpression de deux gènes appartenant aux SST4, les CDS\_ERGA\_00150 et CDS\_ERGA\_05360, qui codent pour les protéines virB10 (FC=3,18) et virB6 (FC=2,87) respectivement. De plus les gènes *typA*, *pstA* et *dnaK* sont tous surexprimés par la souche virulente par rapport à l'atténuée (Tableau 14). *TypA* (homologue de *bipA* gène stade spécifique chez *Bordetella*), constitue un régulateur important de la réponse adaptative aux modifications environnementales (Deora, 2002). *PstA* code pour une perméase transmembranaire du système *Pst* impliqué dans l'acquisition du phosphate et *dnaK* code pour une protéine chaperonne impliquée dans la résistance au choc thermique.

La catégorie de gènes la plus représentée regroupe des gènes hypothétiques (36%), dont beaucoup ont des domaines protéiques de fonction connue et important dans la pathogenèse et la virulence d'autres espèces bactériennes. Une large majorité des gènes inconnus est surexprimée par la souche virulente EGRp39 (figure 23). Parmi ces gènes, on retrouve les CDS\_ERGA\_06440 (FC=3,87) et CDS\_ERGA\_01690 (FC=3,74) qui codent pour les motifs Ankyrines et protéines SNARE respectivement (Tableau 14). Les protéines à répétitions de

motifs ankyrine en tandem sont identifiées chez les Rickettsie et sont potentiellement impliquées dans divers processus tels que, le cycle cellulaire, l'organisation du cytosquelette, la toxicité ou encore la réponse inflammatoire (Rikihisa and Lin, 2010). Les protéines SNARE sont quant à elle responsables de l'inhibition de la fusion phagolysosomiale, empêchant la destruction de la vacuole intracytoplasmique (morula) dans cette phase de croissance exponentielle (Paumet *et al.*, 2009). On retrouve également deux CDS très fortement surexprimées par la souche virulente, les CDS\_ERGA\_02500 (FC=5,01) et CDS\_ERGA\_02360 (FC=4,92) qui possèdent 43% d'homologie avec des hémolysines ou hémagglutinines et 40 % avec les protéines de type RhoGEF respectivement. Egalement très fortement exprimé, le CDS\_ERGA\_07300 (FC=4,91) qui code pour la protéine périplasmique DsbA, indispensable au maintien de l'intégrité et à la stabilisation de protéines exportées *via* le périplasme et de protéines membranaires. Le CDS\_ERGA\_04130 (FC=4,39) qui contient des domaines protéiques similaires aux gènes codant pour des protéines de la famille TRAP (Tripartite ATP-independent Periplasmic Transporter). Ces protéines TRAP sont des transporteurs périplasmiques pour de nombreux solutés indispensables à la bactérie. Ces protéines sont immunogènes chez *E. chaffeensis* et *E. canis* (Barbet *et al.*, 2001). Enfin le CDS\_ERGA\_01020 code pour une protéine qui possède un domaine SoxR et HspR, régulateurs majeurs impliqués dans la réponse au stress oxydatif, la réponse à une augmentation de la température et de l'osmolarité (Compan and Touati, 1993; Nunoshiba, 1996).

Tableau 14 : Gènes surexprimés par la souche virulente ERGp39 à T2

<b>CDS</b>	<b>Fold ch</b>	<b>P-value</b>	<b>Genes</b>	<b>Descriptions</b>
ERGA_CDS_04720	5,49	9,09E-04	mgtE	Magnesium transporter
ERGA_CDS_06470	5,44	3,10E-03	recB	Putative Exodeoxyribonuclease V beta chain
ERGA_CDS_05910	5,13	1,20E-03	dnaB	Replicative DNA helicase
ERGA_CDS_02500	5,01	2,43E-03	unknown	43% homologue with Hemolysins or hemagglutinin
ERGA_CDS_02360	4,93	1,61E-02	unknown	(40%) homologue with RhoGEF protein
ERGA_CDS_07300	4,92	8,09E-03	unknown	Similar to thiol:disulfide interchange protein dsbA"
ERGA_CDS_03920	4,64	6,98E-03	fabG	3-oxoacyl-[acyl-carrier protein] reductase
ERGA_CDS_07980	4,61	3,35E-03	unknown	NifU-like domain, IscR-regulated protein YhgI
ERGA_CDS_03360	4,53	4,32E-03	nifR3	Putative nitrogen regulation protein nifR3
ERGA_CDS_00950	4,46	2,98E-03	ispF	2-C-methyl-D-erythritol 2,4-cyclodiphosphate / synthase
ERGA_CDS_08610	4,42	4,30E-03	dacF	Penicillin-binding protein dacF precursor/ beta lactamase
ERGA_CDS_04130	4,39	1,39E-04	unknown	TRAP transporter solute receptor, TAXI family PBPb
ERGA_CDS_02650	4,38	4,80E-03	sucA	2-oxoglutarate dehydrogenase E1 component
ERGA_CDS_04300	4,26	4,80E-03	fdxB	Ferredoxin, 2Fe-2S
ERGA_CDS_05860	4,25	9,21E-03	thio	Putative thiamine biosynthesis oxidoreductase thiO
ERGA_CDS_05560	4,23	7,40E-03	unknown	protein DnaQ-like exonuclease superfamily
ERGA_CDS_06730	4,22	4,53E-03	bioB	Biotin synthase
ERGA_CDS_06390	4,15	5,28E-03	unknown	Hypothetical protein
ERGA_CDS_09050	4,13	1,17E-03	map1-10	map1-related protein
ERGA_CDS_05680	4,07	1,04E-02	rne	Ribonuclease E
ERGA_CDS_05900	4,03	1,60E-02	unknown	Hypothetical protein
ERGA_CDS_05770	4,01	1,11E-02	unknown	Hypothetical protein
ERGA_CDS_02200	3,97	1,01E-02	coq7	similar to human Ubiquinone biosynthesis protein Coq7
ERGA_CDS_00620	3,96	4,33E-03	unknown	Hypothetical protein
ERGA_CDS_06440	5,49	9,09E-04	unknown	Putative ankyrin repeat protein
ERGA_CDS_03770	5,44	3,10E-03	typA	GTP-binding protein TypA/BipA homolog
ERGA_CDS_01690	5,13	1,20E-03	unknown	Conserved hypothetical protein SNARE
ERGA_CDS_04920	5,01	2,43E-03	nuoJ	NADH-quinone oxidoreductase chain J
ERGA_CDS_03220	4,93	1,61E-02	rluC	Ribosomal large subunit pseudouridine synthase C
ERGA_CDS_09420	4,92	8,09E-03	unknown	conserved hypothetical protein TIGR00103
ERGA_CDS_00150	4,64	6,98E-03	virB10	VirB10 protein
ERGA_CDS_09070	4,61	3,35E-03	map1-8	Map1-related protein
ERGA_CDS_05670	3,01	1,46E-02	dnaK	Chaperone protein dnaK (Heat shock protein)
ERGA_CDS_05360	2,87	1,70E-02	unknown	VirB6 components domain fragment
ERGA_CDS_05430	2,85	1,28E-02	pstA	Phosphate transport system permease protein
ERGA_CDS_00270	2,57	3,05E-03	dapD	carboxylate N succinyltransferase
ERGA_CDS_01020	2,33	9,95E-03	unknown	HspR-like proteins / SoxR domain / merR related family
ERGA_CDS_07630	1,95	1,41E-02	unknown	Hypothetical protein

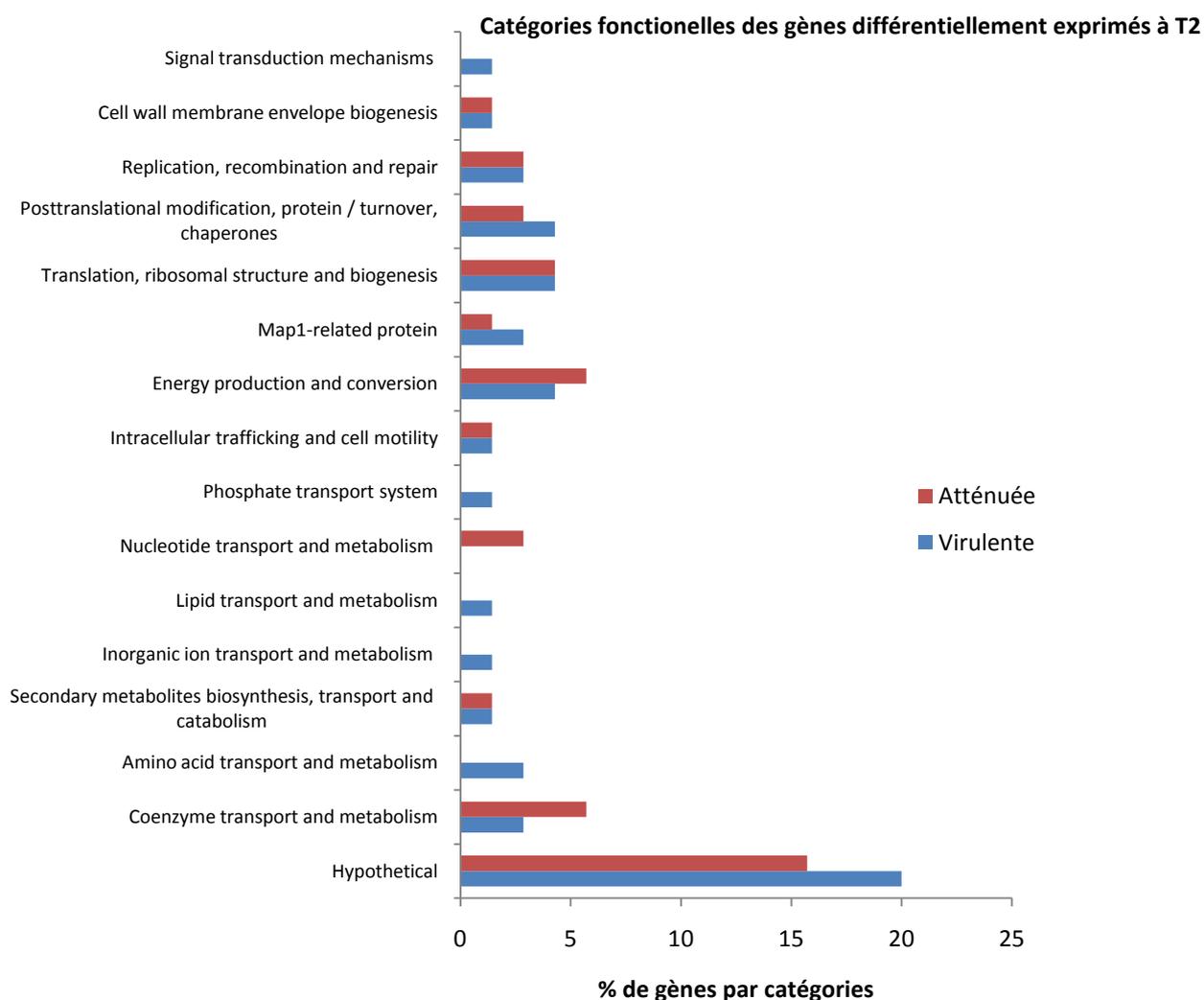


Figure 23 : Catégorie de gènes différentiellement exprimés entre les souches Gardel virulentes et atténuées d'*Ehrlichia ruminantium* au moment de la phase de croissance exponentielle (T2)

#### 3.2.2.4. Gènes surexprimés par la souche Gardel atténuée à T2 (48 hpi vs 72 hpi).

Près de 50% (15/32) des gènes surexprimés par la souche atténuée ERGp243 à ce stade sont de fonction partiellement ou complètement inconnue (Figure 23). Toutefois, parmi les gènes

les plus pertinents exprimés à ce stade *dsbE* (FC=-5,69) et *def* (FC=-2,43) ont été identifiés (Tableau 15). Le premier fortement exprimé par ERGp243 possède une structure et fonction semblable à *dsbA* et le second code pour la protéine peptide deformylase 1. On note également la présence de deux gènes appartenant au complexe Ccm, *ccmA* et *ccmB*, qui exporte l'hème nécessaire pour la biosynthèse du cytochrome c chez les bactéries. Enfin *virB11* (FC=-1,73) qui appartient au SST4 est surexprimé par la souche atténuée (Tableau 15).

Tableau 15 : Gènes sous-exprimés par la souche virulente ERGp243 à T2

<b>CDS</b>	<b>Fold ch</b>	<b>P-value</b>	<b>Genes</b>	<b>Descriptions</b>
ERGA_CDS_08340	-5,80	2,74E-06	unknown	Putative exported protein
ERGA_CDS_02750	-5,78	4,99E-03	unknown	Hypothetical protein
ERGA_CDS_07180	-5,69	1,87E-06	dsbE	Thiol:disulfide interchange protein dsb
ERGA_CDS_03620	-5,32	1,27E-04	unknown	Hypothetical protein
ERGA_CDS_04560	-5,00	3,64E-03	unknown	Conserved hypothetical protein
ERGA_CDS_01810	-4,93	8,82E-03	unknown	Hypothetical protein
ERGA_CDS_08740	-4,30	1,67E-02	unknown	Hypothetical protein
ERGA_CDS_00300	-4,22	3,47E-03	dfp	pantothenate metabolism flavoprotein homolog
ERGA_CDS_06500	-4,16	6,31E-05	folP	Dihydropteroate synthase
ERGA_CDS_00610	-3,98	1,10E-02	unknown	Hypothetical protein
ERGA_CDS_07320	-3,93	7,84E-03	ccmA	Heme exporter protein A
ERGA_CDS_02790	-3,91	1,37E-02	unknown	Hypothetical protein
ERGA_CDS_00340	-3,68	1,34E-02	ccmb	Heme exporter protein B
ERGA_CDS_09150	-3,67	3,43E-03	Map1-1	Map1-related protein
ERGA_CDS_05880	-3,62	1,71E-04	unknown	(putative cytidine deaminase)
ERGA_CDS_01150	-3,49	9,60E-03	unknown	Apolipoprotein N-acyltransferase
ERGA_CDS_05820	-3,48	4,76E-04	unknown	Putative Holliday junction resolvase
ERGA_CDS_00240	-3,46	1,09E-03	xseA	Probable exodeoxyribonuclease VII large subunit
ERGA_CDS_07250	-3,40	1,04E-02	unknown	Hypothetical protein
ERGA_CDS_0840	-3,23	4,11E-04	pth	Peptidyl-tRNA hydrolase
ERGA_CDS_00180	-3,02	7,90E-03	ribA	GTP cyclohydrolase II
ERGA_CDS_07530	-3,02	3,78E-03	frrr	Ribosome recycling factor
ERGA_CDS_01840	-2,99	1,66E-04	sdhC	Succinate dehydrogenase cytochrome B-556 subunit
ERGA_CDS_00470	-2,94	9,87E-04	unknown	hypothetical protein
ERGA_CDS_03180	-2,87	5,90E-03	unknown	Hypothetical protein
ERGA_CDS_01760	-2,43	1,45E-02	def	Peptide deformylase I
ERGA_CDS_02970	-2,09	3,09E-03	thiC	Thiamine biosynthesis protein thiC
ERGA_CDS_07260	-2,00	1,60E-02	dcd	Deoxycytidine triphosphate deaminase
ERGA_CDS_00750	-1,94	7,95E-03	atpH	ATP synthase delta chain
ERGA_CDS_00140	-1,73	1,70E-02	virB11	VirB11 protein
ERGA_CDS_03030	-1,40	1,31E-02	unknown	putative deoxyribonuclease
ERGA_CDS_00660	-1,35	1,10E-02	gltA	Citrate synthase

#### 3.2.2.5. Gènes surexprimés par la souche Gardel virulente à T3 (96 hpi vs 120 hpi)

A T3, 34% des gènes surexprimés par la souche virulente sont de fonction et de classe totalement inconnue (Figure 24). Parmi les 40 gènes de fonction inconnue détectés au moment de la lyse, un très grand nombre fait partie des plus fortement exprimés que ce soit par la souche virulente ou la souche atténuée (Tableau 16 et 17). Les deux gènes les plus différenciellement surexprimés par la souche virulente et de fonction inconnue sont les CDS\_ERGA\_07330 (FC=7,92) et CDS\_ERGA\_06420 (FC=6,48) (Tableau 16). L'annotation du génome et la recherche de motif protéique n'ont pas permis de trouver des domaines fonction connus pour la majorité d'entre eux.

Au moment de la lyse, on détecte une très forte surexpression du gène *atpB* (FC=6,43) qui code pour la sous unité alpha de l'ATP synthétase (Tableau 16). De nombreux transporteurs parmi lesquels *mgtE* (FC=5,75), et *proP* (FC=5,63) ont aussi été identifiés comme fortement surexprimés. *MgtE* code pour un transporteur du magnésium et *proP* pour un transporteur de proline et de bêtaïne connus pour leurs fonctions osmoprotectantes (Csonka, 1989). On retrouve également de nombreux transporteurs ABC (SST1): *lolC* (FC=5,33), le CDS\_ERGA\_01410 (FC=4,03) pour lequel on retrouve des motifs protéiques codants pour une perméase de type ABC et le gène *pstB* (FC=3,20) qui appartient au système Pst (Tableau 16).

Les CDS\_ERGA\_06430 et ERGA\_07160 sont aussi surexprimés par la souche virulente. Le premier CDS possède un domaine band 7 similaire de la flotilline et HflK/C entre autre, impliqué dans divers processus intracellulaires comme la transduction des signaux, la réorganisation du cytosquelette pour l'un ou encore la prise de décision par la cellule d'induire un cycle lytique ou lysogénique en cas d'infection par un bactériophage pour l'autre (Tavernarakis *et al.*, 1999; Hinderhofer *et al.*, 2009). Le deuxième gène code pour des protéines de la famille TRAP, catégorie retrouvée également surexprimée à T2. Parallèlement à cela on identifie le gène *sppA* qui code pour la protéase IV qui est relativement bien connue chez beaucoup de bactéries hautement pathogènes (Poquet *et al.*, 2001). ERGp39 surexprime les gènes *ligA*, *dnaB* et le CDS\_ERGA\_05820 qui codent pour une DNA ligase, une hélicase et une résolvasse de la jonction de holliday respectivement. Enfin le gène *fdxB* (ferredoxine) est également surexprimé par la souche virulente (Tableau 16).

### 3.2.2.6. Gènes surexprimés par la souche Gardel atténuée à T3 (96 hpi vs 120 hpi)

Beaucoup de gènes intervenant dans le métabolisme de base sont détectés (Figure 24). Les gènes de réplication et réparation de l'ADN *recG*, *dnaQ*, *xseA* codant pour une hélicase ADN dépendante, la chaîne epsilon de l'ADN polymérase III et large sous unité de l'exodéoxyribonucléase VII, sont surexprimés par la souche atténuée (Tableau 17). On détecte également le gène *proP* différent de celui surexprimé par la souche virulente à ce stade (CDS\_ ERGA \_02820), *hemH* qui code pour une ferrocélatase et *cspA* une protéine chaperonne induite en réaction au stress thermique. Enfin le gène *bcr* codant pour une protéine de résistance à la bicyclomycine est surexprimé par la souche atténuée (Tableau 17).

