



Système de sécrétion de type IV et protéines à domaines ankyrines dans les interactions Wolbachia-arthropodes

S. Pichon

► To cite this version:

S. Pichon. Système de sécrétion de type IV et protéines à domaines ankyrines dans les interactions Wolbachia-arthropodes. Ecologie, Environnement. Université de Poitiers, 2009. Français. NNT : . tel-00551985

HAL Id: tel-00551985

<https://theses.hal.science/tel-00551985>

Submitted on 5 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

POUR L'OBTENTION DU GRADE DE
DOCTEUR DE L'UNIVERSITE DE POITIERS
FACULTE DES SCIENCES FONDAMENTALES ET APPLIQUEES
(Diplôme National : arrêté du 7 août 2006)

Ecole doctorale : Gay-Lussac
Secteur de Recherche : Biologie de l'environnement, Biologie des Populations, Ecologie

PRESENTEE PAR

Samuel PICHON

Système de sécrétion de type IV et protéines à domaines ankyrines dans les interactions *Wolbachia*-arthropodes

Directeurs de thèse : Professeur **Pierre GREVE**
Professeur **Didier BOUCHON**

Soutenue publiquement le 15 décembre 2009
devant la Commission d'Examen

JURY

David O'CALLAGHAN	Directeur de Recherche, INSERM U-431, Nîmes	Rapporteur
Patrick MAVINGUI	Chargé de Recherche, UMR CNRS 5557, Lyon	Rapporteur
Yann HECHARD	Maître de Conférences, UMR CNRS 6008, Poitiers	Examineur
Elisabeth HUGUET	Maître de Conférences, UMR CNRS 6035, Tours	Examinatrice
Pierre GREVE	Professeur, UMR CNRS 6556, Poitiers	Examineur
Didier BOUCHON	Professeur, UMR CNRS 6556, Poitiers	Examineur

FICHER PDF

Si vous souhaitez imprimer ma thèse à partir de ce fichier pdf, veuillez noter que l'impression est recommandée en recto-verso et que la couleur n'est essentielle que pour les figures des pages 2, 4, 14, 16, 20, 24, 30, 41, 45, 47, 50, 87, 122 du manuscrit (soit les pages 26, 28, 42, 44, 52, 62, 72, 99, 103, 105, 110, 155, 202 de ce pdf). Merci de citer mon travail de la façon suivante :

Pichon S. 2009. Système de sécrétion de type IV et protéines à domaines ankyrines dans les interactions *Wolbachia*-arthropodes. Thèse de doctorat d'Etat, Université de Poitiers, 164 pages.

If you print my thesis manuscript from this file, please note that you can make a duplex printout and that colour is only necessary for pages 2, 4, 14, 16, 20, 24, 30, 41, 45, 47, 50, 87, 122 (*i.e.* pages 26, 28, 42, 44, 52, 62, 72, 99, 103, 105, 110, 155, 202 of this pdf file). Please cite this manuscript this way:

Pichon S. 2009. Type IV secretion system and ankyrin domain-containing proteins in *Wolbachia*-arthropods interactions [in French]. Ph.D Thesis, University of Poitiers, France. 164 pp.

REMERCIEMENTS

A l'heure où j'écris ces remerciements, il est fort possible que j'oublie de citer les noms de certaines personnes qui ont contribué de près ou de loin à ma thèse. J'en suis désolé par avance et n'ai que pour excuse qu'une thèse est une longue épreuve !

Une thèse, c'est un diplôme obtenu après un oral et qui met fin au statut d'étudiant. A cet effet, je remercie en premier lieu les membres de mon jury, David O'Callaghan, Patrick Mavingui, Yann Héchard et Elisabeth Huguet; merci d'avoir accepté d'évaluer mon travail, merci aussi d'avoir été présents lors des deux comités de thèse et de m'avoir prodigué suggestions et encouragements.

Une thèse, c'est également un CDD encadré par deux directeurs. Un grand merci donc à Didier Bouchon et à Pierre Grève pour votre soutien, votre confiance, votre compréhension, le partage de vos connaissances et de votre expérience, la liberté que vous m'avez accordée. Merci de croire en moi. Enfin, un merci plus particulier à Pierre qui m'a supporté dans son bureau (moi aussi je l'ai supporté !!). Merci pour ton dynamisme et ton humour.

Une thèse, c'est donc prendre temporairement le métier de chercheur. Tout cela n'aurait été agréable sans la compagnie de tous les membres de l'équipe « Ecologie, Evolution, Symbiose ». J'ai vu des anciens partir à la retraite, d'autres plus jeunes achever leur thèse ou leurs stages mais j'ai plus souvent vu les gens revenir avec joie, ne serait-ce que pour dire bonjour. A se demander même si certains n'ont pas du mal à couper le cordon ! Pas étonnant ! Vous avez tous largement contribué à mon épanouissement personnel, mon évolution en tant que scientifique et êtes devenus comme une seconde famille pour moi. J'envoie un merci particulier « aux filles » (Alex, Carine, Catherine et Maryline) et Jérôme du staff technique pour leur bonne humeur et leur soutien inconditionnel. Merci à toi aussi Christelle pour me rabibocher avec la paperasse, ce qui n'est pas chose gagnée. Et puis, sur le plan scientifique, je dois aussi beaucoup à Richard Cordaux qui a contribué à ma formation et ma réussite. Une thèse, c'est un sujet de recherche que je me suis largement approprié mais auquel beaucoup ont participé. Merci donc à Sylvaine Cheng et Christine Félix pour leur travail. Merci également aux deux étudiants, Isabelle Mazerie et Hugues de Verdal, que j'ai encadré et qui sont devenus des amis et thésards à leur tour. J'en profite pour remercier tous les stagiaires comme Aline, Clémence, Fred, Laura, Marion, Maureen, Sébastien et Vincent qui m'ont rappelé qu'ambition, travail et hargne étaient sans cesse de rigueur.

Une thèse, ce sont des collaborations et des échanges avec d'autres professionnels. Je dois donc beaucoup à Denis Poinso, Hervé Merçot, Kostas Bourtzis, Panagiotis Ioannidis, Roger Garrett, Chao Liu, Kim Brugger, Paul Hooykaas, Amke den Dulk-Ras, Annette Vergunst, Anne Cantereau, Thierry Bergès et Daniel Guyonnet.

Une thèse, c'est également dans mon cas un « cumul de mandat » avec la fonction temporaire d'enseignant à l'Université. Je remercie les étudiants qui ont su me rendre la pareille des efforts fournis. Ces points positifs n'auraient pu être acquis sans l'accompagnement de tous les enseignants-chercheurs du labo. Merci à vous tous qui formez une équipe soudée. Merci aussi à Tifany Laverre pour ton aide aux TP, et à Vincent Leconte, Amélie Dreiss et autres jeunes intervenants pour le partage d'expériences.

Une thèse, c'est aussi suivre la trace d'anciens thésards et partager avec les actuels. Merci donc à Benjamin, Fred, Gaël, Isabelle, Juline, Marie-Cécile, Marion, Nicolas, Sébastien (x2), Vincent et Wincka qui m'ont permis d'encore mieux vivre ma thèse grâce à de nombreuses soirées mémorables. Quand je repense aux soirées, j'ai une pensée toute particulière pour les étudiants étrangers que j'ai pu côtoyer. Merci donc à Aline, Alison, Dalila, Gipo, Lenka, Momo et Tabata.

Une thèse, ce n'est pas toujours une vocation mais parfois la suite logique après de bonnes expériences de stages réalisés en Licence et en Master. Je remercie donc très chaleureusement Dave Goulson et Mick Hanley (Southampton), Bernard Vaissière (Avignon) et Elisabeth Huguet (Tours) grâce à qui j'ai beaucoup appris et me suis construit scientifiquement.

Une thèse, c'est par ailleurs l'aboutissement d'un parcours universitaire toujours soutenu par ma famille et amis proches. Je ne peux finir ces remerciements sans citer celle qui a suivi toutes les étapes et facettes de ma vie de thésard. Il a fallu parfois que je fasse des concessions mais bien souvent tu as participé au bon déroulement de ma thèse et des « à côtés ». Merci du fond du cœur Elodie.

Système de sécrétion de type IV et protéines à domaines ankyrines dans les interactions *Wolbachia*-arthropodes.

Wolbachia est une bactérie Gram-negative intracellulaire modifiant la reproduction de nombreux arthropodes. Chez l'isopode *Armadillidium vulgare*, la souche *wVulC* entraîne la féminisation des mâles. Nous avons caractérisé deux opérons *vir* s'exprimant dans tous les tissus hôtes et codant un système de sécrétion de type IV (T4SS) pouvant permettre d'exporter des effecteurs bactériens vers le cytoplasme de l'hôte. La comparaison des séquences et de l'organisation des gènes de 37 souches de *Wolbachia* a révélé la forte conservation des deux opérons *vir* suggérant l'importance du T4SS dans la biologie de la bactérie. Nous avons également identifié, dans le génome en cours de séquençage de *wVulC*, 66 gènes codant des protéines à domaines ankyrines. Ces motifs forment des sites d'interactions protéine-protéine chez les eucaryotes et sont supposés être impliqués chez *Wolbachia* dans l'interaction avec des protéines de l'hôte. Nous avons montré qu'une des trois copies du gène *pk2* de *wVulC*, n'est exprimée que chez des souches féminisantes mais chez aucune des 3 souches induisant l'incompatibilité cytoplasmique chez les isopodes terrestres. Ce produit du gène *pk2* pourrait être impliqué dans la féminisation de l'hôte. Toutefois, nous avons réalisé des tests d'interaction par double-hybride en levures et par la méthode CRAFT (Cre-recombinase Reporter Assay for Translocation) entre les protéines du T4SS et cinq protéines à domaines ankyrines dont Pk2 afin de savoir si ces dernières étaient sécrétées par ce système. Les résultats montrent qu'aucun des cinq produits de gènes *ank* testés n'est sécrété par la bactérie mais se révèlent encourageants pour identifier les effecteurs de *Wolbachia*.

Type IV secretion system and ankyrin domain-containing proteins in *Wolbachia*-arthropods interactions.

Wolbachia are intracellular Gram-negative bacteria that are reproductive manipulators of many arthropods. In the isopod *Armadillidium vulgare*, the *Wolbachia wVulC* strain induces male feminization. Here, we characterized two *vir* operons which are expressed in all host tissues and which encode a type IV secretion system (T4SS) used to translocate bacterial effectors into host cytoplasm. Gene organization and sequence comparison in 37 *Wolbachia* strains highlighted the high conservation of both *vir* operons and their importance for the biology of the bacteria. We also identified in the on-going assembly of the *wVulC* genome, 66 ankyrin domain-encoding genes. Ankyrin motifs are known to mediate protein-protein interactions in eukaryotic organisms and thus are suggested to mediate in *Wolbachia* the interaction with host molecules. We showed that one of the three copies of the *wVulC pk2* gene is only expressed in feminizing strains but not in the three strains inducing cytoplasmic incompatibility in terrestrial isopods. The associated Pk2 protein could be involved in male feminization. We thus tested the interaction between three T4SS proteins and five ankyrins (including Pk2) via the yeast two-hybrid and CRAFT (Cre-recombinase Reporter Assay for Translocation) methods. None of the five ankyrin proteins were revealed to be secreted by the *wVulC* strain. Nevertheless, this promising approach may enable us to identify *Wolbachia* effectors.

Discipline : Biologie évolutive

Mots-clés : Interaction hôte-parasite, *Wolbachia*, Système de sécrétion, Protéines à domaines ankyrines, Féminisation, Incompatibilité cytoplasmique, Effecteurs bactériens, Crustacés isopodes.

Key words : Host-parasite interaction, *Wolbachia*, Secretion System, Ankyrin domain-containing proteins, Feminization, Cytoplasmic incompatibility, Bacterial effectors, Crustacean isopod.

Intitulé et adresse du laboratoire

Laboratoire Ecologie, Evolution, Symbiose - UMR CNRS 6556
Bâtiment Sciences Naturelles, 1er étage, Université de Poitiers
40, avenue du Recteur Pineau - 86022 Poitiers Cedex – FRANCE.

Système de sécrétion de type IV et protéines à domaines ankyrines dans les interactions *Wolbachia*-arthropodes.

Wolbachia est une α -protéobactérie intracellulaire obligatoire transmise à la descendance par voie ovocytaire. Elle a été identifiée chez ~65% des arthropodes et chez certaines espèces de nématodes ce qui en fait l'endosymbiote le plus répandu sur Terre. *Wolbachia* modifie la reproduction de son hôte et favorise ainsi sa transmission dans les populations. Chez trois espèces d'isopodes terrestres, *Wolbachia* provoque une incompatibilité cytoplasmique (IC). Chez la plupart des autres espèces d'isopodes, la présence de *Wolbachia* entraîne une féminisation des mâles génétiques en femelles fonctionnelles. Les mécanismes moléculaires de la féminisation restent à ce jour inconnus mais on pense que *Wolbachia* s'oppose à la différenciation de la glande androgène, normalement responsable de la mise en place des caractères sexuels mâles.

La localisation exclusivement cytoplasmique de *Wolbachia* suggère qu'elle puisse sécréter des facteurs de virulence afin d'induire la féminisation des mâles. Or, de nombreuses bactéries Gram-négatives utilisent un système de sécrétion de type IV (T4SS) pour délivrer des effecteurs dans le cytoplasme des cellules hôtes. Ce système, largement conservé chez de nombreuses bactéries pathogènes, a été mis en évidence chez plusieurs souches de *Wolbachia*. Nous avons identifié huit gènes codant un T4SS dans la souche de *Wolbachia* (wVulC) présente chez *Armadillidium vulgare*. Ces gènes s'organisent en deux opérons localisés sur deux loci distincts du génome. Ils sont exprimés dans tous les tissus de l'hôte. Une étude comparative de l'évolution de ces gènes nous a permis de déterminer le haut degré de conservation de leur séquence chez 37 souches de *Wolbachia* induisant différents phénotypes. Ceci suggère fortement que le T4SS pourrait jouer un rôle dans la survie, la transmission et/ou l'expression phénotypique de la bactérie.

Un nombre important de gènes codant des protéines à domaines ankyrines a été mis en évidence dans le génome de différentes souches de *Wolbachia*. Comme les ankyrines possèdent des sites d'interactions protéine-protéine, elles constituent de bons candidats bactériens pouvant interagir avec des protéines de l'hôte. Nous avons identifiés un minimum de 66 gènes *ank* dans le génome de wVulC dont l'assemblage est en cours. Cela en fait le génome bactérien le plus riche en gènes *ank* à ce jour. Nous avons étudié plus spécifiquement l'évolution, la distribution et la transcription des deux gènes ankyrines *pk1* et *pk2* dont le niveau d'expression est connu pour être corrélé au phénotype d'IC chez *Culex pipiens*. Nous montrons qu'une copie du gène *pk2* ne s'exprime que chez 5 souches féminisantes mais chez aucune des 3 souches induisant l'IC chez les isopodes terrestres. Le produit du gène *pk2* pourrait donc être impliqué dans l'effet féminisant.