Tableau 16 : Liste de gènes relevant surexprimés par la souche virulente ERGp39 à T3

CDS	Fold ch	P-value	Genes	Descriptions
ERGA_CDS_07330	7,92	1,53E-10	unknown	Hypothetical protein
ERGA_CDS_06420	6,48	1,08E-07	unknown	Hypothetical protein
ERGA_CDS_08780	6,43	3,28E-09	atpB	ATP synthase A chain
ERGA_CDS_04720	5,75	1,46E-08	mgtE	Magnesium transporter
ERGA_CDS_02740	5,63	2,36E-09	proP	Proline-betaine transporter (tolerance to osmotic stress)
ERGA_CDS_00790	5,33	3,91E-06	lolC	protein lolC - ABC transporter
ERGA_CDS_03060	4,66	5,77E-08	cycJ	Cytochrome c-type biogenesis protein cycJ/ccmE
ERGA_CDS_3890	4,65	2,30E-07	xthA	Exodeoxyribonuclease III
ERGA_CDS_01410	4,03	2,33E-03	unknown	Probable ABC transporter permease protein yqgH
ERGA_CDS_03200	3,82	6,35E-07	suhB	Inositol-1-monophosphatase
ERGA_CDS_06430	3,62	5,91E-05	unknown	band 7 domain of flotillin (reggie) like proteins (HflK/C)
ERGA_CDS_00860	3,58	8,76E-06	comF	Similar to DNA transformation competence protein F
ERGA_CDS_09120	3,39	3,32E-06	map1-3	Map1-related protein
ERGA_CDS_09250	3,31	2,66E-03	soj	Sporulation initiation inhibitor protein soj
ERGA_CDS_05960	3,20	1,95E-04	pstB	ATP-binding protein pstB (ABC phosphate transporter)
ERGA_CDS_05090	3,07	7,88E-06	unknown	similar to Haemophilus competence protein comM
ERGA_CDS_08970	3,00	1,16E-04	pcnB	Poly(A) polymerase
ERGA_CDS_08450	3,00	2,05E-04	rnc	Ribonuclease III
ERGA_CDS_06590	2,88	6,24E-05	purN	Phosphoribosylglycinamide formyltransferase
ERGA_CDS_05570	2,88	3,79E-05	unknown	GTP-binding protein Era homolog
ERGA_CDS_07160	2,86	1,11E-03	unknown	TRAP C4-dicarboxylate transport permease superfamily
ERGA_CDS_00670	2,82	3,38E-03	GshA	Glutamate--cysteine ligase, T. ferrooxidans family
ERGA_CDS_05820	2,76	2,82E-03	unknown	Putative Holliday junction resolvase
ERGA_CDS_09310	2,69	5,60E-06	unknown	Belongs to the cinA family; inducible protein cinA
ERGA_CDS_08770	2,69	1,35E-04	egsA	Probable glycerol-1-phosphate dehydrogenase [NAD(P)]

## *Analyse du transcriptome de la bactérie Ehrlichia ruminantium*

ERGA_CDS_09080	2,67	8,21E-03	map1-7	Map1-related protein
ERGA_CDS_02580	2,62	2,71E-04	znuA	High-affinity zinc uptake system protein znuA
ERGA_CDS_08610	2,59	1,23E-04	dacF	Penicillin-binding protein dacF precursor
ERGA_CDS_04300	2,55	4,88E-05	fdxB	Ferredoxin, 2Fe-2S
ERGA_CDS_01770	2,52	5,40E-05	argH	Argininosuccinate lyase
ERGA_CDS_06350	2,50	5,69E-03	sppA	Putative Protease IV
ERGA_CDS_05910	2,30	1,36E-02	dnaB	Replicative DNA helicase
ERGA_CDS_07310	2,20	1,67E-02	coxW	Cytochrome C oxidase assembly protein
ERGA_CDS_08460	2,15	3,02E-03	ctaG	Cytochrome C oxidase assembly protein ctaG
ERGA_CDS_09050	2,07	1,20E-03	map1-10	Map1-related protein
ERGA_CDS_09110	1,94	2,27E-04	map1-4	Map1-related protein
ERGA_CDS_08080	1,87	2,25E-04	ctaD	Probable cytochrome c oxidase polypeptide I
ERGA_CDS_09070	1,33	6,64E-04	map1-8	Map1-related protein
ERGA_CDS_07210	1,21	1,21E-02	ligA	DNA ligase
ERGA_CDS_05560	1,18	2,57E-03	unknown	DnaQ-like exonuclease superfamily
ERGA_CDS_00510	1,11	1,64E-02	glnA	Glutamine synthetase I
ERGA_CDS_00970	1,09	1,60E-02	unknown	(Ser rich protein)

Tableau 17 : Liste de gènes pertinents sous-exprimés par la souche virulente ERGp39 à T3

<b>CDS</b>	<b>Fold ch</b>	<b>P-value</b>	<b>Genes</b>	<b>Descriptions</b>
ERGA_CDS_08320	-5,87	3,05E-08	unknown	Frameshift version of ERWE_CDS_08430
ERGA_CDS_01830	-5,06	1,51E-05	aroA	3-phosphoshikimate 1-carboxyvinyltransferase
ERGA_CDS_09440	-4,98	1,87E-05	metK	S-adenosylmethionine synthetase
ERGA_CDS_04130	-4,88	2,60E-08	unknown	Hypothetical protein
ERGA_CDS_09230	-4,67	1,95E-07	unknown	Hypothetical protein
ERGA_CDS_05680	-4,56	1,86E-06	rne	Ribonuclease E
ERGA_CDS_03690	-4,50	1,90E-07	unknown	Hypothetical protein
ERGA_CDS_09090	-4,37	1,25E-07	map1-6	Map1-related protein
ERGA_CDS_00310	-4,33	1,60E-06	recG	ATP-dependent DNA helicase RecG
ERGA_CDS_05130	-4,20	1,04E-04	dnaQ	DNA polymerase III, epsilon chain
ERGA_CDS_04310	-3,74	9,28E-09	unknown	Hypothetical protein
ERGA_CDS_02820	-3,66	8,04E-06	proP	Proline/betaine transporter
ERGA_CDS_00240	-3,50	2,20E-04	xseA	Probable exodeoxyribonuclease VII large subunit
ERGA_CDS_08260	-3,24	1,01E-02	gatC	Putative Glutamyl-tRNA(Gln) amidotransferase subunit C
ERGA_CDS_06400	-3,21	3,21E-05	hemH	Ferrochelatase
ERGA_CDS_09030	-3,03	7,62E-05	map1-12	Map1-related protein
ERGA_CDS_05330	-3,03	3,01E-04	ispH	4-hydroxy-3-methylbut-2-enyl diphosphate reductase

*Analyse du transcriptome de la bactérie Ehrlichia ruminantium*

---

ERGA_CDS_01690	-3,00	4,17E-04	unknown	Conserved hypothetical protein SNARE
ERGA_CDS_09150	-2,88	2,47E-03	map1-1	Map1-related protein
ERGA_CDS_02670	-2,85	1,30E-05	dapA	Dihydrodipicolinate synthase
ERGA_CDS_00770	-2,81	1,07E-05	unknown	Hypothetical protein
ERGA_CDS_07580	-2,78	6,34E-03	unknown	Hypothetical protein
ERGA_CDS_05210	-2,77	7,63E-05	znuB	High-affinity zinc uptake system membrane protein znuB
ERGA_CDS_08240	-2,72	3,97E-03	mrp	Mrp protein
ERGA_CDS_08520	-2,59	1,03E-04	ribF	Riboflavin biosynthesis protein ribF
ERGA_CDS_03000	-2,57	5,22E-04	unknown	Hypothetical protein
ERGA_CDS_07410	-2,44	1,78E-04	unknown	Hypothetical protein
ERGA_CDS_03720	-2,36	8,92E-04	unknown	Hypothetical protein
ERGA_CDS_07240	-2,23	3,70E-03	cspA	Putative Cold shock-like protein cspA
ERGA_CDS_06280	-2,12	8,78E-04	rplD	50S ribosomal protein L4
ERGA_CDS_04490	-1,89	7,66E-06	miaA	tRNA delta(2)-isopentenylpyrophosphate transferase
ERGA_CDS_00550	-1,88	4,21E-04	secF	Protein-export membrane protein SECF pas significatif
ERGA_CDS_02750	-1,87	1,67E-02	unknown	Hypothetical protein
ERGA_CDS_00460	-1,82	1,58E-02	rpe	Ribulose-phosphate 3-epimerase
ERGA_CDS_02860	-1,78	1,12E-02	gatB	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase
ERGA_CDS_09210	-1,73	8,06E-03	unknown	Hypothetical protein
ERGA_CDS_02710	-1,31	6,05E-03	nadE	Probable glutamine-dependent NAD(+) synthetase
ERGA_CDS_00010	-1,30	1,17E-02	nadC	Nicotinate-nucleotide pyrophosphorylase
ERGA_CDS_07760	-1,25	1,56E-02	bcr	Bicyclomycin resistance protein
ERGA_CDS_02920	-1,24	5,54E-03	pdxJ	Pyridoxal phosphate biosynthetic protein pdxJ

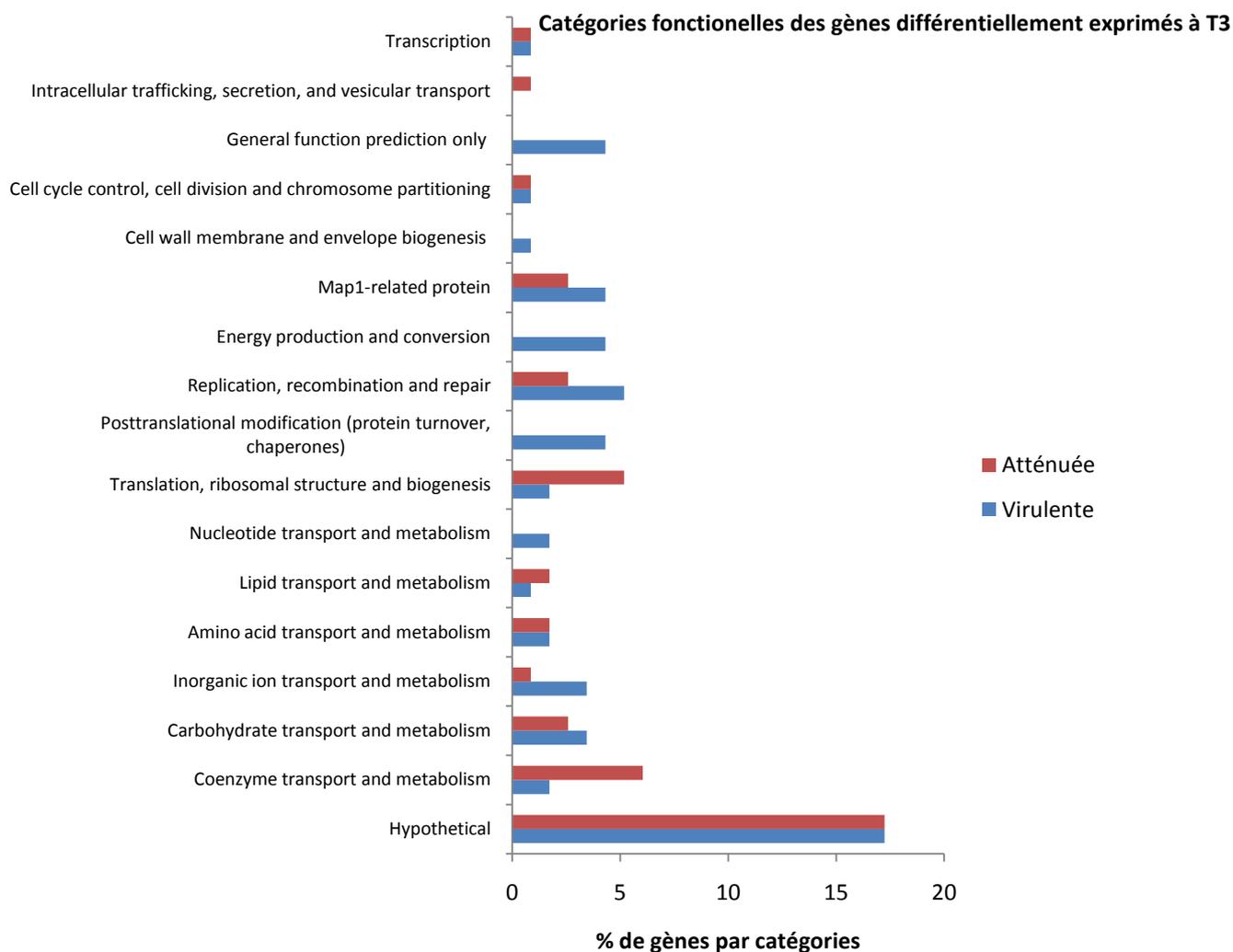


Figure 24 : Catégorie de gènes différentiellement exprimés entre les souches Gardel virulentes et atténuées d’*Ehrlichia ruminantium* moment de la lyse des cellules (T3)

### 3.2.2.7. Gènes communs aux différents temps

On compte en tout 19 gènes exprimés différentiellement de façon significative à deux moments du cycle d’infection (Tableau 18). On identifie uniquement le CDS\_ERGA\_4130 de fonction inconnue, surexprimé à T1 et T3 pour la souche atténuée. Le gène *sdhC* (Succinate dehydrogenase) est surexprimé à T1 et sous-exprimé à T2 pour la souche virulente (Tableau 18B). Trois autres gènes sont exprimés de façon inverse entre T2 et T3 suivant les souches (*rne*, CDS05820 et CDS 04130). Les gènes communs les plus surexprimés pour la souche

*Analyse du transcriptome de la bactérie Ehrlichia ruminantium*

virulente aux temps T2 et T3 sont *mgtE* (transport de magnésium), l'hélicase *dnaB*, *dacF* qui code pour une protéine appartenant à la famille des PBPs (penicillin-binding-proteins) qui permet le maintien de la polymérisation du peptidoglycane à la surface de la membrane de la cellule bactérienne (Popham *et al.*, 1999; Popham and Young, 2003), le gène *fdxB* codant pour la ferrédoxine et les gènes *map1-10* et *map1-8* (Tableau 18B).

Tableau 18 : Listes de gènes pertinents communs aux différents temps (T1/T2 et T2/T3). A : Comparaison T1 et T2. B : Comparaison T2 et T3

A.

CDS	Fold ch (T1)	Fold ch (T2)	Genes
ERGA_CDS_01840	5,72	-2,99	sdhC /Succinate dehydrogenase cytochrome B-556
ERGA_CDS_07980	6,38	4,61	hypothetical protein turnover, chaperones
ERGA_CDS_05770	3,15	4,01	Hypothetical protein

B.

CDS	Fold ch (T2)	Fold ch (T3)	Genes
ERGA_CDS_04720	5,49	5,75	<b>mgtE</b>
ERGA_CDS_05910	5,13	2,30	<b>dnaB</b>
ERGA_CDS_08610	4,42	2,59	<b>dacF</b>
ERGA_CDS_04130	4,39	-4,88	Hypothetical protein
ERGA_CDS_04300	4,26	2,55	<b>fdxB</b>
ERGA_CDS_05560	4,23	1,18	Conserved hypothetical protein
ERGA_CDS_09050	4,13	2,07	<b>map1-10</b>
ERGA_CDS_05680	4,07	-4,56	rne
ERGA_CDS_01690	3,74	-3,00	SNARE-like protein
ERGA_CDS_09070	3,11	1,33	<b>map1-8</b>
ERGA_CDS_00240	-3,46	-3,50	xseA
ERGA_CDS_05820	-3,48	2,76	Putative Holliday junction resolvase
ERGA_CDS_09150	-3,67	-2,88	map1-1
ERGA_CDS_00300	-4,22	-3,13	dfp
ERGA_CDS_02750	-5,78	-1,87	Hypothetical protein

### 3.2.3. La famille multigénique *map*

On identifie plusieurs gènes appartenant à la famille multigénique *map* soit 8 CDS *map* sur 16, modulés à T3 (Figure 25). Outre *map1-10* et *map1-8* qui sont surexprimés à T2 et T3 pour la souche virulente, *map1-3*, *map1-7* et *map1-4* sont aussi surexprimés à T3 (Tableau 16). Inversement, le gène *map1-1* est surexprimé par la souche atténuée à T2 et T3 (Tableau 18B) et *map1-6*, *map1-1* et *map1-12* sont surexprimés uniquement à T3.

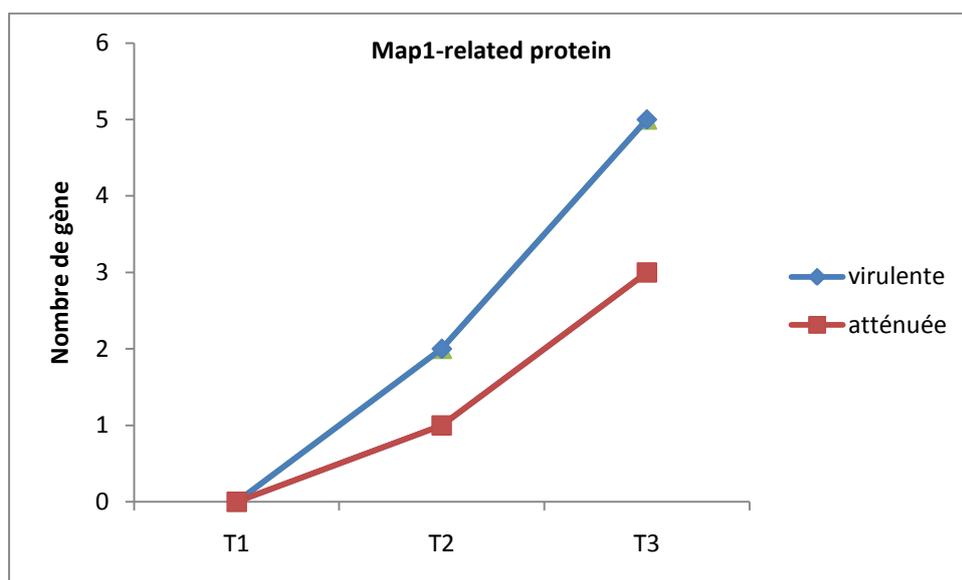


Figure 25 : Nombre de gènes de la famille *map* surexprimés à chaque phase (T1/T2/T3)

### 3.2.4. Le cluster *recB*

A l'aide de l'analyse microarrays et des données de séquençage des génomes des souches Gardel virulente et atténuée, nous avons identifié un ensemble de gènes contigus, co-exprimés avec le gène *recB* (Figure 26). La majorité de ces gènes est surexprimée dans la souche virulente entre T2 et T3. La plupart de ces gènes co-exprimés avec *recB* sont connus pour leur implication dans plusieurs processus pathogéniques décrits dans la littérature chez d'autres bactéries.

Les CDS\_ERGA\_06340 (*rpsA*, ribosomal protein), CDS\_ERGA\_06350 (*sppA*, signal peptide peptidase), CDS\_ERGA\_06390 (protéine de membrane hypothétique), CDS\_ERGA\_06420 (Hypothétique), CDS\_ERGA\_06430 (flotillin/HflK), CDS\_ERGA\_06440 (Ankyrin repeats protein), CDS\_ERGA\_06470 (*recB*), sont tous surexprimés par la souche virulente (Tableau

12, 14 et 16). Seul le gène CDS\_ERGA\_06400 à T3 (ferrochélatase) est surexprimé et détecté uniquement chez la souche atténuée (Tableau 17). Il est intéressant de remarquer qu'en analysant de façon comparative les séquences nucléotidiques de ces différents CDS entre la souche atténuée et la souche virulente, on détecte une mutation dans les gènes *recB* et (*HflK*) CDS\_ERGA\_06430 pour la souche atténuée (Figure 26).