Wolbachia est une bactérie difficilement manipulable hors de son hôte ce qui rend impossible la production de mutants. Toutefois, nous avons pu réaliser des tests d'interaction par double-hybride en levures et CRAfT (Cre-recombinase Reporter Assay for Translocation) entre les protéines du T4SS et les protéines à domaines ankyrines Pk1 et Pk2 afin de savoir si ces dernières étaient sécrétées par ce système. Ces tests se sont révélés infructueux ce qui suggère qu'aucun des 2 produits de gènes *ank* n'est sécrété par la bactérie. Le génome de la souche féminisante wVulC en comparaison avec ceux des souches de *Wolbachia* induisant l'IC pourrait révéler de nouveaux candidats susceptibles d'être impliqués dans la survie, le maintien ou l'expression de la bactérie.

Mots clefs : Interaction hôte-parasite, *Wolbachia*, T4SS, ankyrines, féminisation, Crustacés Isopodes.

SOMMAIRE

Liste des illustrations	i
Liste des abréviations	ii
Avant-propos : <i>Wolbachia</i> et ses applications.	iv
I. Introduction.	1
A. <i>Wolbachia</i>, un symbiote manipulateur.	3
A.1 - Biologie de <i>Wolbachia</i> .	3
A.2 - <i>Wolbachia</i> et les associations obligatoires.	6
A.3 - <i>Wolbachia</i> et les associations facultatives.	6
a) L'incompatibilité cytoplasmique.	8
b) La parthénogenèse.	8
c) Le male-killing.	10
d) La féminisation.	10
e) La transition vers le mutualisme.	13
B. La communication entre <i>Wolbachia</i> et ses hôtes.	15
B.1 - Les barrières membranaires.	15
B.2 - Les différentes stratégies de sécrétion possibles.	15
a) L'export à travers la membrane interne.	17
b) La sécrétion via des systèmes adaptés.	17
B.3 - Les protéines à domaines ankyrines : des candidats impliqués dans la communication entre l'hôte et la bactérie.	21
C. Problématique et objectifs.	22
II. Etude du T4SS chez <i>Wolbachia</i>.	25
A. Objectifs.	25
B. Caractérisation et étude transcriptionnelle du T4SS chez wVulC.	26
C. Localisation génomique et conservation du T4SS chez <i>Wolbachia</i>.	27
D. Conclusions et perspectives.	32
E. Publications.	32
E.1 - Félix <i>et al.</i> 2008.	33
E.2 - Pichon <i>et al.</i> 2009.	35
E.3 - Cordaux <i>et al.</i> 2008.	38

III. Les protéines à domaines ankyrines, des effecteurs impliqués dans les interactions entre <i>Wolbachia</i> et ses hôtes ?	40
A. Objectifs.	40
B. Etude comparative des gènes ankyrines de <i>wVulC</i> .	42
C. Distribution et expression des gènes ankyrines <i>pk1</i> et <i>pk2</i> chez les <i>Wolbachia</i> d'isopodes.	51
D. Conclusions et perspectives.	53
E. Publications.	53
E.1 - Ioannidis <i>et al.</i> En préparation.	54
E.2 - Pichon <i>et al.</i> En préparation.	90
IV. Interaction T4SS / effecteurs de <i>Wolbachia</i>.	117
A. Objectifs.	117
B. Interaction T4SS / effecteurs ciblés de <i>wVulC</i> .	117
B.1 - Test par double-hybride chez la levure.	118
B.2 - Test par CRAFT.	121
C. Elaboration d'une banque d'ADNc de <i>wVulC</i> .	125
D. Conclusions et perspectives.	126
V. Discussion générale et perspectives.	127
A. Perspectives concernant les voies de communication de <i>Wolbachia</i> .	128
B. Recherche <i>in silico</i> d'effecteurs.	131
C. Recherche des cibles hôtes.	132
Références	134
 Annexe 1 : le double-hybride.	 154
Annexe 2 : la méthode CRAFT.	159
Annexe 3 : l'extraction et la purification d'ARN.	160
Annexe 4 : la microscopie confocale.	164

LISTE DES ILLUSTRATIONS

Figure 1: Phylogénie des α -Protéobactéries.....	2
Figure 2: Observation d'ovaires d' <i>Armadillidium vulgare</i>	4
Figure 3: L'incompatibilité cytoplasmique.....	7
Figure 4: Le male-killing.....	9
Figure 5: La féminisation.....	9
Figure 6: Détermination du sexe chez les isopodes.....	11
Figure 7: La sécrétion chez les bactéries Gram-négatives.....	14
Figure 8: Les systèmes de sécrétion chez les bactéries Gram-négatives.....	16
Figure 9: Les projets de génomes de <i>Wolbachia</i>	20
Figure 10: La structure du T4SS chez <i>Agrobacterium tumefaciens</i>	24
Figure 11: Phylogénie des Rickettsiales.....	28
Figure 12: Cartographie génomique des gènes vir de <i>Wolbachia</i>	30
Figure 13: Nombre de gènes ankyrines dans les génomes de Rickettsiales.....	41
Figure 14: Comparaison des caractéristiques des gènes ank de wVulC avec celles des gènes ank des 4 souches de <i>Wolbachia</i>	43
Figure 15: Liste des gènes ankyrines homologues entre les 5 souches de <i>Wolbachia</i> étudiées.....	45
Figure 16: Nombre de protéines ankyrines de <i>Wolbachia</i> en fonction du nombre de motifs ANK qu'elles possèdent.....	47
Figure 17: Motif ANK consensus des protéines ankyrines chez wVulC.....	47
Figure 18: Cartographie génomique des prophages de <i>Wolbachia</i>	50
Figure 19: Méthodes utilisées pour identifier des effecteurs bactériens in vivo.....	116
Figure 20: Résultats du double-hybride en levures.....	119
Figure 21: Résultats du CRAfT chez <i>Agrobacterium tumefaciens</i>	122
Figure 22: Protocole des étapes préliminaires à la construction d'une banque d'ADNc.....	124
Figure 23: Résultats de l'enrichissement d'un culot d'ARN en ARN bactériens.....	124
Figure 24: Liste des systèmes de sécrétion de <i>Wolbachia</i>	129

LISTE DES ABREVIATIONS

aa : acide aminé
ABC : ATP-binding cassette
ADN : Acide désoxyribonucléique
ADNc : ADN complémentaire
ANK : ankyrine
ARN : Acide ribonucléique
ARNm : ARN messenger
ATP : Adénosine triphosphate
BEt : Bromure d'Ethidium
BLAST : Basis Local Alignment Search Tool
BSA : Bovine Serum Albumine
CDS : Coding DNA Sequence
CFU : Colony Forming Unit - Unité Formant Colonie
CRAfT : Cre-recombinase Reporter Assay for Translocation
DEPC: Diéthylpyrocarbonate
dNTP : Mélange des 4 désoxyribonucléotides 5'-triphosphate (dATP, dCTP, dGTP, dTTP)
dATP : Désoxyadénosine 5'-triphosphate
dCTP : Désoxycytidine 5'-triphosphate
dGTP : Désoxyguanine 5'-triphosphate
dTTP : Désoxythymidine 5'-triphosphate
DO : Densité Optique
DTT : Dithiothréitol
EDTA : Acide Éthylène Diamine Tétracétique
EF : Elongating Factor
EST : Expressed Sequence Tag
IC : Incompatibilité Cytoplasmique
IPTG : Isopropyl β -D-thiogalactopyranoside
GA : Glande Androgène
HA : Hormone Androgène
kb : Kilobases
kDa : KiloDaltons
LB : Milieu Luria Bertani
L/D : "Light/Dark" (cycle de lumière/obscurité)
 μ L : microlitre (10^{-6} L)
mL : millilitre (10^{-3} L)
mM : millimolaire (10^{-3} moles)
MOPS : 3[N-morpholino]propane sulfonic acid
O/N : Overnight (toute la nuit)
ORF : Open Reading Frame
pb : Paire de bases

PBS : Phosphate Buffer Saline
PCR : Polymerase Chain Reaction
pH : Potentiel Hydrogène
qsp : Quantité suffisante pour
rpm : Rotation par minute
RT : « Reverse Transcription » (transcription inverse de l'ARN)
SSC : Solution Saline Citratée
SDS : Dodécyl Sulfate de Sodium
T1SS-T8SS : Système de sécrétion de type I à VIII
TAE : Tampon « Tris - acide acétique - EDTA »
TEP : Tampon « Tris - EDTA – PMSF »
TCA : Trichloroacetic acid (acide trichloroacétique)
TRIS : tris[hydroxyméthyl]amino-méthane
U : Unités
UV : Ultraviolets
V : Volume
v/v : Volume pour volume
X-Gal : 5-Bromo-4-Chloro-3-Indolyl- β -D-Galactosidase
Y2H : Yeast-two-Hybrid (Double-hybride chez la levure)
YNB : Yeast Nitrogen Base
YPD : Yeast extract - Peptone - Dextrose
32P : Élément Phosphore radioactif
°C : Degré centigrade
% : Pourcentage

AVANT-PROPOS

AVANT-PROPOS : *WOLBACHIA* ET SES APPLICATIONS.

Wolbachia pipientis est le nom donné à une bactérie découverte à l'intérieur même des cellules d'un moustique *Culex pipiens* par Hertig et Wolbach en 1924. Presque 50 ans plus tard, cette bactérie a été tenue pour responsable des échecs de croisements entre différentes lignées de moustiques appartenant à la même espèce (Yen et Barr 1971). Aujourd'hui, *Wolbachia* est considérée comme l'endoparasite le plus répandu sur Terre car elle infecte des vers nématodes filaires et de nombreuses espèces d'arthropodes tels que des acariens, des crustacés, des araignées et surtout des insectes. Selon une estimation récente, environ 60 % des espèces d'arthropodes pourraient être infectées (Bouchon *et al.* 2008, Hilgenboecker *et al.* 2008).

Endosymbiote obligatoire, *Wolbachia* vit exclusivement dans le cytoplasme des cellules de l'hôte. Elle colonise en particulier les cellules du système reproducteur même si l'ensemble des tissus de l'hôte peut être infecté (O'Neill *et al.* 1992, Rousset *et al.* 1992, Stouthamer *et al.* 1993, Werren 1997). Cette bactérie se transmet essentiellement de façon verticale, par voie transovarienne (de mère à fille), les spermatozoïdes mâtures des mâles en sont dépourvus (Werren *et al.* 1995). Aujourd'hui, l'intérêt de la communauté scientifique pour *Wolbachia* ne se limite pas à sa large distribution mais aussi aux nombreux effets physiologiques qu'elle engendre chez ses hôtes. Les phénotypes induits conduisent entre autres à l'incompatibilité de croisements, à la mort ou la féminisation des mâles. Quel que soit le type de modification induit par *Wolbachia* sur la reproduction de ses hôtes, celle-ci a toujours la même conséquence théorique: sa transmission dans les populations hôtes est maximisée.

Wolbachia possède également des caractéristiques qui laissent entrevoir des applications dans de nombreux domaines. Les chercheurs pensent en effet mettre au point différentes stratégies de lutte en se basant sur le contrôle de la symbiose (« Symbiose based control strategies ») (Rasgon 2008). Il est notamment envisagé d'introduire des souches de *Wolbachia* chez des insectes ravageurs de cultures ou vecteurs de pathogènes humains dans le but de remplacer ou réduire les populations et donc leur incidence (Dobson 2003, Bourtzis 2008). Une autre approche a consisté à utiliser une souche de *Wolbachia* qui réduit fortement l'espérance de vie des adultes : en introduisant à grande échelle cette souche dans les populations, il est possible de diminuer la survie des moustiques vecteurs d'agents pathogènes entraînant leur mort avant que le parasite qu'ils hébergent n'ait eu le temps de se développer (Brownstein *et al.* 2003, Turley *et al.* 2009). Un programme de recherche est d'ailleurs financé par la fondation Bill & Melinda Gates (<http://www.mosquitoage.org>) pour lutter contre la dengue (fièvre hémorragique) transmise à l'homme par l'intermédiaire des moustiques *Aedes aegypti* et dans de rares cas par *A. albopictus*.

Un autre moyen de lutte contre les maladies concerne l'utilisation du potentiel invasif de *Wolbachia* et l'introduction, dans le génome de la bactérie, d'un transgène anti-pathogènes c'est-à-dire d'un gène de résistance aux parasites de l'homme (Ruang-Areerate et Kittayapong 2006, Huang *et al.* 2007). Chez les crustacés, l'étude des interactions *Wolbachia*-hôtes pourrait mener à l'identification des gènes bactériens impliqués dans le processus de féminisation. Cet aspect permettrait d'envisager à terme une application dans les élevages de crustacés d'intérêt économique, où les femelles sont sélectionnées pour leur comportement moins agressif et leur croissance plus rapide. D'un autre côté, certaines parasitoses humaines comme l'éléphantiasis (filariose lymphatique) ou l'onchocercose (cécité des rivières), causées par des nématodes filaires, peuvent être contrées en éliminant, par antibiothérapie, les *Wolbachia* nécessaires à la survie des nématodes (Taylor et Hoerauf 1999). Dans ce cas, supprimer *Wolbachia* par un traitement antibiotique revient à tuer le ver hôte et engendre un effet curatif sur la maladie.

Pour libérer des substances dans les cellules hôtes, les bactéries développent des systèmes de sécrétion permettant le transport de macromolécules impliquées dans les fonctions physiologiques de l'organisme infecté. Depuis ces dix dernières années, des systèmes de sécrétion de type IV sont régulièrement mis en évidence chez les différentes souches de *Wolbachia* et pourraient être utilisés, comme cela peut l'être chez de nombreuses bactéries pathogènes, pour sécréter des produits bactériens vers le cytoplasme de l'hôte. Nous pouvons alors espérer découvrir des molécules capables de bloquer la formation ou bien la fonction de ces pores et par conséquent l'expression des bactéries (Baron 2005, Baron et Coombes 2007, Lopez *et al.* 2007). De telles substances pourraient alors être utilisées comme nouvelles cibles thérapeutiques contre les filarioses, palliant ainsi la diminution d'activité des antibiotiques vis-à-vis desquels les bactéries développent des moyens de résistance accrus.

CHAPITRE I :

INTRODUCTION

I. Introduction.

La symbiose a été définie originellement comme l'association étroite de deux organismes pendant une phase ou la totalité de leur vie (de Bary 1879). Elle peut aussi se définir par la localisation de l'organisme symbiotique par rapport à son hôte. On parlera alors d'ectosymbiose si le symbiote vit à l'extérieur des cellules de l'hôte ou d'endosymbiose si le symbiote est localisé à l'intérieur des cellules hôtes. Une attention particulière peut être portée aux microorganismes intracytoplasmiques car leur survie et leur reproduction dépendent directement de la survie et du succès reproductif de leurs hôtes (Bull et Rice 1991).