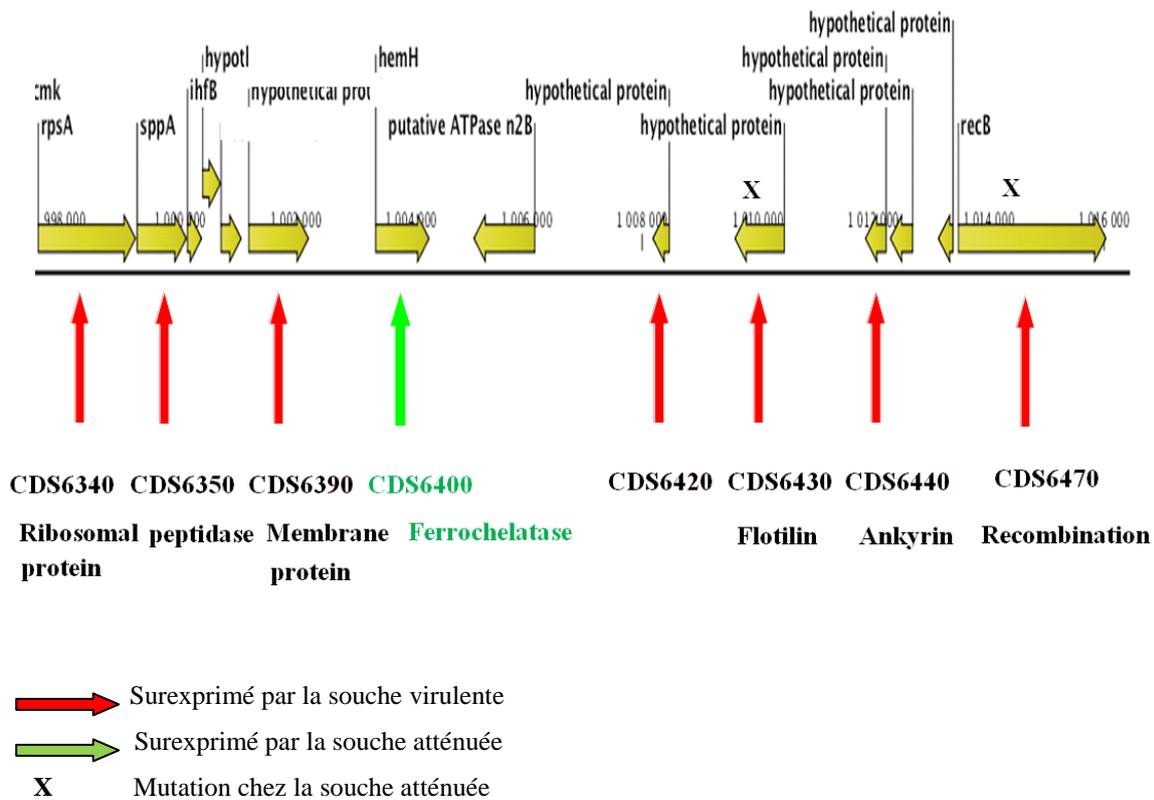


Figure 26 : Représentation schématique du cluster *recB*

### 3.2.5. Transcriptome d'*E. ruminantium* à partir des transcrits SCOTS 0x à T3

Pour les échantillons SCOTS 0x, 43% (766/1772) des CDS présents sur la lame sont détectés contrairement à ce qui est observé avec SCOTS2x (>75%). En effet, seulement 77 gènes ont été identifiés comme différentiellement exprimés entre ERGp39 et ERGp243 avec des ratios inférieurs à 3 et supérieur à -3 pour la surexpression et la sous-expression des gènes pour la souche virulente. 50% (38/77) des gènes différentiellement exprimés sont surexprimés par la

souche virulente et 50% (39/77) sous exprimés. Seuls 14 gènes détectés sont communs avec l'analyse SCOTS 2x à T3.

### 3.3. Validation des CDS d'intérêts par qRT-PCR

La validation du différentiel d'expression des gènes par PCR quantitative a été faite dans un premier temps sur les gènes les plus surexprimés par la souche virulente aux différents temps et/ou dont la fonction peut être associée à des mécanismes de pathogenèse. Au temps T1, nous avons choisi de tester l'expression de 5 CDS : CDS00440, 08000, 05660, 07980 et 05400. Au temps T2, nous avons testé 6 CDS, les CDS 06470, 03770, 06440, 02360, 05360, 04720 et au temps T3 8 CDS, les CDS 06420, 02740, 08780, 04720, 07330, 09900, 09080 et 06790 (annexe 1).

Afin de faire une mesure relative d'expression de gènes, l'ensemble des PCR ciblant ces gènes ont été testés et validés en mesurant l'efficacité de la PCR pour chaque couple d'amorces à partir des dilutions sériées d'ADN génomique d'*E. ruminantium*. L'efficacité obtenue était comprise entre 90 et 100% selon les amorces utilisées. Les échantillons de cDNA ont été testés sans capture par qRT-PCR avec comme gènes normalisateurs les 3 gènes déjà établis comme étant des gènes d'expression relativement stable : *pCS20*, *recA* et *16S* (voir chapitre II). L'efficacité des PCR ciblant les 3 gènes endogènes est équivalente à celle des gènes cibles nous permettant d'utiliser la méthode des  $\Delta\Delta CT$ . Quelque soit le gène normalisateur utilisé et quelque soit l'échantillon testé (replicat biologique ou échantillon utilisé en parallèle pour la capture SCOTS et l'analyse microarrays), nous avons obtenu une surexpression systématique des gènes pour la souche atténuée par rapport à la souche virulente ce qui est l'inverse des résultats obtenus par microarrays.

La stabilité de l'expression de ces gènes sur la souche atténuée a été évaluée sur deux réplicats biologiques ERGp241 et 243 en utilisant la méthode décrite au chapitre II. Le nombre de transcrits de *pCS20*, *recA* et *16S* a été mesuré et une normalisation par le nombre de bactéries pour chaque échantillon a été faite. Le rapport R (nombre de copies ADN (*mapI*)/nombre de copies cADNs) est représenté dans le tableau 19. Les résultats obtenus montrent que l'expression des 3 gènes choisis est instable pour la souche atténuée au cours du temps. Les rapports sont plus faibles pour la souche virulente que pour la souche atténuée ce

## *Analyse du transcriptome de la bactérie Ehrlichia ruminantium*

qui traduit un niveau d'expression des gènes beaucoup plus important pour la souche virulente par rapport à la souche atténuée.

La recherche de gènes candidats stables au cours du temps chez les deux souches est en cours. Par conséquent, 10 gènes ne variant pas au cours du temps d'après les données de microarrays ont été choisis en comparant l'ensemble des résultats SCOTS2x et SCOTS0x pour pouvoir trouver le ou les candidats idéaux dans le but de valider les résultats microarrays (Tableau 20). En parallèle des CDS normalement surexprimés par la souche atténuée au cours du temps ont été également choisis pour vérification par qRT-PCR (annexe 2).

Tableau 19 : Stabilité de l'expression des gènes (R) chez ERGp243 au cours du temps

<b>R</b>	24 hpi			48 hpi			72 hpi			96 hpi			120 hpi
	Gp38	Gp243	Gp241	Gp38	Gp243	Gp241	Gp38	Gp243	Gp38	Gp243	Gp241	Gp38	
<i>recA1</i>	21	8	79	10	278	471	4	102	7	26	57	25	252
<i>pCS20</i>	9	23		12	893		7	482	9	117	72		
<i>16SI</i>	19	12	73	17	157	128	16	17	13	5	5	7	25

Tableau 20 : Liste des gènes choisis pour la validation de gènes de ménages sur la base des résultats microarrays

<b>sondes</b>	<b>genes/fonctions</b>	<b>M_T1_scot2x</b>	<b>M_T2_scot2x</b>	<b>M_T3_scot2x</b>	<b>M_T3_scot0x</b>
ERGA_CDS_01640_b	50S ribosomal protein L7/L12	-0,87	-0,73	-0,16	-0,72
ERGA_CDS_01640_a		-0,36	-0,79	0,12	-0,70
ERGA_CDS_02130_b	3-oxoacyl- synthase II	-0,27	-0,50	0,58	0,40
ERGA_CDS_02130_a		0,20	-0,15	0,34	0,42
ERGA_CDS_07690_a	Hypothetical protein	-0,26	-0,37	-0,41	0,00
ERGA_CDS_07690_b		-0,85	-0,69	-0,73	0,21
ERGA_CDS_07610_a	hypothetical protein	-0,11	0,48	0,27	-0,25
ERGA_CDS_06070_a	30S ribosomal protein S11	0,81	-0,18	-0,06	-0,87
ERGA_CDS_06070_b		0,99	-0,17	-0,27	-0,46
ERGA_CDS_06170_a	50S ribosomal protein L5	-0,79	-1,03	0,25	-0,57
ERGA_CDS_06170_b		0,78	-0,14	0,07	-0,44
ERGA_CDS_07140_a	30S ribosomal protein S6	0,81	0,13	-0,48	-0,49
ERGA_CDS_07140_b		1,01	-0,25	-0,85	-1,42
ERGA_CDS_04390_a	Hypothetical protein	0,57	-0,76	-0,87	-0,34
ERGA_CDS_04390_b		0,58	-1,02	-0,79	0,69
ERGA_CDS_07660_a	Hypothetical protein	-0,67	-0,32	-0,45	0,81

## **CHAPITRE IV.**

### **Discussion générale**

Dans cette étude, nous avons voulu mieux comprendre la pathogénèse de la bactérie *E. ruminantium* et notamment identifier les gènes reliés aux mécanismes de survie, à l'adaptation au milieu intracellulaire mammalien et surtout à la virulence. Pour ce faire, notre modèle expérimental et notre analyse reposaient essentiellement sur la comparaison *in vitro* du niveau d'expression des gènes, entre deux souches Gardel virulente et atténuée, aux génomes quasi identiques, mais ayant deux comportements différents *in vivo* vis-à-vis de l'animal infecté conduisant à la mort de l'animal infecté par la souche virulente et à sa survie en cas d'infection par la souche atténuée. Ces souches possèdent toutes les deux un cycle de développement *in vitro* identique et complexe comportant différents stades évolutifs mais avec une lyse cellulaire plus rapide pour la souche atténuée par rapport à la virulente. Nous avons déterminé le profil transcriptomique comparé des deux souches au cours des différentes phases du cycle d'infection *in vitro*. Cette analyse transcriptomique a pu être conduite par l'utilisation de la méthode SCOTS permettant de contourner le problème, caractéristique des pathogènes intracellulaires obligatoires, de faible quantité d'ARN bactérien récolté parmi l'ARN eucaryote en particulier aux stades précoces de développement (Graham and Clark-Curtiss, 1999; Hinton *et al.*, 2004).

Pour la réalisation de ce travail, plusieurs cinétiques (répétitions biologiques) ont été générées pour la souche virulente et pour la souche atténuée en gardant toujours des passages similaires dans les cellules endothéliales afin que la variation génétique observée au cours des passages cellulaires successifs n'affecte pas le comportement de la bactérie entre 2 répliquats biologiques. Les souches ERGp39 (virulente) et ERGp243 (atténuée) ont été utilisées pour l'application de la méthode SCOTS suivie de l'hybridation des échantillons (cDNA SCOTS2x) sur les puces à ADN couvrant l'ensemble des séquences codantes d'*E. ruminantium*. L'hybridation des ADNs génomiques des deux souches Gardel virulente et atténuée démontre une très forte colinéarité ( $R^2 = 0,98$ ) des génomes. Les différences attendues en utilisant l'analyse microarrays pour la comparaison des deux souches relèvent donc plus du niveau d'expression différentiel des gènes, plutôt que des différences au niveau du génome.

L'analyse globale du transcriptome des deux souches *in vitro* a révélé de nombreux gènes différentiellement exprimés aux différents stades de développement avec une majorité au

stade infectieux. L'identification de la fonction des gènes modulés est un indicateur de leur rôle dans l'invasion, le développement de la bactérie, le détournement de la machinerie cellulaire de l'hôte, l'échappement à la réponse de l'hôte. Les nombreux gènes identifiés permettent d'obtenir un schéma global des différents mécanismes développés par la bactérie au cours de l'infection *in vitro* (figure 27).

## **1. Expression des gènes impliqués dans les fonctions métaboliques**

Au cours de l'infection des cellules endothéliales, les deux souches vont pénétrer dans la cellule hôte par un moyen qui est encore mal connu pour la virulente et non étudié chez l'atténuée. *E. ruminantium* va devoir très rapidement s'adapter au milieu intracellulaire et va donc devoir synthétiser durant les différentes phases du cycle de développement, de nouveaux composants impliqués dans son métabolisme notamment des cofacteurs, des enzymes nécessaires à la production d'énergie et à l'acquisition de nutriments :

### **1.1. Biogénèse des ribosomes et la traduction**

L'analyse globale révèle 29 gènes différentiellement exprimés qui sont reconnus pour leurs implications dans la traduction, la structure et la biogénèse des ribosomes. Au cours de la phase précoce de croissance de la bactérie qui correspond au 1<sup>er</sup> jour du cycle de développement pour les deux souches d'*E. ruminantium* (24 hpi), on retrouve 15 gènes impliqués dans la traduction dont 11 sont surexprimés par la souche virulente. La surexpression de ces gènes impliqués dans la traduction et la biogénèse des ribosomes est probablement significative de la différence de comportement qu'il y a entre la souche virulente et la souche atténuée. La capacité que possède la souche virulente à surexprimer des gènes intervenant dans la synthèse de protéines lui confère un avantage sélectif important dans la préparation aux autres phases de développement au cours desquelles la bactérie est dans une vacuole intracytoplasmique puis libre sous forme de corps élémentaire. On peut supposer que (i) beaucoup plus de protéines seront sécrétées pour pouvoir acquérir des nutriments nécessaires à la phase de croissance (ii) la synthèse d'enzymes permettant la production d'énergie sera également plus conséquente. La souche virulente surexprime le

gène *infC* qui code le facteur de démarrage de la traduction IF3 qui se fixe à la petite sous unité 30S du ribosome qui contrôle la fidélité de la traduction des ARNm en protéines. Des gènes codant pour les protéines constitutives de la petite sous-unité du ribosome (30S), impliquée dans la lecture de l'ARNm font également parti des plus surexprimés par la souche virulente en particulier un régulateur majeur de la traduction et de l'élongation chez les bactéries Gram négative *rpsA* (Sengupta *et al.*, 2001). La mutation de ce gène chez *E. coli* induit une réduction progressive mais drastique de la croissance bactérienne. La protéine ribosomique S1 codée par *rpsA* est un composant essentiel du ribosome 30S et est par conséquent une cible de choix pour beaucoup d'antibiotiques (Herring and Blattner, 2004). Les gènes *rpmF*, *rplR*, *rplM*, *rplT*, *rpmA*, *rplU*, *rplQ* codant pour des protéines (sous unités constitutives) de la grande sous unité du ribosome (50S) qui est le catalyseur direct dans la synthèse protéique sont significativement surexprimés par la souche virulente. La souche virulente surexprime très fortement le gène codant pour une peptidyl deformylase (def ou PDF) durant cette phase de latence. La fonction de l'enzyme PDF dans la maturation des protéines est essentielle et son inhibition est létale chez les bactéries. Pour cette raison, cette enzyme présente uniquement chez les procaryotes est une cible thérapeutique de choix pour la conception de nouveaux antibiotiques, et est très étudiée pour tenter de répondre par exemple au problème de multi-résistance des bactéries pathogènes aux antibiotiques classiquement utilisés (Mamelli *et al.*, 2009). En phase exponentielle de croissance, l'expression des gènes impliqués dans la synthèse protéique diminue, avec 6 CDS détectés dont 3 sont surexprimés par la souche virulente et 3 surexprimés par la souche atténuée. En revanche au moment de la lyse cellulaire, la souche atténuée surexprime 6 gènes impliqués dans la synthèse protéique alors que la souche virulente n'en surexprime que 2. Les résultats obtenus *in vitro* suggèrent que la phase critique nécessitant une grande synthèse protéique et permettant d'amorcer le cycle pathogénique est la phase de latence des bactéries. Si les observations réalisées *in vitro* montrent que la souche atténuée semble avoir un cycle de développement similaire à celui de la souche virulente, une différence de niveau d'expression des gènes impliqués dans la traduction et la biogénèse des ribosomes pourrait se traduire *in vivo* par une baisse du niveau de la traduction des protéines réduisant ainsi la capacité de la bactérie atténuée à se multiplier.

## 1.2. La production d'énergie

La production d'énergie est essentielle chez la bactérie pour l'élaboration des nombreux processus nécessaires au cycle de développement, qu'elle soit virulente ou atténuée. En phase de latence on observe chez la souche virulente, une surexpression significative de deux gènes clés intervenant dans le cycle des acides tricarboxyliques, *sucC* codant pour la chaîne bêta de la succinyl-CoA synthetase responsable du clivage de la succinyl-CoA en succinate avec production d'ATP et la succinate déshydrogénase (*sdhC*) qui catalyse l'oxydation du succinate en fumarate. À T2, *sdhC* est surexprimé mais par la souche atténuée de même que le gène *atpH* codant pour la chaîne delta de l'ATP synthétase. Des études récentes menées chez la souche SR-11 de *Salmonella enterica* serovar Typhimurium ont révélé qu'une mutation de *sucCD* ou de *sdhCDA* rendait la bactérie modérément atténuée pour la première mutation et fortement atténuée pour la seconde (Mercado-Lubo *et al.*, 2008). À la vue de cette étude, il est possible que ces deux gènes jouent un rôle important dans la virulence d'*E. ruminantium*. La bactérie *E. ruminantium* possède l'ensemble des éléments constituant ces deux groupes géniques. À T3, une énorme production d'énergie semble être induite par la souche virulente par la surexpression du gène *atpB* codant pour la chaîne alpha de l'ATP synthétase, qui synthétise l'ATP lors de la phosphorylation oxydative.

## 1.3. Réparation et réplication de l'ADN

En pénétrant dans la cellule hôte, les bactéries peuvent être endommagées ce qui nécessite des mécanismes de réparation, de réorganisation et de réplication de l'ADN. Dans notre modèle, des gènes impliqués dans ces mécanismes de réparation et de réplication sont plus fortement induits chez la souche virulente (11/17) au cours du cycle de développement, en particulier après l'entrée dans la cellule et à la lyse cellulaire. On peut supposer qu'au moment de la lyse des cellules hôtes, la bactérie virulente maintienne l'intégrité de son matériel génétique, d'où l'importance de l'expression de ces gènes. La souche virulente surexprime *ihfA* qui code pour le facteur IHF, faisant partie de la famille des protéines de type Histone. *ihfA* est impliqué dans l'expression des SST3 de *Salmonella typhimurium*. L'absence d'IHF, protéine hétérodimérique composée de sous-unités codées par *ihfA* et *ihfB*, conduit à la diminution de

l'expression des gènes codant les protéines intervenant aux différentes étapes de la biosynthèse des trois SST3 (Mangan *et al.*, 2006). Des profils globaux de transcription ont été réalisés chez *typhimurium* afin de comparer la souche de référence et des mutants dépourvus d'IHF (Mangan *et al.*, 2006). De profonds changements de la transcription ont été observés suite à l'absence d'IHF, les gènes affectés étant notamment impliqués dans la transition entre les phases exponentielles et stationnaires, mais aussi dans la virulence et l'invasion des cellules épithéliales (Mangan *et al.*, 2006). Même si dans notre modèle, aucun SST3 n'a été identifié, il est possible qu'*ihfA* joue un rôle important dans la pathogénèse.