Cette « vie commune » entraîne en général une co-évolution du parasite et de l'hôte menant parfois les deux partenaires à se livrer à une course aux armements (Van Valen 1973). L'interaction durable entre le parasite et l'hôte dépend du degré de virulence du premier et de la résistance du second (Combes 1995). Si le parasite est extrêmement virulent, l'interaction aboutit généralement à la mort de l'hôte et la transmission du symbiote est assurée de façon horizontale, d'espèce à espèce. Un parasite moins virulent utilise la reproduction de son hôte pour se transmettre verticalement, de parents à descendants. Ce trait induit une pression de sélection au niveau des symbiotes : ceux dont l'hôte possède le meilleur succès reproducteur vont se transmettre le plus (Ewald 1987, Lipsitch *et al.* 1995). En conséquence, en optimisant leur propre transmission, les endosymbiotes optimisent également leur survie et leur valeur adaptative. Les bactéries du genre *Wolbachia* en sont l'exemple typique. Leur potentialité à perturber la reproduction des hôtes semble être extrêmement large et de nouveaux effets induits par ces bactéries sont découverts de façon récurrente. Aussi, ces interactions toutes particulières peuvent s'inscrire selon un spectre allant du mutualisme jusqu'au parasitisme reproductif (Ishikawa 2003). Dans ce cadre, la symbiose à *Wolbachia* présente plusieurs caractéristiques exceptionnelles qui en font un modèle unique et pertinent dans l'analyse des interactions eucaryotes-bactéries.

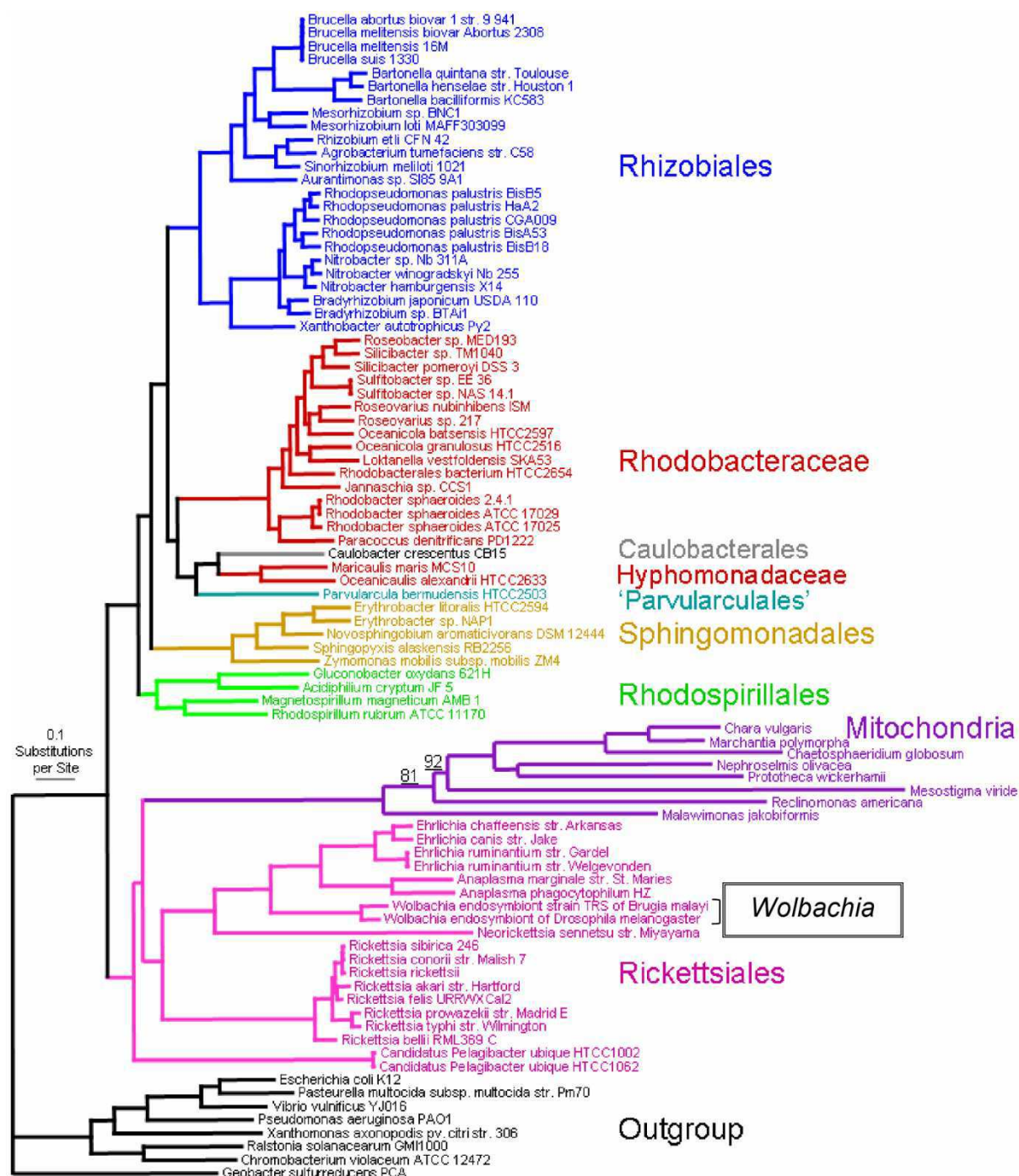


Figure 1: Phylogénie des α -Protéobactéries.

Arbre phylogénétique des alpha-Protéobactéries (80 bactéries) généré par des méthodes bayésiennes sur 104 protéines dont 16 ont servi à établir un sous-arbre de 8 mitochondries secondairement greffé sur le premier. D'après Williams *et al.* (2007).

A. *Wolbachia*, un symbiote manipulateur.

A.1 - Biologie de *Wolbachia*.

Wolbachia pipientis (par la suite nommée *Wolbachia*) est une bactérie intracellulaire obligatoire qui se localise dans le cytoplasme des cellules hôtes, généralement à l'intérieur d'une vacuole et se transmet par voie transovarienne de la mère aux descendants. Ces bactéries peuvent également se transmettre, dans une proportion non négligeable, de façon horizontale et se propager encore plus rapidement dans les populations (Rigaud et Juchault 1995, Poinsoy *et al.* 1998, Veneti *et al.* 2003). Le spectre d'hôtes de *Wolbachia* est, en conséquence, extrêmement étendu ce qui en fait l'endosymbiote le plus répandu sur Terre. Elle est observée chez des Nématodes (principalement chez des filaires et chez l'espèce non-filaire *Angiostrongylus cantonensis*) (Bandi *et al.* 1998, Tsai *et al.* 2007) et dans de nombreux arthropodes tels les acariens (Breeuwer et Jacobs 1996, Tsagkarakou *et al.* 1996), les crustacés (Bouchon *et al.* 2008), les insectes (Werren 1997, Bourtzis et Miller 2003), les araignées et les scorpions (Rowley *et al.* 2004, Goodacre *et al.* 2006, Baldo *et al.* 2007). Ce large spectre reflète sans doute une forte adaptabilité à des hôtes différents (O'Neill *et al.* 1997, Rigaud 1999, Werren *et al.* 2008).

Grâce aux analyses réalisées à partir du gène 16S rDNA, il a été montré que *Wolbachia* est une bactérie à Gram négatif dont les plus proches parents sont des bactéries intracellulaires des genres *Ehrlichia*, *Anaplasma*, *Rickettsia* et *Neorickettsia* (Breeuwer *et al.* 1992, O'Neill *et al.* 1992, Rousset *et al.* 1992). Elle fait partie d'un groupe monophylétique (ordre des Rickettsiales) situé à la base de l'arbre des α -protéobactéries (Williams *et al.* 2007) (Figure 1). D'autres phylogénies réalisées avec des marqueurs plus variables permettent de regrouper les différentes souches de *Wolbachia* en 8 supergroupes nommés de A à H plus ou moins bien établis (Lo *et al.* 2002, Casiraghi *et al.* 2005, Lo *et al.* 2007). La majorité des souches de *Wolbachia* infectant les arthropodes appartient aux supergroupes A et B (Werren *et al.* 1995). Toutes les souches de *Wolbachia* identifiées jusqu'à maintenant chez les crustacés isopodes terrestres appartiennent au supergroupe B (Bouchon *et al.* 1998, Cordaux *et al.* 2004). Les études de distribution des souches de *Wolbachia* d'isopodes ont permis d'identifier plusieurs souches infectant un même hôte et de mettre en évidence, *via* les phylogénies, des transferts horizontaux entre espèces comme chez *Armadillidium vulgare* (Bouchon *et al.* 1998, Cordaux *et al.* 2004, Verne *et al.* 2007).