Des gènes impliqués dans la réplication et la réparation de l'ADN sont aussi surexprimés tels que les gènes codant pour des gyrases (*gyrB*) et de *ssb* qui maintiennent tout deux l'ADN dans une conformation stable afin de faciliter la transcription immédiate des gènes. En parallèle, la bactérie doit continuer à se multiplier et lutter pour maintenir l'intégrité de son ADN lors de la phase exponentielle et stationnaire. Les hélicases *recB* et *dnaB* et 6 autres gènes, *dnaQ* (exonucléase), *xthA* (Exodeoxyribonuclease III), le\_CDS\_01880, le CDS\_05820 (Holliday junction resolvase), *ligA* (DNA ligase), participant à la réparation et/ou la réplication de l'ADN sont surexprimées par la souche virulente durant la phase exponentielle. Cette fonction semble être aussi importante pour la souche atténuée qui surexprime 6 gènes de cette catégorie mais avec des différentiels d'expression qui sont relativement faibles.

## **2. Expression des gènes impliqués dans la survie intracellulaire : échappement aux systèmes de défense de l'hôte.**

### 2.1. Les stress oxydatifs, thermiques et osmotiques

Dans la phase précoce du cycle de développement, un gène étroitement lié à l'échappement de la réponse à la cellule hôte codant pour la thioredoxine (*trx* ; (ERGA\_CDS\_07850) et qui permet de lutter contre la poussée respiratoire est fortement surexprimé chez la souche virulente. Son expression pourrait expliquer qu'*in vivo* lors de l'infection, la souche virulente échappe à la première réponse des monocytes et macrophages, tandis que la souche atténuée pourrait avoir perdu cette capacité. A l'intérieur de la vacuole intracytoplasmique la bactérie doit lutter contre des modifications éventuelles de température, de pH et le stress oxydatif. Dans ce contexte, sa capacité à répondre à ces différents stress pourrait être corrélée à son

niveau de virulence. Nous avons mis en évidence des gènes surexprimés chez la souche virulente à T2 et impliqués dans la régulation du stress thermique et oxydatif: *dnaK* CDS\_ERGA\_07670 et le CDS\_ERGA\_01020 (Domaine SoxR/HspR). Au moment de la lyse cellulaire, une modification brusque de l'osmolarité du milieu externe peut engendrer des dommages délétères et irréversibles pour la bactérie. Par conséquent, celle-ci possède des systèmes de transport d'osmoprotectants qui semblent être activés pour limiter l'effet inhibiteur de l'hyperosmolarité. Ainsi, un des 4 CDS du gène *proP* codant un système de transport d'osmoprotectants tels que la proline et la bétaine (Csonka, 1989) fait partie des 5 gènes les plus surexprimés (FC= 5,63) par la souche virulente par rapport à l'atténuée. Il semble qu'il participe au maintien de l'homéostasie et de l'intégrité cellulaire chez la bactérie. La souche atténuée surexprime également un autre gène *proP* (CDS\_ERGA\_02820) en réponse au stress osmotique et *cspA* en réponse au stress thermique, au moment de la lyse avec un plus faible différentiel d'expression (FC= -3,65 et -2,23 respectivement).

## 2.2. Acquisition de fer, magnésium et autres éléments

Entre la phase exponentielle et stationnaire, la souche virulente induit très fortement des systèmes d'acquisitions et de transport du fer (*fdxB*, ferredoxine) et du magnésium (*mgtE*) qui sont tous deux essentiels pour le pouvoir pathogène et des transporteurs périplasmiques de la famille TRAP, assurant le transport d'acides aminés, peptides, sucres, vitamines et ions inorganiques. La souche atténuée exprime également à ces stades de développement des éléments d'acquisition du fer notamment une ferrochélatase, *hemH* et les gènes *ccmA* et *ccmB* qui participent respectivement à la séquestration du fer et au transport de l'hème.

## 2.3. Transporteurs ABC: SST1

Quatre éléments appartenant à différents groupes de SST1 ou transporteurs ABC, dont 3 sont bien connus, sont surexprimés à la phase exponentielle et à la lyse cellulaire. Les gènes *pstA/B* appartenant au système Pst (importateur) sont surexprimés par la souche virulente permettant probablement une meilleure acquisition du phosphate par rapport à l'atténuée. L'inactivation ou la mutation de cet importateur semble entraîner une sensibilité accrue à plusieurs

xénobiotiques chez *Mycobacterium smegmatis* et diminue la pathogénicité des *E. coli* associés à des pathologies évoluant vers une septicémie chez les porcs (Daigle et al., 1995; Banerjee et al., 2000). Le gène *lolC* du système ABC est également surexprimé. Ce système particulièrement bien connu chez les bactéries notamment chez *E. coli* est indispensable pour la croissance bactérienne et l'exportation des lipoprotéines de la membrane interne vers la membrane externe. Le 4<sup>ème</sup> système transporteur ABC identifié, le CDA\_ERGA\_01410, appartient à cette grande famille de transporteurs mais l'annotation n'a pas permis de révéler quel type d'élément était exporté ou importé.

#### 2.4. Inhibition de la fusion phagolysosomiale

Le profil d'expression des gènes de la bactérie obtenu à la phase exponentielle et à la lyse cellulaire apporte des éléments importants pour la compréhension des mécanismes de virulence. Au cours de ces deux phases, la bactérie va se multiplier de façon exponentielle au sein d'une vacuole intracytoplasmique, la *morula*, jusqu'à la lyse des cellules hôtes. A l'intérieur de cette niche de survie et de prolifération, les bactéries doivent mettre en place différentes stratégies pour éviter la maturation de la vacuole en phagolysosomes qui leur serait délétère. Ainsi, nous avons identifié un gène codant pour une protéine SNARE connue pour son rôle dans l'inhibition de la fusion phagolysosomiale (Paumet et al., 2009) et qui est surexprimé en phase exponentielle et sous exprimé en phase de lyse pour la souche virulente. La faible expression de ce gène chez la souche atténuée au moment où elle en a le plus besoin (phase intracellulaire) pourrait induire une reconnaissance précoce du pathogène par les SNARE de la cellule hôte provoquant une lyse anticipée des cellules. Ce faible niveau d'expression pourrait expliquer nos observations concernant la lyse plus précoce des cellules par la souche atténuée *in vitro*. Ainsi, l'inhibition de la fusion phagolysosomiale dans la cellule est fortement corrélée à la virulence des souches.

#### 2.5. Résistance aux antibiotiques

La souche virulente surexprime de façon significative à la phase exponentielle et à la lyse cellulaire, le gène *dacF* qui code pour une bêta-lactamase appartenant à la famille des PBPs

(penicillin-binding-proteins) et qui permet le maintien de la polymérisation du peptidoglycane à la surface de la membrane de la cellule bactérienne. Or, la synthèse du peptidoglycane est essentielle pour la survie et la virulence de la bactérie (Popham *et al.*, 1999; Popham and Young, 2003). En revanche la souche atténuée surexprime à la lyse cellulaire le gène *bcr* codant pour une protéine de résistance à la bicyclomycine qui bloque la protéine rho ATPase au niveau du transcrit.

### **3. La famille multigénique *map***

La famille protéique MAP a été étudiée de part ses propriétés polymorphes et de protéines membranaires qui font de ces protéines des candidats vaccinaux de choix. Huit protéines de la famille MAP appartenant à la membrane externe sont exprimées de façon différentielle. Certaines, codées par *map1-6*, *map1-1* et *map1-12*, sont exclusivement surexprimées par la souche atténuée à la phase exponentielle et à la lyse cellulaire. Les 5 autres gènes, *map1-3*, *map1-7*, *map1-10*, *map1-4* et *map1-8*, sont surexprimés uniquement par la souche virulente. Un essai vaccinal réalisé au laboratoire avec un vaccin recombinant contenant quatre protéines de la famille MAP (MAP1+1, MAP1-6, MAP1-2 et MAP1-14), a été réalisé. Il a permis de protéger 50 % des animaux contre une infection létale réalisée en conditions contrôlées. Une amélioration de la production et de la purification des protéines est en cours avant un nouvel essai vaccinal. Les MAP identifiées au cours de l'analyse du transcriptome devraient être intégrées aux nouveaux essais vaccinaux. Si l'on considère que certaines MAP ont un fort pouvoir immunogène, celles qui sont surexprimées de façon spécifique par la souche atténuée pourraient participer éventuellement à la reconnaissance de cette souche par le système de défense de l'hôte induisant son élimination comme observé *in vivo*.

### **4. Le cluster *recB***

Le gène *recB* semble être impliqué dans divers processus en particulier la réponse au stress oxydatif, la réparation de l'ADN et la colonisation de l'hôte mammalien chez *Helicobacter pylori*. Il a été montré que la mutation de ce gène entraîne une diminution importante de la survie dans différentes conditions et peut entraîner la mort de la bactérie (Wang and Maier, 2009). Un ensemble de gènes contigus ou proches, semble être co-régulé entre T1 et T3, les

CDS\_ERGA\_06340 (*rpsA*), CDS\_ERGA\_06350 (*sppA*), CDS\_ERGA\_06390 (protéine de membrane hypothétique), CDS\_ERGA\_06420 (Hypothétique), CDS\_ERGA\_06430 (*Band7*) et CDS\_ERGA\_06440 (Ankyrine). Quatre des 6 gènes co-exprimés avec *recB* sont connus pour leurs implications dans plusieurs processus pathogéniques décrits dans la littérature chez les bactéries. CDS\_ERGA\_06440 code pour une protéine contenant des répétitions de motifs de type Ankyrine. Ces protéines, qui sont des médiateurs de nombreuses interactions protéine-protéine chez les eucaryotes, sont transloquées dans le cytoplasme de la cellule hôte par le SST4 chez les bactéries intracellulaires *Legionella pneumophila*, *Coxiella burnetii*, *Anaplasma phagocytophilum* et *E. chaffeensis* (Pan *et al.*, 2008; Rikihisa and Lin, 2010). Même si la translocation de ces protéines par le SST4 n'a pas encore été identifiée chez *E. ruminantium*, des études dans le genre *Ehrlichia* et *Anaplasma* montrent qu'elles sont toutes des substrats du SST4 chez ces bactéries (Rikihisa and Lin, 2010). Fait encore plus marquant, l'analyse microarrays révèle qu'à la phase exponentielle, le gène codant pour une protéine à motifs ankyrine est surexprimé chez la souche virulente simultanément avec deux gènes codant pour les protéines VirB10 et VirB6 appartenant au STT4 d'*E. ruminantium*. Un autre gène co-exprimé avec *recB* et codant la protéase IV (*sppA*) correspond à un facteur de virulence majeur chez les bactéries qui sert aux clivages de peptide signaux des protéines exportées (Poquet *et al.*, 2001). Cette protéase semble participer à de nombreuses fonctions extracytoplasmiques (i) la nutrition azotée et la croissance, par la fourniture de peptides et d'acides aminés assimilables, (ii) le renouvellement et le recyclage des protéines exportées et (iii) la fonction « de ménage » dans l'enveloppe, qui permet de se débarrasser des protéines exportées anormales ou mal repliées. Un autre gène surexprimé chez la souche virulente, le CDS\_ERGA\_06430 qui code pour une protéine appartenant à la famille Band7 qui compte des domaines comme la stomatine, la flotilline et HflC/K qui sont impliqués dans la virulence bactérienne. Lors de la lyse, ce gène pourrait participer à la lyse des cellules infectées ou à la réorganisation du cytosquelette des cellules adjacentes pour y pénétrer afin d'initier un nouveau cycle infectieux. Chez les procaryotes, les homologues des protéines stomatine sont généralement organisés en opéron avec une protéine de membrane et une protéase, élément que l'on retrouve dans ce groupe génique.

## 5. Virulence bactérienne

Nous avons identifié un gène ayant des motifs protéiques similaires à DsbA et fortement surexprimé pour la souche virulente lors de la phase exponentielle. Or, DsbA fait partie des protéines Dsb, responsables de la formation et de l'isomérisation des ponts disulfures au sein des protéines en maturation dans le périplasme. De nombreux homologues sont retrouvés chez les bactéries gram négatifs et ils sont importants pour l'exportation de nombreux facteurs de virulence. Chez la souche uropathogénique *Escherichia coli* CFT073, il existe deux systèmes Dsb impliqués dans la virulence et constitués de plusieurs protéines (Totsika *et al.*, 2009). Chez *Ehrlichia*, l'annotation du génome n'a permis d'identifier que *dsbA*, *E* et *D*. L'activité de *dsbA* est nécessaire pour la formation d'un pili de type IV ou fimbriae qui sert à l'adhérence à la surface des cellules hôtes chez *Neisseria meningitidis* et *Vibrio cholerae* (Peek and Taylor, 1992; Tinsley *et al.*, 2004). Beaucoup de mutants pour le gène *dsbA* ne peuvent pas (i) produire ou sécréter des toxines, (ii) produire les composants constituant les STT, notamment SST3, chez *Pseudomonas* spp., *Shigella flexneri* et *Salmonella enterica* serovar Typhimurium et (iii) ont une diminution drastique de la survie (Rosadini *et al.*, 2008). Au cours de la phase exponentielle, la souche virulente co-exprime avec *dsbA* 2 CDS de fonction inconnue mais qui possède une faible homologie avec une hémolysine et la protéine RhoGEF. Ces deux protéines sont impliquées dans des mécanismes de virulence. En effet, l'hémolysine est une toxine qui se fixe à des cellules cibles et provoque la formation de canaux membranaires laissant passer les ions aboutissant à la destruction de la cellule cible. Certaines toxines secrétées par les bactéries telles que l'hémolysine alpha sont bien connues pour détériorer les cellules endothéliales. Le facteur RhoGEF est quand à lui un activateur des protéines RhoGTPases et joue un rôle important dans la réorganisation du cytosquelette, la lyse et l'invasion cellulaire en interagissant avec l'actine (Boquet and Lemichez, 2003). Lors de la phase de latence, 2 gènes importants dans la pathogenèse sont surexprimés par la souche atténuée tels que les *clpB*, qui servent à désagréger les protéines agrégées, instables ou mal repliées, *secB*, le constituant cytoplasmique essentiel à la translocation de protéines via les systèmes de sécrétions sec dépendant. Les gènes *dsbE* et *virB11* sont également surexprimés par la souche atténuée en phase exponentielle. La surexpression par la souche atténuée de ces

gènes importants pour la sécrétion et/ou l'exportation d'effecteurs intervenants dans la pathogenèse ne doit pas être négligée pour la formulation de nos hypothèses.

## **6. Les gènes de fonction inconnue**

L'annotation du génome de Gardel avait révélé que le génome comportait plus de 30% de CDS de fonctions inconnus. L'analyse transcriptomique a permis de détecter 76 CDS différemment exprimés entre la souche virulente et la souche atténuée qui sont de fonction inconnue, soit 32% des CDS détectés. Parmi ces gènes, se trouvent autant de pistes à explorer dans l'optique de l'analyse fonctionnelle pour pouvoir identifier des gènes éventuellement impliqués dans l'atténuation ou dans la virulence d'*E. ruminantium*. Des fonctions nouvelles pourraient être mises en évidence et permettront de mieux comprendre la pathogenèse des deux souches.

Comparativement aux études transcriptomiques précédentes notamment chez *Rickettsia conorii*, des éléments et systèmes communs ou similaires ont été identifiés et correspondent pour la plupart à des gènes exprimés en phase exponentielle et surtout au moment de la lyse cellulaire. On retrouve notamment les thiorédoxines, les transporteurs proline/betaine, les transporteurs ABC, *recB* et les hélicases, les *virB 4-11-10* (La *et al.*, 2007; Renesto *et al.*, 2008). En parallèle, des travaux de protéomique réalisés par Marcelino et collaborateurs sur le stade corps élémentaire pour la souche virulente, révèlent la présence de protéines correspondant à 20% de celles détectées en microarrays comme étant surexprimées. L'adéquation entre l'analyse du transcriptome et les données du protéome est essentielle pour valider l'importance des gènes modulés dans les mécanismes de virulence (données en cours de soumission).

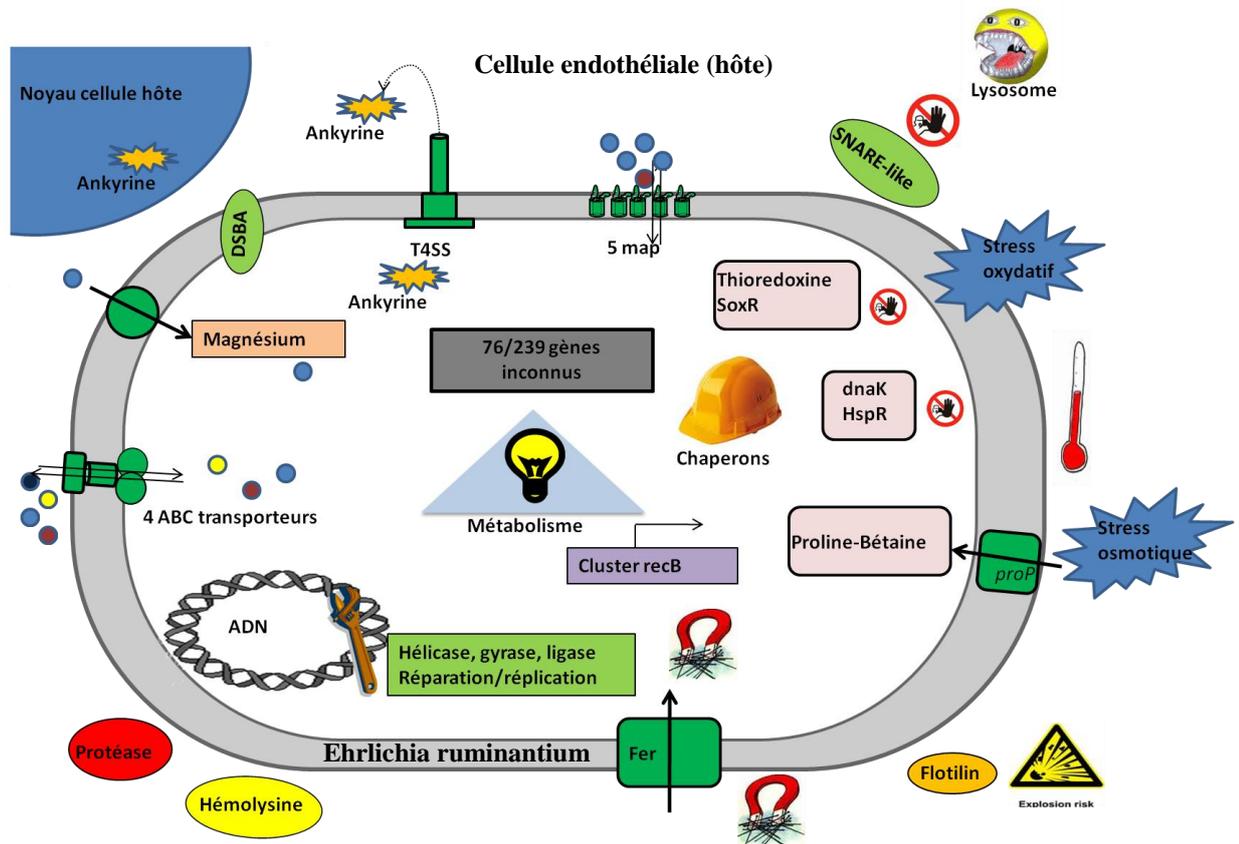


Figure 27 : Schéma global de la biologie d'*Ehrlichia ruminantium*

## **CHAPITRE V.**

### **Conclusions et perspectives**

L'adaptation de SCOTS à notre modèle *E. ruminantium* et l'utilisation des microarrays nous ont permis d'avoir une meilleure vision sur le pouvoir pathogène de la bactérie et les différences de comportements observées entre souches. Les résultats obtenus permettent de mettre en évidence les différences d'expressions existant entre la souche virulente et la souche atténuée de Gardel au cours de la cinétique *in vitro*, mais ne permettent pas d'extrapoler à ce qui se passe *in vivo*. Ils nous permettent de formuler des hypothèses pour expliquer la différence dans le pouvoir pathogène observé *in vivo* entre les deux souches. En effet, seul des analyses complémentaires permettant de valider chaque CDS détecté et considéré comme étant important pour la pathogenèse pourront nous apporter des certitudes dans l'interprétation des résultats.

Parmi toutes les hypothèses que nous pouvons formuler, l'une d'entre elles consiste à considérer que certains des gènes surexprimés chez la souche atténuée permettraient d'induire une réponse immunitaire précoce et/ou favorable à son élimination *in vivo*. La modulation de l'expression des gènes peut être la résultante d'une adaptation à la culture *in vitro* après de nombreux passages et d'une spécialisation à un environnement donné mais défavorable à son développement *in vivo*. De plus, très peu de facteurs reliés à la virulence sont surexprimés au cours du développement de la souche atténuée alors qu'elle surexprime un nombre considérable de gènes intervenant dans le métabolisme ou la réparation et la réplication de l'ADN. Ceci pourrait signifier que la bactérie atténuée cherche d'abord à assurer son métabolisme de base. On pourrait par conséquent émettre l'hypothèse que la souche virulente bien qu'étant dans des conditions *in vitro* « milieu artificiel », mime les conditions d'infections *in vivo* puisque de très nombreux gènes intervenant dans la virulence sont retrouvés aux différents stades du cycle de développement contrairement à la souche atténuée. De façon très générale, on constate que la souche virulente surexprime un très grand nombre de gènes lui permettant d'améliorer son développement optimal et sa survie dans un environnement intracellulaire qui lui est hostile. Par ailleurs on peut supposer que les systèmes de reconnaissance de la bactérie mis en place par la cellule hôte sont plus efficaces en cas d'une infection par la souche atténuée. Dans les conditions d'infection *in vitro*, la surexpression de la protéine SNARE par la souche virulente, protéine qui empêche la fusion phagolysosomiale est un élément important allant dans le sens de notre hypothèse. Il serait intéressant d'évaluer le niveau d'expression de ce gène *in vivo* pour vérifier si la fusion

phagolysosomiale n'est pas restaurée dans le cas d'une infection par la souche atténuée. La mise en évidence des gènes impliqués dans les systèmes de sécrétion (SST1 et STT4) dans l'échappement à la réponse de l'hôte, SNARE et le transport du fer nous ouvre des perspectives d'études fonctionnelles intéressantes.

La première étape doit consister à valider ces résultats en utilisant la méthode qRT-PCR. Cette validation a représenté une partie importante de ce travail de thèse. La stabilité des gènes de ménages choisis pour la normalisation a été vérifiée au cours de la cinétique d'infection de la souche virulente uniquement, l'expression des gènes *pCS20*, *recA* et surtout *16S* ne variaient pas au cours du temps. Cependant nous avons détecté a posteriori une forte variabilité dans l'expression de ces gènes au cours de la cinétique de la souche atténuée. Par conséquent, les ratios obtenus étaient incohérents (inversés) par rapport à ceux observés suite à l'analyse microarrays. De nombreux gènes candidats pour la normalisation ont été choisis pour la validation de nos résultats sur la base de leur stabilité lors de l'analyse microarrays.

Même si la mutagenèse dirigée reste encore compliquée chez ces bactéries intracellulaires obligatoires, des travaux se développent sur ces méthodes (O. Driskell *et al* 2009, Binet et Maurelli 2009), ce qui nous offre des possibilités de validation fonctionnelle de nos gènes d'intérêts. Les facteurs qui contribuent à la virulence choisis sur les bases de nos résultats peuvent être validés par des approches de complémentation dans un hôte hétérologue tel que *Bartonella henselae*. Le modèle semble bien adapté pour tester la fonctionnalité de certains gènes conservés d'*E. ruminantium*. Beaucoup de gènes de fonction inconnue sont surexprimés par les souches virulentes et atténuées. La recherche de l'impact de ces gènes dans le développement et la pathogenèse de la bactérie, par mutagenèse dirigée semble être nécessaire. L'organisation et l'expression du groupe de gènes liés à *recB* est particulièrement intéressante. A l'aide de la mutagenèse, nous pourrions vérifier s'il s'agit de gènes co-régulés. L'analyse des séquences sur le génome de la souche atténuée effectuée sur ce groupe de gènes a permis d'identifier 2 mutations pour le gène *recB* et le CDS\_ERGA\_06430 (*hflK/flotilline*). Ces mutations pourraient être à l'origine de la sous-expression de ce groupe de gène chez la souche atténuée s'ils sont co-régulés. Le séquençage récent de la souche atténuée nous permettra d'effectuer l'analyse en génomique comparative pour l'ensemble des gènes différenciellement exprimé que ce soit pour la souche virulente ou la souche atténuée.

La réponse aux différents stress intracellulaires (osmotique, thermique, oxydatif) semble être importante tout au long du cycle pathogénique. Les gènes permettant une réponse efficace semblent surexprimés dans des conditions d'infection *in vitro* pour la souche virulente par rapport à la souche atténuée. Des essais *in vitro* utilisant des monocytes purifiés infectés avec la souche virulente et atténuée permettraient de mettre en évidence les différences de comportement des souches vis-à-vis de la poussée respiratoire. D'autre part, des différences d'expression de gènes *in vivo* sur animaux infectés par les deux souches sont maintenant réalisables permettant de valider la modulation de l'expression de gènes. Des expériences peuvent être également faites en soumettant les cellules infectées à un stress thermique dans des conditions contrôlées, pour analyser la réponse de l'hôte et de la bactérie.

Les transporteurs et systèmes d'acquisition de métabolites semblent être importants pour *E. ruminantium* surtout pour la souche virulente qui surexprime une large gamme de transporteurs. Des expériences (effets doses) d'inhibition des éléments transportés tels que le fer, le magnésium et le phosphate peuvent être réalisées pour confirmer leur importance dans la survie et la virulence intracellulaire à partir du stade *morula* en utilisant par exemple des chélateurs qui réquisitionneraient les éléments disponibles dans le cytoplasme de la cellule hôte, *in vitro*. L'utilisation d'anticorps bloquant spécifiquement certaines fonctions dans des tests *in vitro* devrait être réalisable.

Les données de microarrays vont aussi être exploitées temps par temps pour une souche donnée afin d'identifier les gènes nécessaires à l'établissement de chaque stade de développement physiologique de la bactérie. La validation fonctionnelle permettrait de mettre en évidence des gènes « stades spécifiques » et nécessaires à chaque stade de développement. L'analyse du protéome complet de la souche Gardel semble nécessaire pour évaluer la proportion de gènes dont les protéines sont synthétisées au cours d'une cinétique d'infection et pour définir si les différentiels d'expression observés pour les gènes sont corrélés aux quantités de protéines synthétisées entre souches virulentes et atténuées.

L'annotation du génome et l'étude du transcriptome des souches Sénégal virulente et atténuée sont en cours. L'analyse transcriptomique de ces deux souches nous apportera des éléments de réponse notamment sur les mécanismes de pathogenèse et d'atténuation qui peuvent être différents de ce qui est observé pour la souche Gardel. L'hybridation de l'ADN

génomique de Gardel et Sénégal sur les puces a révélé une grande différence génomique, avec une faible colinéarité entre Gardel et Sénégal (virulente/atténuée). Par conséquent, d'autres gènes « souches spécifiques » pourraient être identifiés.

Grâce au travail réalisé sur le transcriptome des souches Gardel, d'autres souches pourront être analysées sur ce même modèle en optimisant de nouvelles lames d'*E. ruminantium*.

# **ANNEXES**

## Annexe 1. Gènes d'intérêts surexprimés par la souche virulente ERGp39 T1/T2/T3

---

### Amorce pour validation microarrays SCOTS2x T1/T2/T3

---

ERGA\_CDS\_07980

ERGA\_CDS\_07980F2: TCTTTCGCATCAAAAACTCTTGT

ERGA\_CDS\_07980R2: GAATGGGATGTTTTAAAGCCAGAA

ERGA\_CDS\_04720

ERGA\_CDS\_04720F2: ATTCTAATTCACCTTAGTACCAGATTTGCG

ERGA\_CDS\_04720R2: TAAGTCTTCAATTATGACTACTATGTCTTCAGTGT

ERGA\_CDS\_05910

ERGA\_CDS\_05910F1: CCCATTGCATCGTCGTATCTTT

ERGA\_CDS\_05910R1: CTCCTCCGGATTCAGTCAGTG

ERGA\_CDS\_08610

ERGA\_CDS\_08610F1: AGCATGGCAACAAGGTGGAC

ERGA\_CDS\_08610R1: TTCTGCAAGTGTGATAGATGCATC

ERGA\_CDS\_04300

ERGA\_CDS\_04300F2: TGAAGCTTATGATGGAGAACTCTATTAA

ERGA\_CDS\_04300R2: CCAACTTGGATCAATTATCACGTG

ERGA\_CDS\_00440

ERGA\_CDS\_00440F1: CAAGCCAGCCTTGAGCTTTT

ERGA\_CDS\_00440R1: CAAGAAGGATGCTTATCTGTACCAGAT

ERGA\_CDS\_05660

ERGA\_CDS\_05660F1: CAACTTTTATCCACCCTGGCA

ERGA\_CDS\_05660R1: CGTAGCGATGTGATGATAGTTGCT

ERGA\_CDS\_07850F1: AACCTCAATTAGAAAAGTTAGCCAATCA

ERGA\_CDS\_07850R1: AGTATAGTGGGTACTGCACTAACACCA

ERGA\_CDS\_05400

ERGA\_CDS\_05400F2: ATACAGGAGCAAGAGATACTGGATCTT

ERGA\_CDS\_05400R2: GATGGATCTCATGCAGCAATATTT

ERGA\_CDS\_08000

ERGA\_CDS\_08000F2: GTTTTAGCATCCTTCCATTACTGACC

ERGA\_CDS\_08000R2: CCTGAGGATAAAGGAGG AGAGGTTA

ERGA\_CDS\_06470

ERGA\_CDS\_06470F1: GCAAAACAGCGCACCATAAGT

ERGA\_CDS\_06470R1: TATCTCGTTCTAGTGTTGGAATGTAAAGTT

ERGA\_CDS\_06440

ERGA\_CDS\_06440F1: TCCATATCTTACAGCATAAGTTAACAAAGTATC

ERGA\_CDS\_06440R1: ATGTTTCTGGTTTACGTGTTTTGATT

ERGA\_CDS\_06440F2: AGTTGGCAACATTAGAATCAACACC

ERGA\_CDS\_06440R2: AACTAGATATTGGTTTGCCTGTTGTATC

ERGA\_CDS\_03770

ERGA\_CDS\_03770F1: AAGGTGATATGCAAGACTTGCGT

ERGA\_CDS\_03770R1: ACCTGTACCATGAGAATCAGTCAAAA

ERGA\_CDS\_00150

ERGA\_CDS\_00150F1: CTGATGCTCCTTGCTTCCTAGTTC

ERGA\_CDS\_00150R1: AATGCCAGCCCAGGAAAAA

ERGA\_CDS\_05360

ERGA\_CDS\_05360F2: TAATCGTTTCATTTGGTTCAGGATC

ERGA\_CDS\_05360R2: TGTGAAATGATTTCTGGAGTTCAAGA

ERGA\_CDS\_07330

ERGA\_CDS\_07330F1: TTGCACCAATCCAGCAATAAGT

ERGA\_CDS\_07330R1: TGAAGTACAATCTGAAAGGGAGAAAATT

ERGA\_CDS\_07330F2: TTTTGCACCAATCCAGCAATA

ERGA\_CDS\_06420

ERGA\_CDS\_06420F1: CAAACAGTAACAAAATCCTCTGAGTAATC

ERGA\_CDS\_06420R1: GGAAGGCAGTAATACTTCAAGCAATT

ERGA\_CDS\_08780

ERGA\_CDS\_08780F2: CACCAAGGAACTCATTTCCCTTCATA

ERGA\_CDS\_08780R2: CGAGTCGTATTGATAGGCTTACAGG

ERGA\_CDS\_02740

ERGA\_CDS\_02740F1: ACGTACTTTGCAAGGATTATCTTTTG

ERGA\_CDS\_02740 R1: CAGCTGCACCTCCAAGTATCG

ERGA\_CDS\_00790  
ERGA\_CDS\_00790F2: CTGGATCTTGACAGGCAGCTT  
ERGA\_CDS\_00790R2: TTTTTTGTGATACCTTACCCTCCATCT

ERGA\_CDS\_06430  
ERGA\_CDS\_06430F1: ACTTTCCAGACAATTACAGCAGCTATT  
ERGA\_CDS\_06430R1: TTCCATTTGTGCGTATGAGAACA

---

## Annexe 2. Gènes d'intérêts surexprimés par la souche atténuée ERGp243 T1/T2/T3

### A.

---

Amorce pour validation microarrays SCOTS2x T1	
ERGA_CDS_00640	F1 : GCTGAGCTGATTAACGGTTCT R1: TCTGTAGATCAAAGCTTAGAAGATG
ERGA_CDS_05930	F1 : GCACCACCCATAATTGCAAAGTGA R1 : GGGTGGGGATTTTGCCCCAGAA
ERGA_CDS_06620	F1 : CTCCTATCTCGATTTTGTAGCTGCAT R1 : CGCAGCGTTATATTACGGATAGGT
ERGA_CDS_08410	F1: GGACATGGTGAAGGTGTGATGTTA R1: CTTGTGATATTGCAGCTCTTCCTTT
ERGA_CDS_02560	F1 : AATTACCAAATTTAGGGTCATCTTC R1: ATCACGCTTAACATCAGAACAAA
ERGA_CDS_04240	R1 : TAACACTCCTGCTCCAGCAATG F1 : CAAACAATGAAACTGGTGCTATTCA
ERGA_CDS_07740	F1 : TCGTTTTGCCCAACACCTTGT R1: TGTTACTAGTGGCTGCTCCAAG

---

### B.

---

Amorce pour validation microarrays SCOTS 2x T2	
ERGA_CDS_08340	F1 : TACAAAGGCCCAATGAGA R1 : ATCGTGCGTTTAAATGCGTA
ERGA_CDS_07180	F1 : CCGTTCCTGATGTTTCAAGT R1 : CATGCTCTTCTGGCATGTT
ERGA_CDS_07320	F1 : CCCACGAATTAATAGCAGCA R1 : GATGTGTGAAAGGCTCATCG
ERGA_CDS_0340	F1 : GCACTGGGTTTATGTTGGAA R1 : AAGCCAATGCATGACCTACA
ERGA_CDS_05820	F1 : ACCCTAGTAGCCATAGCAGT R1 : AAAATGATGCAGGTTCAATG
ERGA_CDS_00140	F1 : GCCATCATTGGTGTATCAGC

---

---

R1 : CAAAGGAGGGCAAGGTAGAG

---

**C.**

---

**Amorce pour validation microarrays SCOTS 2x T3**

---

ERGA_CDS_08320	F1 : ACGATCTCCCAATACAAACC R1 : AGTCGATGATGGATTCCAAG
ERGA_CDS_05130	F1 : CCCCATTAATTGCACATAC R1 : GGAAGAAGTTTCCTGGAGTG
ERGA_CDS_02820	F1 : AAGGGTTGTCTTTTGCAGGT R1 : CCAACAGCACCACCAAGTAT
ERGA_CDS_0240	F1 : TCCTCCTCTGGCAATAATGA R1 : GGTGCTGCTATCAACGACAT
ERGA_CDS_06400	F1 : TCTAGACGGCAAATCCCCTA R1 : TGTTGGATGGCTGAAAATGT
ERGA_CDS_07240	F1 : TTCGAGGTTAATTGCTGAGA R1 : TCTCAACGGTTGGTAGATGA

---

## **REFERENCES BIBLIOGRAPHIQUES**

- Adakal, H., Stachurski, F., Konkobo, M., Zoungrana, S., Meyer, D.F., Pinarello, V., Aprelon, R., Marcelino, I., Alves, P.M., Martinez, D., Lefrancois, T., Vachierey, N., 2010, Efficiency of inactivated vaccines against heartwater in Burkina Faso: Impact of *Ehrlichia ruminantium* genetic diversity. *Vaccine*.
- Alexander, R.A. 1931. Heartwater. The present state of our knowledge on the disease. In Director of Veterinary Services and Animal Industry (South Africa, Union of South Africa), pp. 89-150.
- Allsopp, B.A., 2010, Natural history of *Ehrlichia ruminantium*. *Veterinary parasitology* 167, 123-135.
- Allsopp, M.T., Louw, M., Meyer, E.C., 2005, *Ehrlichia ruminantium*: an emerging human pathogen? *Annals of the New York Academy of Sciences* 1063, 358-360.
- Baldrige, G.D., Burkhardt, N.Y., Felsheim, R.F., Kurtti, T.J., Munderloh, U.G., 2008, Plasmids of the pRM/pRF family occur in diverse Rickettsia species. *Applied and environmental microbiology* 74, 645-652.
- Banerjee, S.K., Bhatt, K., Misra, P., Chakraborti, P.K., 2000, Involvement of a natural transport system in the process of efflux-mediated drug resistance in *Mycobacterium smegmatis*. *Mol Gen Genet* 262, 949-956.
- Bao, W., Kumagai, Y., Niu, H., Yamaguchi, M., Miura, K., Rikihisa, Y., 2009, Four VirB6 paralogs and VirB9 are expressed and interact in *Ehrlichia chaffeensis*-containing vacuoles. *Journal of bacteriology* 191, 278-286.
- Barbet, A.F., Whitmire, W.M., Kamper, S.M., Simbi, B.H., Ganta, R.R., Moreland, A.L., Mwangi, D.M., McGuire, T.C., Mahan, S.M., 2001, A subset of *Cowdria ruminantium* genes important for immune recognition and protection. *Gene* 275, 287-298.
- Barloy-Hubler, F., Cheron, A., Hellegouarch, A., Galibert, F., 2004, Smc01944, a secreted peroxidase induced by oxidative stresses in *Sinorhizobium meliloti* 1021. *Microbiology (Reading, England)* 150, 657-664.
- Barré, N., Uilenberg, G., Morel, P.C., Camus, E., 1987, Danger of introducing heartwater onto the American mainland: potential role of indigenous and exotic Amblyomma ticks. *Onderstepoort J Vet Res* 54, 405-417.
- Bekker, C.P., Postigo, M., Taoufik, A., Bell-Sakyi, L., Ferraz, C., Martinez, D., Jongejan, F., 2005, Transcription analysis of the major antigenic protein 1 multigene family of three in vitro-cultured *Ehrlichia ruminantium* isolates. *Journal of bacteriology* 187, 4782-4791.
- Bezuidenhout, J.D., 1987, Natural transmission of heartwater. *Onderstepoort J Vet Res* 54, 349-351.
- Binet, R., Maurelli, A.T., 2009, Transformation and isolation of allelic exchange mutants of *Chlamydia psittaci* using recombinant DNA introduced by electroporation. *Proceedings of the National Academy of Sciences of the United States of America* 106, 292-297.
- Bjur, E., Eriksson-Ygberg, S., Aslund, F., Rhen, M., 2006, Thioredoxin 1 promotes intracellular replication and virulence of *Salmonella enterica* serovar Typhimurium. *Infection and immunity* 74, 5140-5151.
- Boquet, P., Lemichez, E., 2003, Bacterial virulence factors targeting Rho GTPases: parasitism or symbiosis? *Trends in cell biology* 13, 238-246.