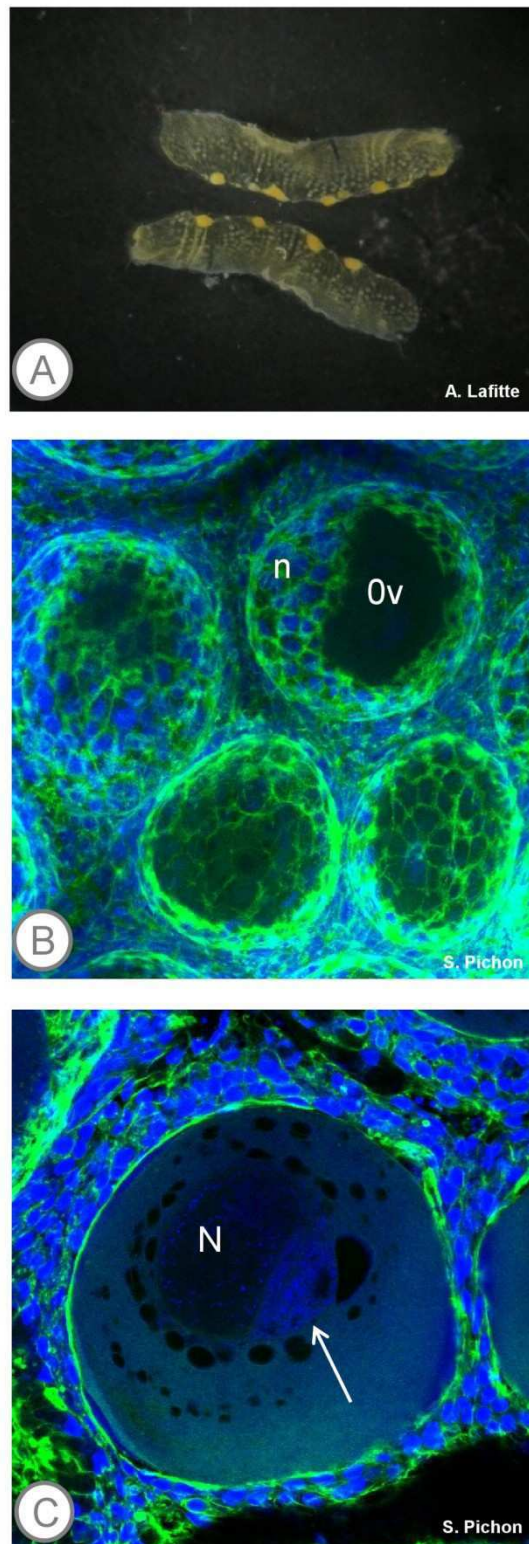


Figure 2: Observation d'ovaires d'*Armadillidium vulgare*.

(A) Photographies d'ovaires d'*A. vulgare*. (B) Observation de la surface d'un ovaire en microscopie confocale. Ov: ovocytes; n: cellules nourricières entourant les ovocytes. (C) Ovocyte observé en microscopie confocale dans lequel on distingue le noyau (N) et à la périphérie de celui-ci un « croissant » (flèche) constitué par l'accumulation de *Wolbachia*. La plupart des cellules nourricières sont également infectées par *Wolbachia*. Les noyaux et les bactéries sont colorés en bleu et les filaments d'actine en vert. Cette expérimentation, dont le protocole est détaillé en Annexe 4 (p. 164), a été réalisée dans le cadre d'une formation d'une semaine proposée par l'Ecole Doctorale ICBG de Poitiers.

A l'instar de nombreux micro-organismes intracellulaires, *Wolbachia* est internalisée dans une vacuole que l'on pourrait qualifier de niche de survie et/ou de réplication (Martin *et al.* 1973, Méresse *et al.* 1999). De nombreuses observations en microscopie électronique ont montré que tous les tissus de l'hôte *A. vulgare* sont infestés (Martin *et al.* 1973, Rigaud *et al.* 1991). C'est néanmoins dans les ovaires que la densité des bactéries est la plus importante. Dans les ovocytes, les *Wolbachia* s'accumulent à la périphérie du noyau et forment un croissant observable en microscopie confocale par coloration au TOPRO-3 (Figure 2). La plupart des cellules nourricières sont également infectées par *Wolbachia*. Outre les ovocytes, *Wolbachia* colonise les cellules de nombreux tissus (cerveau, chaîne nerveuse, tissu adipeux, cœur, caeca digestifs) mais également les cellules responsables de la défense immunitaire (les hémocytes) et les organes hématopoïétiques qui les produisent (Braquart-Varnier *et al.* 2008, Herbinère *et al.* 2008). Chez d'autres espèces, la localisation de *Wolbachia* est restreinte aux ovaires comme chez les nématodes ou l'embryon de drosophile (Bressac et Rousset 1993, Taylor *et al.* 2005).

Wolbachia constitue un véritable paradoxe dans le monde des bactéries symbiotiques qui sont transmises verticalement. En théorie, cette association devrait évoluer vers le mutualisme puisque la survie et la reproduction de *Wolbachia* dépendent aussi de celles de l'hôte (Combes 1995). Or, la valeur sélective des individus hébergeant *Wolbachia* ne semble pas nécessairement augmentée. Au contraire, l'endosymbiose induit différentes altérations de la reproduction de l'hôte qui se traduisent généralement par des baisses de fécondité ou du nombre de descendants (Stouthamer *et al.* 1993, Wade et Chang 1995, Moreau *et al.* 2001). De plus, une évolution progressive vers le mutualisme suppose théoriquement que l'interaction entre les deux partenaires ait des répercussions sur l'évolution du génome des endosymbiotes. Les bactéries intracellulaires obligatoires sont d'ailleurs connues pour posséder des génomes plus réduits que leurs homologues évoluant dans les environnements extracellulaires. Cette réduction génomique est le résultat de phénomènes de recombinaison entre séquences répétées conduisant à des délétions de régions contenant des pseudogènes devenus non fonctionnels (Moran et Plague 2004). Les génomes de différentes souches de *Wolbachia* contiennent une forte prévalence de séquences de phage (virus parasitant les bactéries), de séquences d'insertion (IS) et d'éléments répétés. Or, ces éléments sont connus pour favoriser les événements de recombinaison homologues et représentent autant de facteurs susceptibles d'entraîner et d'affecter la dynamique des réarrangements génomiques.

Néanmoins, du point de vue évolutif, deux caractéristiques majeures expliquent le succès de *Wolbachia* : les transferts dits horizontaux (d'espèce hôte à espèce hôte) et la manipulation de la reproduction des hôtes (Werren 1997).