- Braun, V., Schonherr, R., Hobbie, S., 1993, Enterobacterial hemolysins: activation, secretion and pore formation. *Trends in microbiology* 1, 211-216.
- Brouqui, P., 1997, L'ehrlichiose humaine. Une maladie infectieuse émergente. *Med. Mal. Infect* 27, 256-266.
- Camilli, A., Mekalanos, J.J., 1995, Use of recombinase gene fusions to identify *Vibrio cholerae* genes induced during infection. *Molecular microbiology* 18, 671-683.
- Camus, E., Barre, N., 1995, Vector situation of tick-borne diseases in the Caribbean Islands. *Veterinary parasitology* 57, 167-176.
- Camus, E., Maillard, J.C., Ruff, G., Pepin, L., Naves, M., Matheron, G., 1996, Genetic resistance of Creole goats to cowdriosis in Guadeloupe. Status in 1995. *Annals of the New York Academy of Sciences* 791, 46-53.
- Cescau, S., Cwerman, H., Letoffe, S., Delepelaire, P., Wandersman, C., Biville, F., 2007, Heme acquisition by hemophores. *Biometals* 20, 603-613.
- Chan, K., Kim, C.C., Falkow, S., 2005, Microarray-based detection of *Salmonella enterica* serovar Typhimurium transposon mutants that cannot survive in macrophages and mice. *Infection and immunity* 73, 5438-5449.
- Chan, P.F., O'Dwyer, K.M., Palmer, L.M., Ambrad, J.D., Ingraham, K.A., So, C., Lonetto, M.A., Biswas, S., Rosenberg, M., Holmes, D.J., Zalacain, M., 2003, Characterization of a novel fucose-regulated promoter (P<sub>fcsK</sub>) suitable for gene essentiality and antibacterial mode-of-action studies in *Streptococcus pneumoniae*. *Journal of bacteriology* 185, 2051-2058.
- Collins, H.L., 2003, The role of iron in infections with intracellular bacteria. *Immunol Lett* 85, 193-195.
- Collins, N.E., Liebenberg, J., de Villiers, E.P., Brayton, K.A., Louw, E., Pretorius, A., Faber, F.E., van Heerden, H., Josemans, A., van Kleef, M., Steyn, H.C., van Strijp, M.F., Zweygarth, E., Jongejan, F., Maillard, J.C., Berthier, D., Botha, M., Joubert, F., Corton, C.H., Thomson, N.R., Allsopp, M.T., Allsopp, B.A., 2005, The genome of the heartwater agent *Ehrlichia ruminantium* contains multiple tandem repeats of actively variable copy number. *Proceedings of the National Academy of Sciences of the United States of America* 102, 838-843.
- Compan, I., Touati, D., 1993, Interaction of six global transcription regulators in expression of manganese superoxide dismutase in *Escherichia coli* K-12. *Journal of bacteriology* 175, 1687-1696.
- Conway, T., Schoolnik, G.K., 2003, Microarray expression profiling: capturing a genome-wide portrait of the transcriptome. *Molecular microbiology* 47, 879-889.
- Cornelis, G.R., 2006, The type III secretion injectisome. *Nature reviews* 4, 811-825.
- Cossart, P., 2000, Actin-based motility of pathogens: the Arp2/3 complex is a central player. *Cellular microbiology* 2, 195-205.
- Csonka, L.N., 1989, Physiological and genetic responses of bacteria to osmotic stress. *Microbiological reviews* 53, 121-147.
- Cummings, C.A., Relman, D.A., 2000, Using DNA microarrays to study host-microbe interactions. *Emerg Infect Dis* 6, 513-525.
- Daigle, F., Fairbrother, J.M., Harel, J., 1995, Identification of a mutation in the *pst-phoU* operon that reduces pathogenicity of an *Escherichia coli* strain causing septicemia in pigs. *Infection and immunity* 63, 4924-4927.

- Daigle, F., Graham, J.E., Curtiss, R., 3rd, 2001, Identification of *Salmonella typhi* genes expressed within macrophages by selective capture of transcribed sequences (SCOTS). *Molecular microbiology* 41, 1211-1222.
- Deora, R., 2002, Differential regulation of the *Bordetella bipA* gene: distinct roles for different BvgA binding sites. *Journal of bacteriology* 184, 6942-6951.
- Diatchenko, L., Lukyanov, S., Lau, Y.F., Siebert, P.D., 1999, Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes. *Methods Enzymol* 303, 349-380.
- Didier, G., Brezellec, P., Remy, E., Henaut, A., 2002, GeneANOVA--gene expression analysis of variance. *Bioinformatics (Oxford, England)* 18, 490-491.
- Domann, E., Wehland, J., Rohde, M., Pistor, S., Hartl, M., Goebel, W., Leimeister-Wachter, M., Wuenscher, M., Chakraborty, T., 1992, A novel bacterial virulence gene in *Listeria monocytogenes* required for host cell microfilament interaction with homology to the proline-rich region of vinculin. *The EMBO journal* 11, 1981-1990.
- Dozois, C.M., Daigle, F., Curtiss, R., 3rd, 2003, Identification of pathogen-specific and conserved genes expressed in vivo by an avian pathogenic *Escherichia coli* strain. *Proceedings of the National Academy of Sciences of the United States of America* 100, 247-252.
- Driskell, L.O., Yu, X.J., Zhang, L., Liu, Y., Popov, V.L., Walker, D.H., Tucker, A.M., Wood, D.O., 2009, Directed mutagenesis of the *Rickettsia prowazekii* *pld* gene encoding phospholipase D. *Infection and immunity* 77, 3244-3248.
- Du Plessis, J.L., 1985, A method for determining the *Cowdria ruminantium* infection rate of *Amblyomma hebraeum*: effects in mice injected with tick homogenates. *Onderstepoort J Vet Res* 52, 55-61.
- Du Plessis, J.L., Malan, L., 1987, The application of the indirect fluorescent antibody test in research on heartwater. *Onderstepoort J Vet Res* 54, 319-325.
- Dumler, J.S., 2005, Anaplasma and ehrlichia infection. *Annals of the New York Academy of Sciences* 1063, 361-373.
- Dumler, J.S., Barbet, A.F., Bekker, C.P., Dasch, G.A., Palmer, G.H., Ray, S.C., Rikihisa, Y., Rurangirwa, F.R., 2001, Reorganization of genera in the families Rickettsiaceae and Anaplasmataceae in the order Rickettsiales: unification of some species of Ehrlichia with Anaplasma, Cowdria with Ehrlichia and Ehrlichia with Neorickettsia, descriptions of six new species combinations and designation of Ehrlichia equi and 'HGE agent' as subjective synonyms of *Ehrlichia phagocytophila*. *Int J Syst Evol Microbiol* 51, 2145-2165.
- Economou, A., 1999, Following the leader: bacterial protein export through the Sec pathway. *Trends in microbiology* 7, 315-320.
- Egile, C., Loisel, T.P., Laurent, V., Li, R., Pantaloni, D., Sansonetti, P.J., Carrier, M.F., 1999, Activation of the CDC42 effector N-WASP by the *Shigella flexneri* IcsA protein promotes actin nucleation by Arp2/3 complex and bacterial actin-based motility. *The Journal of cell biology* 146, 1319-1332.
- Emboule, L., Daigle, F., Meyer, D.F., Mari, B., Pinarello, V., Sheikboudou, C., Magnone, V., Frutos, R., Viari, A., Barbry, P., Martinez, D., Lefrancois, T., Vachery, N., 2009, Innovative approach for transcriptomic analysis of obligate intracellular pathogen: selective capture of transcribed sequences of *Ehrlichia ruminantium*. *BMC Mol Biol* 10, 111.

- Estrada-Pena, A., Pegram, R.G., Barre, N., Venzal, J.M., 2007, Using invaded range data to model the climate suitability for *Amblyomma variegatum* (Acari: Ixodidae) in the New World. *Experimental & applied acarology* 41, 203-214.
- Euzeby, J.P., Tindall, B.J., 2001, Nomenclatural type of orders: corrections necessary according to Rules 15 and 21a of the Bacteriological Code (1990 Revision), and designation of appropriate nomenclatural types of classes and subclasses. Request for an opinion. *Int J Syst Evol Microbiol* 51, 725-727.
- Faburay, B., Geysen, D., Ceesay, A., Marcelino, I., Alves, P.M., Taoufik, A., Postigo, M., Bell-Sakyi, L., Jongejan, F., 2007, Immunisation of sheep against heartwater in The Gambia using inactivated and attenuated *Ehrlichia ruminantium* vaccines. *Vaccine* 25, 7939-7947.
- Faraldo-Gomez, J.D., Sansom, M.S., 2003, Acquisition of siderophores in gram-negative bacteria. *Nat Rev Mol Cell Biol* 4, 105-116.
- Faucher, S.P., 2007. Identification et caractérisation de gènes impliqués dans la virulence de *Salmonella typhi* suite à une analyse globale par biopuces de l'infection de macrophages humains en culture. Université de Montréal, Montréal.
- Faucher, S.P., Porwollik, S., Dozois, C.M., McClelland, M., Daigle, F., 2006, Transcriptome of *Salmonella enterica* serovar Typhi within macrophages revealed through the selective capture of transcribed sequences. *Proceedings of the National Academy of Sciences of the United States of America* 103, 1906-1911.
- Filloux, A., Bally, M., Ball, G., Akrim, M., Tommassen, J., Lazdunski, A., 1990, Protein secretion in gram-negative bacteria: transport across the outer membrane involves common mechanisms in different bacteria. *The EMBO journal* 9, 4323-4329.
- Filloux, A., Michel, G., Bally, M., 1998, GSP-dependent protein secretion in gram-negative bacteria: the Xcp system of *Pseudomonas aeruginosa*. *FEMS microbiology reviews* 22, 177-198.
- Frutos, R., Viari, A., Ferraz, C., Bensaid, A., Morgat, A., Boyer, F., Coissac, E., Vachiery, N., Demaille, J., Martinez, D., 2006a, Comparative genomics of three strains of *Ehrlichia ruminantium*: a review. *Annals of the New York Academy of Sciences* 1081, 417-433.
- Frutos, R., Viari, A., Ferraz, C., Morgat, A., Eychenie, S., Kandassamy, Y., Chantal, I., Bensaid, A., Coissac, E., Vachiery, N., Demaille, J., Martinez, D., 2006b, Comparative genomic analysis of three strains of *Ehrlichia ruminantium* reveals an active process of genome size plasticity. *Journal of bacteriology* 188, 2533-2542.
- Frutos, R., Viari, A., Vachiery, N., Boyer, F., Martinez, D., 2007, *Ehrlichia ruminantium*: genomic and evolutionary features. *Trends Parasitol* 23, 414-419.
- Fuller, T.E., Shea, R.J., Thacker, B.J., Mulks, M.H., 1999, Identification of in vivo induced genes in *Actinobacillus pleuropneumoniae*. *Microb Pathog* 27, 311-327.
- Fuxelius, H.H., Darby, A., Min, C.K., Cho, N.H., Andersson, S.G., 2007, The genomic and metabolic diversity of *Rickettsia*. *Res Microbiol* 158, 745-753.
- Ge, Y., Rikihisa, Y., 2007, Identification of novel surface proteins of *Anaplasma phagocytophilum* by affinity purification and proteomics. *Journal of bacteriology* 189, 7819-7828.
- Goldberg, M.B., Theriot, J.A., 1995, *Shigella flexneri* surface protein IcsA is sufficient to direct actin-based motility. *Proceedings of the National Academy of Sciences of the United States of America* 92, 6572-6576.

- Goosen, N., van de Putte, P., 1995, The regulation of transcription initiation by integration host factor. *Molecular microbiology* 16, 1-7.
- Gouin, E., Egile, C., Dehoux, P., Villiers, V., Adams, J., Gertler, F., Li, R., Cossart, P., 2004, The RickA protein of *Rickettsia conorii* activates the Arp2/3 complex. *Nature* 427, 457-461.
- Graham, J.E., Clark-Curtiss, J.E., 1999, Identification of *Mycobacterium tuberculosis* RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). *Proceedings of the National Academy of Sciences of the United States of America* 96, 11554-11559.
- Graham, J.E., Peek, R.M., Jr., Krishna, U., Cover, T.L., 2002, Global analysis of *Helicobacter pylori* gene expression in human gastric mucosa. *Gastroenterology* 123, 1637-1648.
- Gurskaya, N.G., Diatchenko, L., Chenchik, A., Siebert, P.D., Khaspekov, G.L., Lukyanov, K.A., Vagner, L.L., Ermolaeva, O.D., Lukyanov, S.A., Sverdlov, E.D., 1996, Equalizing cDNA subtraction based on selective suppression of polymerase chain reaction: cloning of Jurkat cell transcripts induced by phytohemagglutinin and phorbol 12-myristate 13-acetate. *Anal Biochem* 240, 90-97.
- Handfield, M., Levesque, R.C., 1999, Strategies for isolation of in vivo expressed genes from bacteria. *FEMS microbiology reviews* 23, 69-91.
- Hantke, K., 2001, Iron and metal regulation in bacteria. *Current opinion in microbiology* 4, 172-177.
- Herring, C.D., Blattner, F.R., 2004, Conditional lethal amber mutations in essential *Escherichia coli* genes. *Journal of bacteriology* 186, 2673-2681.
- Hinderhofer, M., Walker, C.A., Friemel, A., Stuermer, C.A., Moller, H.M., Reuter, A., 2009, Evolution of prokaryotic SPFH proteins. *BMC evolutionary biology* 9, 10.
- Hinton, J.C., Hautefort, I., Eriksson, S., Thompson, A., Rhen, M., 2004, Benefits and pitfalls of using microarrays to monitor bacterial gene expression during infection. *Current opinion in microbiology* 7, 277-282.
- Holmgren, A., 1989, Thioredoxin and glutaredoxin systems. *The Journal of biological chemistry* 264, 13963-13966.
- Huang, H., Lin, M., Wang, X., Kikuchi, T., Mottaz, H., Norbeck, A., Rikihisa, Y., 2008, Proteomic analysis of and immune responses to *Ehrlichia chaffeensis* lipoproteins. *Infection and immunity* 76, 3405-3414.
- Hubberstey, A.V., Mottillo, E.P., 2002, Cyclase-associated proteins: CAPacity for linking signal transduction and actin polymerization. *Faseb J* 16, 487-499.
- Hueck, C.J., 1998, Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev* 62, 379-433.
- Hutcheon, D., 1900, History of heartwater. *Agricultural Journal of the Cape of Good Hope* 17, 410-417.
- Jongejan, F., Vogel, S.W., Gueye, A., Uilenberg, G., 1993, Vaccination against heartwater using in vitro attenuated *Cowdria ruminantium* organisms. *Rev Elev Med Vet Pays Trop* 46, 223-227.
- Jongejan, F., Zandbergen, T.A., van de Wiel, P.A., de Groot, M., Uilenberg, G., 1991, The tick-borne rickettsia *Cowdria ruminantium* has a Chlamydia-like developmental cycle. *Onderstepoort J Vet Res* 58, 227-237.
- Kaufmann, S.H., 1993, Immunity to intracellular bacteria. *Annu Rev Immunol* 11, 129-163.

- Kelly, P.J., Yunker, C.E., Mason, P.R., Tagwira, M., Matthewman, L.A., 1992, Absence of antibody to *Cowdria ruminatum* in sera from humans exposed to vector ticks. *S Afr Med J* 81, 578.
- Kessels, M.M., Engqvist-Goldstein, A.E., Drubin, D.G., Qualmann, B., 2001, Mammalian Abp1, a signal-responsive F-actin-binding protein, links the actin cytoskeleton to endocytosis via the GTPase dynamin. *The Journal of cell biology* 153, 351-366.
- Kocks, C., Marchand, J.B., Gouin, E., d'Hauteville, H., Sansonetti, P.J., Carlier, M.F., Cossart, P., 1995, The unrelated surface proteins ActA of *Listeria monocytogenes* and IcsA of *Shigella flexneri* are sufficient to confer actin-based motility on *Listeria innocua* and *Escherichia coli* respectively. *Molecular microbiology* 18, 413-423.
- Kostakioti, M., Newman, C.L., Thanassi, D.G., Stathopoulos, C., 2005, Mechanisms of protein export across the bacterial outer membrane. *Journal of bacteriology* 187, 4306-4314.
- La, M.V., Francois, P., Rovey, C., Robineau, S., Barbry, P., Schrenzel, J., Raoult, D., Renesto, P., 2007, Development of a method for recovering rickettsial RNA from infected cells to analyze gene expression profiling of obligate intracellular bacteria. *J Microbiol Methods* 71, 292-297.
- La Scola, B., Raoult, D., 1997, Laboratory diagnosis of rickettsioses: current approaches to diagnosis of old and new rickettsial diseases. *J Clin Microbiol* 35, 2715-2727.
- Landfald, B., Strom, A.R., 1986, Choline-glycine betaine pathway confers a high level of osmotic tolerance in *Escherichia coli*. *Journal of bacteriology* 165, 849-855.
- Lawley, T.D., Chan, K., Thompson, L.J., Kim, C.C., Govoni, G.R., Monack, D.M., 2006, Genome-wide screen for *Salmonella* genes required for long-term systemic infection of the mouse. *PLoS Pathog* 2, e11.
- Le Brigand, K., Barbry, P., 2007, Mediante: a web-based microarray data manager. *Bioinformatics (Oxford, England)* 23, 1304-1306.
- Leroy, Q., Raoult, D., 2010, Review of microarray studies for host-intracellular pathogen interactions. *J Microbiol Methods*.
- Liu, Z.M., Tucker, A.M., Driskell, L.O., Wood, D.O., 2007, Mariner-based transposon mutagenesis of *Rickettsia prowazekii*. *Applied and environmental microbiology* 73, 6644-6649.
- Ma, Q., Zhai, Y., Schneider, J.C., Ramseier, T.M., Saier, M.H., Jr., 2003, Protein secretion systems of *Pseudomonas aeruginosa* and *P. fluorescens*. *Biochimica et biophysica acta* 1611, 223-233.
- Mahan, M.J., Slauch, J.M., Mekalanos, J.J., 1993, Selection of bacterial virulence genes that are specifically induced in host tissues. *Science (New York, N.Y)* 259, 686-688.
- Mahan, S.M., Peter, T.F., Simbi, B.H., Burrige, M.J., 1998, PCR detection of *Cowdria ruminantium* infection in ticks and animals from heartwater-endemic regions of Zimbabwe. *Annals of the New York Academy of Sciences* 849, 85-87.
- Mamelli, L., Petit, S., Chevalier, J., Giglione, C., Lieutaud, A., Meinnel, T., Artaud, I., Pages, J.M., 2009, New antibiotic molecules: bypassing the membrane barrier of gram negative bacteria increases the activity of peptide deformylase inhibitors. *PLoS one* 4, e6443.
- Mangan, M.W., Lucchini, S., Danino, V., Croinin, T.O., Hinton, J.C., Dorman, C.J., 2006, The integration host factor (IHF) integrates stationary-phase and virulence gene

- expression in *Salmonella enterica* serovar Typhimurium. *Molecular microbiology* 59, 1831-1847.
- Marchand, J.B., Moreau, P., Paoletti, A., Cossart, P., Carlier, M.F., Pantaloni, D., 1995, Actin-based movement of *Listeria monocytogenes*: actin assembly results from the local maintenance of uncapped filament barbed ends at the bacterium surface. *The Journal of cell biology* 130, 331-343.
- Martinez, D., Maillard, J.C., Coisne, S., Sheikboudou, C., Bensaid, A., 1994, Protection of goats against heartwater acquired by immunisation with inactivated elementary bodies of *Cowdria ruminantium*. *Vet Immunol Immunopathol* 41, 153-163.
- Martinez, D., Vachier, N., Stachurski, F., Kandassamy, Y., Raliniaina, M., Aprelon, R., Gueye, A., 2004, Nested PCR for detection and genotyping of *Ehrlichia ruminantium*: use in genetic diversity analysis. *Annals of the New York Academy of Sciences* 1026, 106-113.
- Mazel, D., Pochet, S., Marliere, P., 1994, Genetic characterization of polypeptide deformylase, a distinctive enzyme of eubacterial translation. *The EMBO journal* 13, 914-923.
- Mercado-Lubo, R., Gauger, E.J., Leatham, M.P., Conway, T., Cohen, P.S., 2008, A *Salmonella enterica* serovar typhimurium succinate dehydrogenase/fumarate reductase double mutant is avirulent and immunogenic in BALB/c mice. *Infection and immunity* 76, 1128-1134.
- Musoke, A.J., McKeever, D., Nene, V., 1997, Subunit vaccines for the control of tick-borne diseases: implications for the future. *Parassitologia* 39, 131-137.
- Nicholson, T.L., Olinger, L., Chong, K., Schoolnik, G., Stephens, R.S., 2003, Global stage-specific gene regulation during the developmental cycle of *Chlamydia trachomatis*. *Journal of bacteriology* 185, 3179-3189.
- Norval, R.A., Andrew, H.R., Yunker, C.E., 1990, Infection rates with *Cowdria ruminantium* of nymphs and adults of the bont tick *Amblyomma hebraeum* collected in the field in Zimbabwe. *Veterinary parasitology* 36, 277-283.
- Nunoshiba, T., 1996, Two-stage gene regulation of the superoxide stress response soxRS system in *Escherichia coli*. *Critical reviews in eukaryotic gene expression* 6, 377-389.
- Nyika, A., Barbet, A.F., Burridge, M.J., Mahan, S.M., 2002, DNA vaccination with map1 gene followed by protein boost augments protection against challenge with *Cowdria ruminantium*, the agent of heartwater. *Vaccine* 20, 1215-1225.
- Nyika, A., Mahan, S.M., Burridge, M.J., McGuire, T.C., Rurangirwa, F., Barbet, A.F., 1998, A DNA vaccine protects mice against the rickettsial agent *Cowdria ruminantium*. *Parasite Immunol* 20, 111-119.
- Oberem, P.T., Bezuidenhout, J.D., 1987a, Heartwater in hosts other than domestic ruminants. *Onderstepoort J Vet Res* 54, 271-275.
- Oberem, P.T., Bezuidenhout, J.D., 1987b, The production of heartwater vaccine. *Onderstepoort J Vet Res* 54, 485-488.
- Ogata, H., Renesto, P., Audic, S., Robert, C., Blanc, G., Fournier, P.E., Parinello, H., Claverie, J.M., Raoult, D., 2005, The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biol* 3, e248.

- Pan, X., Luhrmann, A., Satoh, A., Laskowski-Arce, M.A., Roy, C.R., 2008, Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* (New York, N.Y) 320, 1651-1654.
- Paradkar, P.N., De Domenico, I., Durchfort, N., Zohn, I., Kaplan, J., Ward, D.M., 2008, Iron depletion limits intracellular bacterial growth in macrophages. *Blood* 112, 866-874.
- Parola, P., Paddock, C.D., Raoult, D., 2005, Tick-borne rickettsioses around the world: emerging diseases challenging old concepts. *Clin Microbiol Rev* 18, 719-756.
- Patchett, R.A., Kelly, A.F., Kroll, R.G., 1994, Transport of glycine-betaine by *Listeria monocytogenes*. *Archives of microbiology* 162, 205-210.
- Paumet, F., Wesolowski, J., Garcia-Diaz, A., Delevoeye, C., Aulner, N., Shuman, H.A., Subtil, A., Rothman, J.E., 2009, Intracellular bacteria encode inhibitory SNARE-like proteins. *PloS one* 4, e7375.
- Peek, J.A., Taylor, R.K., 1992, Characterization of a periplasmic thiol:disulfide interchange protein required for the functional maturation of secreted virulence factors of *Vibrio cholerae*. *Proceedings of the National Academy of Sciences of the United States of America* 89, 6210-6214.
- Peter, T.F., BurrIDGE, M.J., Mahan, S.M., 2002, *Ehrlichia ruminantium* infection (heartwater) in wild animals. *Trends Parasitol* 18, 214-218.
- Poirier, K., Faucher, S.P., Beland, M., Brousseau, R., Gannon, V., Martin, C., Harel, J., Daigle, F., 2008, *Escherichia coli* O157:H7 survives within human macrophages: global gene expression profile and involvement of the Shiga toxins. *Infection and immunity* 76, 4814-4822.
- Popham, D.L., Gilmore, M.E., Setlow, P., 1999, Roles of low-molecular-weight penicillin-binding proteins in *Bacillus subtilis* spore peptidoglycan synthesis and spore properties. *Journal of bacteriology* 181, 126-132.
- Popham, D.L., Young, K.D., 2003, Role of penicillin-binding proteins in bacterial cell morphogenesis. *Current opinion in microbiology* 6, 594-599.
- Poquet, I., Bolotin, A., Gruss, A., 2001, Optimisation de la production de protéines hétérologues exportées chez *Lactococcus lactis* par inactivation de HtrA, son unique protéase de ménage de surface. *INRA, EDP Sciences, 2001 Lait* 81 (2001) 37-47.
- Pretorius, A., Liebenberg, J., Louw, E., Collins, N.E., Allsopp, B.A., 2010, Studies of a polymorphic *Ehrlichia ruminantium* gene for use as a component of a recombinant vaccine against heartwater. *Vaccine* 28, 3531-3539.
- Prozesky, L., 1987a, Diagnosis of heartwater at post-mortem in ruminants and the confirmation of *Cowdria ruminantium* in mice. *Onderstepoort J Vet Res* 54, 301-303.
- Prozesky, L., 1987b, The pathology of heartwater. III. A review. *Onderstepoort J Vet Res* 54, 281-286.
- Qin, J., Peng, Z., McLeod, M.V., 2004, In vitro mutagenesis to define functional domains. *Methods Mol Biol* 241, 189-194.
- Raliniaina, M., Meyer, D.F., Pinarello, V., Sheikboudou, C., Emboule, L., Kandassamy, Y., Adakal, H., Stachurski, F., Martinez, D., Lefrancois, T., Vachiery, N., 2010, Mining the genetic diversity of *Ehrlichia ruminantium* using map genes family. *Veterinary parasitology* 167, 187-195.
- Ramsay, G., 1998, DNA chips: state-of-the art. *Nat Biotechnol* 16, 40-44.

- Renesto, P., Roveery, C., Schrenzel, J., Leroy, Q., Huyghe, A., Li, W., Lepidi, H., Francois, P., Raoult, D., 2008, *Rickettsia conorii* transcriptional response within inoculation eschar. PLoS one 3, e3681.
- Richards, A.L., 2004, Rickettsial vaccines: the old and the new. Expert Rev Vaccines 3, 541-555.
- Rikihisa, Y., Lin, M., 2010  
*Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* type IV secretion and Ank proteins. Current opinion in microbiology 13, 59-66.
- Rockett, J.C., Christopher Luft, J., Brian Garges, J., Krawetz, S.A., Hughes, M.R., Hee Kirn, K., Oudes, A.J., Dix, D.J., 2001, Development of a 950-gene DNA array for examining gene expression patterns in mouse testis. Genome Biol 2, RESEARCH0014.
- Rodionov, D.A., Gelfand, M.S., Todd, J.D., Curson, A.R., Johnston, A.W., 2006, Computational reconstruction of iron- and manganese-responsive transcriptional networks in alpha-proteobacteria. PLoS Comput Biol 2, e163.
- Rosadini, C.V., Wong, S.M., Akerley, B.J., 2008, The periplasmic disulfide oxidoreductase DsbA contributes to *Haemophilus influenzae* pathogenesis. Infection and immunity 76, 1498-1508.
- Sassetti, C.M., Boyd, D.H., Rubin, E.J., 2001, Comprehensive identification of conditionally essential genes in mycobacteria. Proceedings of the National Academy of Sciences of the United States of America 98, 12712-12717.
- Sassetti, C.M., Boyd, D.H., Rubin, E.J., 2003, Genes required for mycobacterial growth defined by high density mutagenesis. Molecular microbiology 48, 77-84.
- Savli, H., Karadenizli, A., Kolayli, F., Gundes, S., Ozbek, U., Vahaboglu, H., 2003, Expression stability of six housekeeping genes: A proposal for resistance gene quantification studies of *Pseudomonas aeruginosa* by real-time quantitative RT-PCR. Journal of medical microbiology 52, 403-408.
- Schmidt, H., Hensel, M., 2004, Pathogenicity islands in bacterial pathogenesis. Clin Microbiol Rev 17, 14-56.
- Schnappinger, D., Ehrt, S., Voskuil, M.I., Liu, Y., Mangan, J.A., Monahan, I.M., Dolganov, G., Efron, B., Butcher, P.D., Nathan, C., Schoolnik, G.K., 2003, Transcriptional Adaptation of *Mycobacterium tuberculosis* within Macrophages: Insights into the Phagosomal Environment. J Exp Med 198, 693-704.
- Sengupta, J., Agrawal, R.K., Frank, J., 2001, Visualization of protein S1 within the 30S ribosomal subunit and its interaction with messenger RNA. Proceedings of the National Academy of Sciences of the United States of America 98, 11991-11996.
- Smith, L.T., 1996, Role of osmolytes in adaptation of osmotically stressed and chill-stressed *Listeria monocytogenes* grown in liquid media and on processed meat surfaces. Applied and environmental microbiology 62, 3088-3093.
- Sumption, K.J., Paxton, E.A., Bell-Sakyi, L., 2003, Development of a polyclonal competitive enzyme-linked immunosorbent assay for detection of antibodies to *Ehrlichia ruminantium*. Clin Diagn Lab Immunol 10, 910-916.
- Takle, G.W., Toth, I.K., Brurberg, M.B., 2007, Evaluation of reference genes for real-time RT-PCR expression studies in the plant pathogen *Pectobacterium atrosepticum*. BMC plant biology 7, 50.

- Tampakaki, A.P., Fadouloglou, V.E., Gazi, A.D., Panopoulos, N.J., Kokkinidis, M., 2004, Conserved features of type III secretion. *Cellular microbiology* 6, 805-816.
- Tavernarakis, N., Driscoll, M., Kyrpidis, N.C., 1999, The SPFH domain: implicated in regulating targeted protein turnover in stomatins and other membrane-associated proteins. *Trends in biochemical sciences* 24, 425-427.
- Teysier, C., Marchandin, H., Jumas-Bilak, E., 2004, [The genome of alpha-proteobacteria : complexity, reduction, diversity and fluidity.]. *Can J Microbiol* 50, 383-396.
- Theriot, J.A., Rosenblatt, J., Portnoy, D.A., Goldschmidt-Clermont, P.J., Mitchison, T.J., 1994, Involvement of profilin in the actin-based motility of *L. monocytogenes* in cells and in cell-free extracts. *Cell* 76, 505-517.
- Tinsley, C.R., Voulhoux, R., Beretti, J.L., Tommassen, J., Nassif, X., 2004, Three homologues, including two membrane-bound proteins, of the disulfide oxidoreductase DsbA in *Neisseria meningitidis*: effects on bacterial growth and biogenesis of functional type IV pili. *The Journal of biological chemistry* 279, 27078-27087.
- Totsika, M., Heras, B., Wurpel, D.J., Schembri, M.A., 2009, Characterization of two homologous disulfide bond systems involved in virulence factor biogenesis in uropathogenic *Escherichia coli* CFT073. *Journal of bacteriology* 191, 3901-3908.
- Totte, P., Blankaert, D., Marique, T., Kirkpatrick, C., Van Vooren, J.P., Werenne, J., 1993a, Bovine and human endothelial cell growth on collagen microspheres and their infection with the rickettsia *Cowdria ruminantium*: prospects for cells and vaccine production. *Rev Elev Med Vet Pays Trop* 46, 153-156.
- Totte, P., De Gee, A.L., Werenne, J., 1993b, Role of interferons in infectious diseases in the bovine species: effect on viruses and rickettsias. *Rev Elev Med Vet Pays Trop* 46, 83-86.
- Totte, P., McKeever, D., Martinez, D., Bensaid, A., 1997, Analysis of T-cell responses in cattle immunized against heartwater by vaccination with killed elementary bodies of *Cowdria ruminantium*. *Infection and immunity* 65, 236-241.
- Totte, P., Nyanjui, J., Bensaid, A., McKeever, D., 1999, Bovine CD4+ T-cells lines reactive with soluble and membrane antigens of *Cowdria ruminantium*. *Vet Immunol Immunopathol* 70, 269-276.
- Tseng, T.T., Tyler, B.M., Setubal, J.C., 2009, Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC microbiology* 9 Suppl 1, S2.
- Tuanyok, A., Tom, M., Dunbar, J., Woods, D.E., 2006, Genome-wide expression analysis of *Burkholderia pseudomallei* infection in a hamster model of acute melioidosis. *Infection and immunity* 74, 5465-5476.
- Uilenberg, G., 1982, Experimental transmission of *Cowdria ruminantium* by the Gulf coast tick *Amblyomma maculatum*: danger of introducing heartwater and benign African theileriasis onto the American mainland. *Am J Vet Res* 43, 1279-1282.
- Uilenberg, G., 1983, Heartwater (*Cowdria ruminantium* infection): current status. *Adv Vet Sci Comp Med* 27, 427-480.
- Uilenberg, G., 1990, Methods currently used for the control of heartwater: their validity and proposals for future control strategies. *Parassitologia* 32, 55-62.
- Uilenberg, G., Camus, E., Barre, N., 1985, [A strain of *Cowdria ruminantium* isolated in Guadeloupe (French West Indies)]. *Rev Elev Med Vet Pays Trop* 38, 34-42.

- Vachier, N., Lefrancois, T., Esteves, I., Molia, S., Sheikboudou, C., Kandassamy, Y., Martinez, D., 2006, Optimisation of the inactivated vaccine dose against heartwater and in vitro quantification of *Ehrlichia ruminantium* challenge material. *Vaccine* 24, 4747-4756.
- Van Amstel, S.R., Oberem, P.T., 1987, The treatment of heartwater. *Onderstepoort J Vet Res* 54, 475-479.
- Van de Pypekamp, H.E., Prozesky, L., 1987, Heartwater. An overview of the clinical signs, susceptibility and differential diagnoses of the disease in domestic ruminants. *Onderstepoort J Vet Res* 54, 263-266.
- Van Heerden, H., Steyn, H.C., Allsopp, M.T., Zwegarth, E., Josemans, A.I., Allsopp, B.A., 2004, Characterization of the pCS20 region of different *Ehrlichia ruminantium* isolates. *Vet Microbiol* 101, 279-291.
- van Vliet, A.H., Jongejan, F., van der Zeijst, B.A., 1992, Phylogenetic position of *Cowdria ruminantium* (Rickettsiales) determined by analysis of amplified 16S ribosomal DNA sequences. *Int J Syst Bacteriol* 42, 494-498.
- van Vliet, A.H., Jongejan, F., van Kleef, M., van der Zeijst, B.A., 1994, Molecular cloning, sequence analysis, and expression of the gene encoding the immunodominant 32-kilodalton protein of *Cowdria ruminantium*. *Infection and immunity* 62, 1451-1456.
- Viljoen, G.J., Vermeulen, N.M., Neitz, A.W., 1987, Theoretical aspects of the enzyme-linked immunosorbent assay technique and its use in the detection of *Cowdria ruminantium* antigen and antibody in reacting animals. *Onderstepoort J Vet Res* 54, 305-312.
- Wakeel, A., Zhu, B., Yu, X.J., McBride, J.W., 2010, New insights into molecular *Ehrlichia chaffeensis*-host interactions. *Microbes Infect*.
- Walker, D.H., Feng, H.M., Popov, V.L., 2001, Rickettsial phospholipase A2 as a pathogenic mechanism in a model of cell injury by typhus and spotted fever group rickettsiae. *The American journal of tropical medicine and hygiene* 65, 936-942.
- Walker, D.H., Yu, X.J., 2005, Progress in rickettsial genome analysis from pioneering of *Rickettsia prowazekii* to the recent *Rickettsia typhi*. *Annals of the New York Academy of Sciences* 1063, 13-25.
- Walker, J.B., Olwage, A., 1987, The tick vectors of *Cowdria ruminantium* (Ixodoidea, Ixodidae, genus *Amblyomma*) and their distribution. *Onderstepoort J Vet Res* 54, 353-379.
- Wandersman, C., Delepelaire, P., 2004, Bacterial iron sources: from siderophores to hemophores. *Annual review of microbiology* 58, 611-647.
- Wang, C., Zhou, G.L., Vedantam, S., Li, P., Field, J., 2008, Mitochondrial shuttling of CAP1 promotes actin- and cofilin-dependent apoptosis. *J Cell Sci* 121, 2913-2920.
- Wang, G., Maier, R.J., 2009, A RecB-like helicase in *Helicobacter pylori* is important for DNA repair and host colonization. *Infection and immunity* 77, 286-291.
- Wells, M.Y., Rikihisa, Y., 1988, Lack of lysosomal fusion with phagosomes containing *Ehrlichia risticii* in P388D1 cells: abrogation of inhibition with oxytetracycline. *Infection and immunity* 56, 3209-3215.
- Wilhelm, S., Tommassen, J., Jaeger, K.E., 1999, A novel lipolytic enzyme located in the outer membrane of *Pseudomonas aeruginosa*. *Journal of bacteriology* 181, 6977-6986.
- Winkler, H.H., Daugherty, R.M., 1989, Phospholipase A activity associated with the growth of *Rickettsia prowazekii* in L929 cells. *Infection and immunity* 57, 36-40.

- Zeller, T., Klug, G., 2006, Thioredoxins in bacteria: functions in oxidative stress response and regulation of thioredoxin genes. *Die Naturwissenschaften* 93, 259-266.
- Zweygarth, E., Josemans, A.I., Van Strijp, M.F., Lopez-Rebollar, L., Van Kleef, M., Allsopp, B.A., 2005, An attenuated *Ehrlichia ruminantium* (Welgevonden stock) vaccine protects small ruminants against virulent heartwater challenge. *Vaccine* 23, 1695-1702.