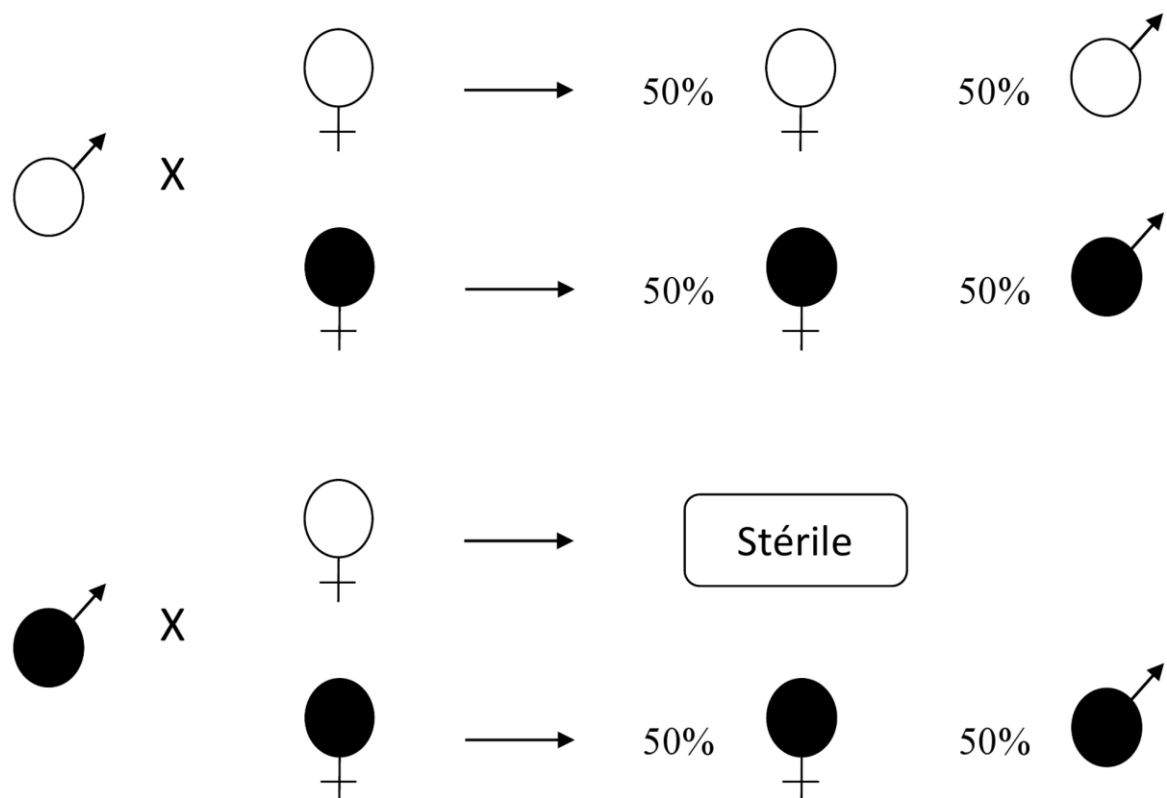
Comme d'autres bactéries symbiotiques d'arthropodes dites « parasites du sexe » tels que *Cardinium* spp., *Arsenophonus* spp., *Flavobacteria* spp. et *Spiroplasma ixodetis*, *Wolbachia* induit divers effets sur la physiologie reproductive des hôtes (Duron *et al.* 2008). Ceux-ci sont comparables à la définition des « phénotypes étendus » (Dawkins 1982) car l'expression des effets de *Wolbachia* se lit à travers le phénotype de l'hôte.

A.2 - *Wolbachia* et les associations obligatoires.

Chez les nématodes filaires, aucune altération de la reproduction induite par *Wolbachia* n'a été observée. Au contraire, sa présence est nécessaire au développement de son hôte (Hoerauf *et al.* 1999, Taylor *et al.* 2005). En effet, les premières études effectuées sur le nématode *Litomosoides sigmodontis* ont montré que l'élimination de *Wolbachia* par des traitements antibiotiques provoquait l'arrêt de croissance et la stérilité de l'hôte (Hoerauf *et al.* 1999). D'autres résultats ont ensuite confirmé les conséquences néfastes de l'élimination de *Wolbachia* sur la physiologie de l'hôte incluant l'inhibition de l'embryogenèse ou encore la mort des adultes (Taylor *et al.* 2005). Aujourd'hui, la communauté scientifique s'accorde à considérer les relations entre *Wolbachia* et nématodes filaires comme mutualistes.

A.3 - *Wolbachia* et les associations facultatives.

A ce jour, les effets de *Wolbachia* les mieux documentés sont ceux conduisant à l'incompatibilité cytoplasmique (IC), à la parthénogenèse, à la mort ou la féminisation des mâles. Les effets de *Wolbachia* ne se résument pas à ces cas détaillés ci-après mais il semble que, selon le type d'association facultative, différents degrés soient observés dans la gamme allant des effets bénéfiques aux effets pathogènes. Dans certains cas la présence de *Wolbachia* peut être corrélée à l'augmentation de la survie et/ou de la fécondité de son hôte (Vavre *et al.* 1999, Dobson *et al.* 2004, Fry *et al.* 2004). La bactérie peut aussi avoir des effets bénéfiques sur des traits qui ne concernent pas la reproduction de l'hôte en participant à sa protection contre des virus ou champignons pathogènes (Panteleev *et al.* 2007, Hedges *et al.* 2008, Teixeira *et al.* 2008). A l'opposé, une souche pathogène de *Wolbachia* (*wMel* popcorn) infectant *Drosophila melanogaster* entraîne en proliférant chez l'adulte la dégénérescence des tissus hôtes et réduit sévèrement son espérance de vie (Min et Benzer 1997, McGraw *et al.* 2002).



Incompatibilité cytoplasmique (unidirectionnelle)

Figure 3: L'incompatibilité cytoplasmique.

Illustration du phénotype d'incompatibilité cytoplasmique unidirectionnelle. Est représentée la proportion d'individus infectés à la descendance suite à différents croisements. On suppose que les symbiotes sont transmis à tous les descendants. Les individus infectés par *Wolbachia* sont en noir, les non infectés en blanc, les mâles sont représentés par ♂, les femelles par ♀. D'après Rigaud et Rousset 1996.

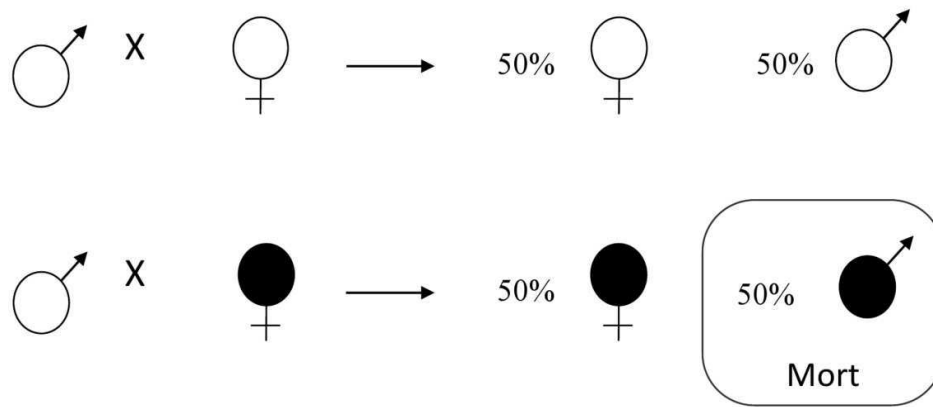
a) L'incompatibilité cytoplasmique.

L'IC se manifeste par une mortalité embryonnaire anormalement élevée dans la descendance issue de croisements entre des mâles infectés par *Wolbachia* et des femelles non-infectées (Moret *et al.* 2001, Bourtzis et Miller 2003) (Figure 3). L'IC peut également s'exprimer lors de croisements entre mâles et femelles porteurs de souches ou variants bactériens différents (O'Neill et Karr 1990) (Figure 3). L'intensité de l'IC induite par *Wolbachia* varie de 50 à 100% et est proportionnelle à la quantité de symbiotes hébergés et à l'âge des mâles (Charlat *et al.* 2001). L'IC constitue à ce jour le phénotype le plus fréquemment décrit qui a été démontré principalement chez des insectes mais également chez trois espèces d'isopodes terrestres : *Porcellio dilatatus petiti* (Legrand *et al.* 1978), *P. d. dilatatus* (P. Grève, communication personnelle) et *Cylisticus convexus* (Moret *et al.* 2001).