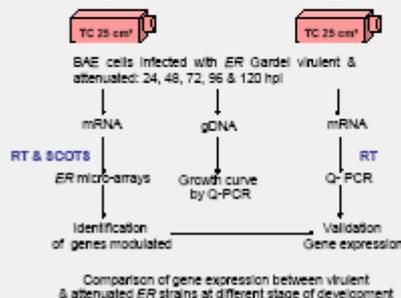
**Posters présentés dans les congrès internationaux au cours de la  
thèse**

**« FEMS 2009 et TTP6 2008 »**

## Introduction

Rickettsial *Ehrlichia ruminantium* (ER), the causative agent of heartwater, is an obligate intracellular pathogen (1). This characteristic explains that little is known about the genetic determinants and molecular mechanisms of ER pathogenesis. Recently a method of selective capture of transcripts sequences (SCOTS) (2) was successfully adapted to ER. This step was essential for any future transcriptomic study on obligate intracellular pathogen. The annotation of 2 genomes of Gardel and Weigevonden strains revealed that 30% of the CDS had unknown functions (3). The objectives of this study was to do high throughput gene expression analysis through the development of ER microarrays and then identify genes involved in the pathogenesis and attenuation of ER strains and comprehension of these mechanisms. An *in vitro* model using both virulent and attenuated Gardel strains in bovine endothelial cells allowed us to compare for the first time ER gene expression. In this study, we reported a set of genes differentially expressed between virulent and attenuated strains of ER Gardel.

### Experimental model & Strategy



### Growth kinetics of ER strains by Q-PCR

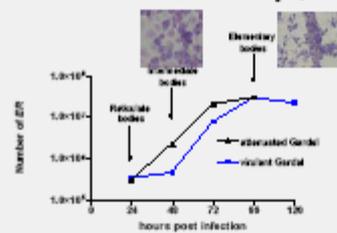


Figure 1: Q-PCR of target gene on gDNA samples

Optical microscopy observations of ER development depending on time of collect & following up of ER growth by Q-PCR. The kinetics of both strains showed typical bacteria growth curve and allowed us to chose the different stage of development in association with post infection time.

### map1, recA and rpoD expression after SCOTS



Figure 2: Q-RT-PCR of map1, recA and rpoD genes on ER gDNA samples before and after 3 rounds of capture. The level of expression was different depending on gene: 1 copy for rpoD and 10<sup>3</sup> copies for map1 at 24hpi. SCOTS after 3 capture allowed to detect a higher number of copies for each time post infection. A difference of expression between genes was maintained after capture (1 log10).

### ER Micro-arrays validation

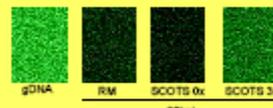


Figure 3: Hybridization of ER gDNA, cDNA after RNasease (RM), SCOTS 0x and SCOTS 3x at 96 hpi

### Hybridization on ER micro-arrays :

- ✓ for gDNA: 59% of detected probes
- ✓ for cDNA after removal of eucaryotic rRNA (RM): 1% at 24hpi & 20% at 96hpi
- ✓ for cDNA without capture: <1% at 24hpi & 40% of detected probes at 96hpi
- ✓ for cDNA after 3 captures: 24% at 24 hpi & 81% of detected probes at 96hpi

ER gDNA hybridization demonstrated the efficiency and specificity of ER micro-arrays. SCOTS method is crucial to increase the quantity of transcripts in order to measure gene expression at early stage of development. At late stage, the percentage of genes detected doubled after capture.

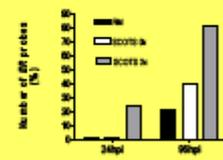


Figure 4: Detection of ER transcripts by micro-arrays

### Genes differentially expressed between virulent & attenuated strains

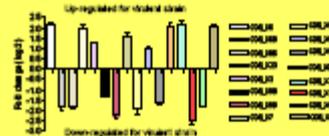


Figure 5: Fold expression (log 2) of gene between virulent and attenuated Gardel strains by micro-arrays

ID	Ratio	Function	Category
CD8_X8	2.18	ATP1 transport	pathogenicity
CD8_X29	-1.91	SDS released protein	
CD8_X89	-1.92	Signal transduction mechanism	
CD8_X23	1.97	-	
CD8_X3	1.97	-	metabolism
CD8_X5	-1.98	-	
CD8_X89	-1.98	Hypothetical transcriptional regulator	
CD8_X7	1.98	Hypothetical extracellular metalloprotease involved in virulence transport	
CD8_X38	1.92	Hypothetical major virulence proteins	hypothetical
CD8_X88	-1.98	Hypothetical pathogen virulence factors	
CD8_X54	-1.97	Hypothetical transcription and metabolism	
CD8_X8	1.91	-	
CD8_X7	1.94	Secondary metabolite biosynthetic transport mechanism	metabolism
CD8_X5	-1.98	Energy production and conversion	
CD8_X27	1.94	-	
CD8_X89	2.06	-	
CD8_X79	2.14	Unknown	hypothetical
CD8_X14	-2.01	-	
CD8_X39	-1.94	-	
CD8_X71	2.04	-	
CD8_X2	-2.17	-	hypothetical
CD8_X45	-2.42	-	
CD8_X39	-1.77	-	
CD8_X39	-1.77	-	

Table 1: 22 Genes differentially expressed between virulent and attenuated strains of ER Gardel at 96hpi. Ratio is expressed in log2.

34 genes with modulated expression at 96hpi: 4 genes involved in virulence for intracellular pathogens *S. Enterica typhimurium*, *S. flexneri*, *M. tuberculosis*; 3 genes potentially involved in pathogenicity; 19 genes with unknown function; 6 genes involved in the metabolism

## Discussion & Perspectives

The adaptation of SCOTS method to our *in vitro* model, allowed to obtain a large amount of ER transcripts and do further transcriptomic study. For the first time, ER micro-arrays was validated by hybridization of ER gDNA and cDNA. SCOTS method allowed the detection of the lowest expressed genes such as rpoD or recA. There was a strong increase of detection of transcripts after SCOTS both demonstrated by Q-RT-PCR and micro-array analysis. We demonstrated that SCOTS method was crucial for expression analysis during early stationary phase of growth of ER (24 hpi) with 24% of transcripts detected compare to 1% of transcripts before capture. Preliminary results on ER micro-arrays at late stage of development demonstrated that 34 genes were strongly modulated depending on virulent or attenuated strains. 19 of the genes were significantly up-regulated for virulent strain. Through comparative genomic study, on 34 genes, 19 had unknown function and 6 were involved in metabolism pathway. Interestingly, 4 genes were identified to be virulent factors for other intracellular pathogens and 2 as hypothetical virulent factor specific of ER. Three of these genes should be strongly involved in the virulence (CD8\_X23, CD8\_X3, CD8\_X7; up-regulated for virulent strain) and 3 involved in the attenuation (CD8\_X86, CD8\_X88, CD8\_X89; down-regulated). The differential gene expression is under validation by Q-RT-PCR. Further analysis on ER earlier stage of development using SCOTS method are under process. This strategy paves the way for new insights in pathogenicity of obligate intracellular pathogens.

### References

1. Scaevolone et al., 1995, Onderstepoort J Vet Res 62, 119-120.
2. Faucher et al., PNAS 2006, Vol 103, n° 6, p1906-1911.
3. Fobos et al., Journal of Bacteriology, 2006, p2523-2542.



Centre de coopération internationale en recherche agronomique pour le développement

## Adaptation of SCOTS method for *Ehrlichia ruminantium*: Innovative approach for transcriptomic analysis of obligatory intracellular pathogen

EMBOULE L<sup>1</sup>, DAIGLE F<sup>2</sup>, FRUTOS R<sup>3</sup>, MEYER D<sup>1</sup>, SHEIKBOUDOU C<sup>1</sup>, PINARELLO V<sup>1</sup>, MARTINEZ D<sup>3</sup>, LEFRANÇOIS T<sup>1</sup>, VACHIER Y N<sup>1</sup>.

<sup>1</sup>UMR 15 CIRAD/INRA, Site de Duclos, Frise d'Eau, 97170 Petit Bourg Guadeloupe, loic.emboule@cirad.fr

<sup>2</sup>Département de microbiologie et immunologie, Université de Montréal, Canada

<sup>3</sup>CIRAD, Montpellier, France.

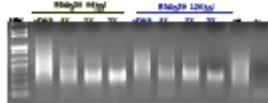
### Introduction

The Rickettsia *Ehrlichia ruminantium* ER is the causative agent of heartwater, which affects wild and domestic ruminants and is transmitted by ticks of the genus *Amblyomma*. Recently, the genomic organisation of this microorganism was revealed by the sequencing of two strains: Gardel and Weigevonden (1). Genes involved in virulence, host cell penetration or invasion and intracellular growth process are still unidentified. Transcriptomic analysis using high throughput microarray analysis and Q RT-PCR will be done in order to identify the genes involved in the attenuation and the ER pathogenesis. A large amount of RNA from virulent and attenuated Gardel strains can be obtained using an *in vitro* model (bovine endothelial cells). Due to the obligate intracellular properties of ER, the main obstacle to do accurate transcriptomic study is the production of purified ER transcripts without host cell contaminant. Selective capture of transcribed sequences (SCOTS) was originally developed by Graham and Clark-Curtis in 1989 for the non-obligatory intracellular pathogen *Mycobacterium tuberculosis* (2), and was further used for transcriptomic microarray analysis (3). In this study, we report for the first time the successful adaptation of the SCOTS method to an obligate intracellular bacterium.

### Nkpn1 PCR after SCOTS

After each capture, a PCR using Nkpn1 primer was performed to confirm the efficiency of the cDNA hybridization to gDNA and then the selective captures. Figure 1 shows amplicons after Nkpn1 PCR obtained for Gardel passage 39 at 96 and 120 hpi. There is a diminution of the size of the amplified transcribed sequences after each capture.

Fig. 1. Nkpn1 PCR products on cDNA at 96 and 120hpi before and after 1, 2 or 3 rounds of SCOTS

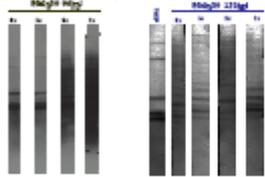


1X, 2X and 3X: Amplification signals correspond to results obtained by RT-PCR on cDNA after one, two and three rounds of captures.  
cDNA: Amplification signals correspond to results obtained by RT-PCR on total cDNA generated before SCOTS.  
M: Molecular weight 100bp DNA ladder

### Enrichment of ER cDNA after SCOTS

Southern blots were done on cDNAs before and after each capture for each time of infection for virulent Gardel. Figure 2 shows southern blots obtained using cDNA at 96 hpi and 120 hpi as probes on ER gDNA. DIG-gDNA 238+59 and 16S are used to reveal the bands corresponding to ER rDNA (lane 1). We observe a decrease of contaminant rRNA after the second and third capture. After successive capture, there is a progressive increase of bacterial cDNA in diversity and in amount. Similar results were obtained at different time of infection.

Fig. 2. Southern blot using DIG labeling cDNA at 96hpi and 120hpi before and after SCOTS



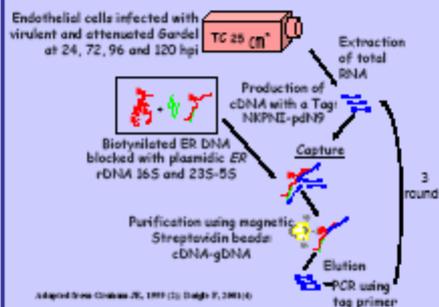
1X, 2X and 3X: Southern blot showing bacterial cDNA obtained from endothelial cells after one, two and three rounds of SCOTS respectively at 96hpi and 120hpi.  
c: Southern blot showing bacterial cDNA before SCOTS in initial sample at 96hpi and 120hpi respectively.  
cDNA: Southern blot showing ER ribosomal DNA 16S, 23S and 5S hybridization probes.



### Results



### SCOTS method for production of cDNA samples from ER



Adapted from Graham JC, 1989 (2); Daigle F, 2001 (4)

Adaptation of SCOTS from Salmonella to ER model:  
- Temperature of hybridization 50°C;  
- Modification of hybridization buffer components

### Detection of ER 16S by RTPCR after SCOTS

Detection of ER 16S by RTPCR to assess the decrease in contaminant transcripts: There was a significant diminution of 16S contaminant following successive capture for 96 and 120 hpi (table 2).

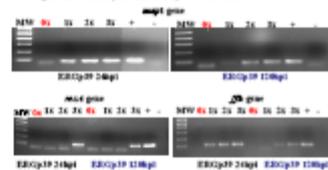
Table 2. Quantitative RTP-PCR of specific ribosomal 16S gene of ER

Samples	16S rDNA copy		Cycle threshold
	Before SCOTS	After SCOTS	
96	1x	1680	24
hpi	2x	2694	23
	3x	1265	24
120	Before SCOTS	2995	23
1x	1090	24	
hpi	2x	92	26
3x	42	26	

### Detection of ER transcripts by RTPCR after SCOTS

Detection of transcripts of 5 ER genes were done following successive capture to assess the increase in ER specific transcripts (Figure 3 and Table 1)  
- there was no detection of ER transcripts before capture for all genes  
- there was an increase of the amount of cDNA for each target genes after each capture  
- 3 rounds of capture were necessary to obtain a maximal amount of cDNAs for variable expressed genes (table 1)

Fig. 3. ER transcripts detected by RT-PCR after SCOTS



1X, 2X and 3X: amplification signals correspond to results obtained by RT-PCR on cDNA after one, two and three rounds of captures.  
c: amplification signals correspond to results obtained by RT-PCR on total cDNA generated before SCOTS  
M: Molecular weight 100bp DNA ladder

Table 1. Detection of ER specific genes by RT-PCR on total cDNA before / after SCOTS

Gene	96hpi			120hpi			96hpi			120hpi		
	1X	2X	3X	1X	2X	3X	1X	2X	3X	1X	2X	3X
map1	-	+	+	-	+	+	-	+	+	-	+	+
recA	-	+	+	-	+	+	-	+	+	-	+	+
fh	-	+	+	-	+	+	-	+	+	-	+	+
pCG20	-	+	+	-	+	+	-	+	+	-	+	+
rpoD	-	+	+	-	+	+	-	+	+	-	+	+
cDNA	-	+	+	-	+	+	-	+	+	-	+	+

(-) : absence of amplicon after RTPCR.

(+) : presence of amplicon after RTPCR.

1X, 2X and 3X correspond to results obtained on cDNA after 1, 2 or 3 rounds of captures.

### References

- Frutos et al., Journal of Bacteriology, 2006, p2023-2042.
- James E. Graham and Jacqueline B. Clark-Curtis, PNAS, 1989, Vol 86, p 11554-11558.
- Faucier et al., PNAS, 2006, Vol 103, n° 9, p1906-1911.
- Daigle et al., Molecular Microbiology, 2001, 41(5), p1211-1220.
- Daigle et al., Veterinary Microbiology, 2001, n°123, p119-121.
- Doores et al., 2000, PNAS, 100, p447-452.

### Discussion & Perspectives

The innovative method of SCOTS which has been used previously for non-obligatory intracellular parasites (4,5,6) was adapted to our ER model for further transcriptomic analysis. Hybridization temperature and hybridization buffer which conditioned an efficient hybridization in stringent conditions were modified. The diminution of the size of transcripts for Nkpn1 PCR following successive captures of ER was observed as previously for Salmonella (4). We observed by Southern blot a progressive enrichment of ER cDNA between first and third capture. There was also a diminution of ribosomal contaminant after capture as shown by QRT-PCR targeting 16S transcripts. Increased amounts of 5 different ER transcripts were observed by RTPCR following successive capture. Furthermore, this study highlighted the differential expression of map1, RecA, fh, pCG20, rpoD since variable number of captures were necessary to detect specific transcripts. Map1 was strongly expressed whereas pCG20 was the lowest expressed gene independently of the time of infection.

Both RTPCR results and southern blot validated the successful adaptation of SCOTS method to ER and should enable us to perform transcriptomic study on attenuated and virulent Gardel strains. This should lead to the identification of new virulence factors.

## Résumé

La bactérie intracellulaire obligatoire *Ehrlichia ruminantium* (ER) est l'agent pathogène responsable de la cowdriose ou heartwater chez les ruminants domestiques et sauvages. Pour élucider la pathogenèse de la bactérie dans les cellules endothéliales, l'analyse du transcriptome a été entreprise en comparant le niveau d'expression des souches Gardel virulente et atténuée dont les génomes ont été séquencés. La sélection des transcrits bactériens a été réalisée grâce à l'adaptation de la méthode SCOTS à notre modèle. Les cDNA obtenus par SCOTS ont été hybridés sur des puces à ADN pangénomique d'ER spécifiques des souches Gardel et Welgevonden. Cette étude a révélé 239 gènes différentiellement exprimés entre souche virulente et atténuée dans les cellules infectées aux 3 stades évolutifs étudiés. 116 gènes ont été détectés au temps le plus tardif de l'infection (corps élémentaires libres). Parmi les gènes surexprimés par la souche virulente, certains sont connus chez d'autres bactéries intracellulaires (*R. conorii*) pour leurs implications dans l'échappement à la réponse de l'hôte : réparation de l'ADN (helicase, ligase), échappement au stress oxydatif et osmotique (thioredoxine), ainsi que dans l'inhibition de la fusion des endosomes (SNARE) et dans le ralentissement de l'apoptose (Ank A). Parmi les gènes surexprimés chez la souche atténuée, certains peuvent induire une réponse immunitaire protectrice. En combinant les données génomiques des groupes de gènes potentiellement co-régulés ont été identifiés notamment le « cluster » *recB*. La validation des résultats par qRT-PCR et l'analyse du protéome d'ER sont en cours, ce qui nous offre une perspective d'analyse fonctionnelle.

**Mots clés :** *E. ruminantium*, intracellulaire, diversité génétique, SCOTS, transcriptome, vaccin

## Summary

The obligatory intracellular bacterium *Ehrlichia ruminantium* (ER) is a pathogen responsible for heartwater which affects in wild and domestic ruminants. To elucidate the pathogenesis of ER in endothelial cells, a high throughput analysis of its transcriptome was done. A comparison of gene expression between two sequenced strains was done: Gardel virulent and attenuated. Selective capture of ER transcribed sequences was done using the SCOTS method which has been adapted to our model. The cDNA were hybridized on pangenomic ER microarrays specific of both Gardel and Welgevonden strains. 239 genes were identified to be differentially expressed between virulent and attenuated Gardel strains in host cells within the three developmental stages studied. 116 genes were detected at the late stage corresponding to free elementary bodies. Among over-expressed genes identified for virulent strain, some genes have been described previously in other intracellular bacteria (*R. conorii*) to be involved in the host cell immune response escape. They were mainly involved in DNA repair (helicase, ligase), diminution of the oxidatif (thioredoxine) and osmotic stress, the inhibition of endosome fusion (SNARE) and diminution of apoptosis (Ank A). Some interesting genes over-expressed for attenuated strain could induce a protective immune response. Genomic and transcriptomic comparison of data from both strains allowed identification of several cluster of genes potentially co-regulated such as the cluster *recB*. The qRT-PCR validation of differential gene expression identified by microarrays is currently done. In parallel, ER proteomic analysis paves the way for further functional analysis.

**Key words:** *E. ruminantium*, intracellular, genetic diversity, SCOTS, transcriptome, vaccine