Selon certaines hypothèses, l'IC provoque une asynchronie du cycle cellulaire entre les noyaux paternel et maternel lors de la première division de mitose après fécondation (Tram et Sullivan 2002, Tram *et al.* 2003). Lors de croisements incompatibles, la condensation incomplète des chromosomes paternels en prophase entraîne, en anaphase, un retard de migration des chromatides mâles vers les pôles de la cellule par rapport à celles de la femelle. Le résultat se traduit par un arrêt du développement des embryons et un avortement (O'Neill et Karr 1990, Callaini *et al.* 1997, Tram et Sullivan 2002). Peu d'éléments nous permettent de connaître de façon claire les mécanismes moléculaires de l'IC. Chez les mâles, il est probable que la présence de *Wolbachia* modifie l'expression de certains gènes au cours de la spermatogenèse par interaction avec la chromatine de l'hôte. Une étude récente révèle que les effets de la bactérie s'installent bien avant l'entrée en mitose (Landmann *et al.* 2009). Les résultats démontrent qu'une répllication incomplète et qu'un retard d'incorporation des histones lors de la condensation des chromosomes paternels sont à l'origine de la perturbation du cycle cellulaire (Landmann *et al.* 2009).

b) La parthénogenèse.

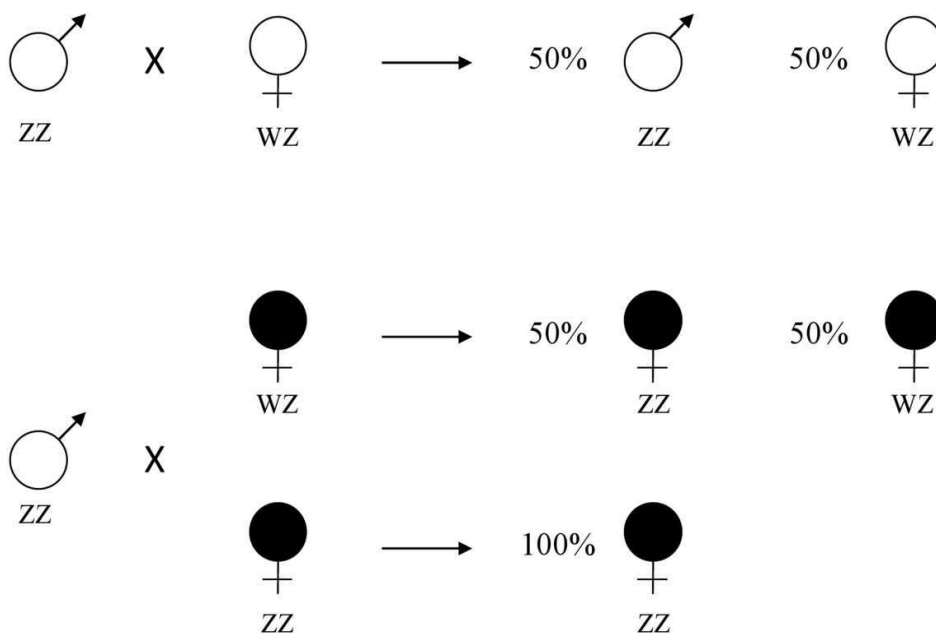
La parthénogénèse thélythoïque se traduit par le développement exclusivement de femelles dans la descendance de certaines espèces d'hyménoptères, collembolés, thrips et acariens (Stouthamer *et al.* 1993, Schilthuizen et Stouthamer 1997, Bourtzis et Miller 2003, Kremer *et al.* 2009a). Habituellement, chez ces espèces, les œufs fécondés (diploïdes) se développent en femelles alors que les œufs non-fécondés (haploïdes) se développent en mâles (Stouthamer *et al.* 1993). *Wolbachia* induit la diploïdisation des œufs non fécondés entraînant leur développement en femelles. Les femelles infectées produisent ainsi jusqu'à deux fois plus de filles (thélytoquie) que les femelles non infectées favorisant ainsi la transmission de *Wolbachia* (Schilthuizen et Stouthamer 1997).



Male killing

Figure 4: Le male-killing.

Illustration du phénotype de « male-killing ». Est représentée la proportion d'individus infectés à la descendance suite à différents croisements. On suppose que les symbiotes sont transmis à tous les descendants. Les individus infectés par *Wolbachia* sont en noir, les non infectés en blanc, les mâles sont représentés par ♂, les femelles par ♀.



Féminisation

Figure 5: La féminisation.

Illustration du phénotype de féminisation. Est représentée la proportion d'individus infectés à la descendance suite à différents croisements. On suppose que les symbiotes sont transmis à tous les descendants. Les génotypes des individus (WZ ou ZZ) sont indiqués. Les individus infectés par *Wolbachia* sont en noir, les non infectés en blanc, les mâles sont représentés par ♂, les femelles par ♀. D'après Rigaud et Rousset 1996.

c) *Le male-killing.*

Le male-killing (MK) correspond à la mort des embryons mâles chez certaines espèces d'insectes infectées par *Wolbachia* (Hurst et Jiggins 2000, Jiggins *et al.* 2000) (Figure 4). Infecter les mâles apparaît alors comme une impasse évolutive pour les bactéries. Une transmission inférieure à 100% empêche néanmoins l'extinction des populations naturelles. Compte tenu du faible nombre de mâles, la forte consanguinité entre individus au sein d'une population favorise le maintien et la transmission de la souche native de *Wolbachia* infectant la population (Werren 1997). Dans certains cas, les œufs mâles non-éclos servent de source d'énergie aux individus femelles après leur éclosion. Cela augmente la valeur sélective des femelles et bénéficie du même coup aux *Wolbachia* dont elles sont porteuses (Hurst et Randerson 2002).

d) *La féminisation.*

Des sex-ratios déséquilibrés en faveur des femelles ont été longtemps observés chez les crustacés isopodes terrestres (Vandel 1941). La féminisation des mâles génétiques par *Wolbachia* a été historiquement démontrée chez plusieurs espèces d'isopodes terrestres (Martin *et al.* 1973, Bouchon *et al.* 2008). Afin d'assurer sa transmission à la descendance, *Wolbachia* colonise principalement les ovocytes où sa présence induit la différenciation des embryons génétiquement mâles en femelles fonctionnelles capables de se reproduire et d'engendrer une descendance (Rigaud *et al.* 1997). Ainsi, le sex-ratio de la descendance est fortement biaisé vers les femelles (Figure 5). Le taux de transmission n'est pas de 100% si bien que quelques ovocytes (~10 à 20%) échappent à l'infection et assurent, par ce biais, la persistance de mâles dans les populations (Bouchon *et al.* 2008).

Aujourd'hui, la prévalence de *Wolbachia* est estimée à 61% des espèces d'isopodes terrestres (Bouchon *et al.* 2008). Le phénomène de féminisation n'a été clairement démontré que chez 3 espèces d'isopodes terrestres: *A. vulgare* et *A. nasatum* (Rigaud *et al.* 2001, Cordaux *et al.* 2004) et *Oniscus asellus* (Rigaud *et al.* 1999). Ce phénotype a également été fortement suspecté chez neuf autres espèces infectées : *Porcellionides pruinosus*, *Porcellio scaber*, *A. album*, *Sphaeroma hookeri*, *S. rugicauda*, *Asellus aquaticus*, *Ligia oceanica*, *Chaetophiloscia elongata* et *Philoscia muscorum* (Bouchon *et al.* 2008). Parmi les 39 espèces d'isopodes infectées par *Wolbachia* répertoriées à ce jour, *A. vulgare* et *P. pruinosus* présentent l'originalité d'être infectés par trois souches bactériennes différentes (Bouchon *et al.* 1998, Michel-Salzat *et al.* 2001, Cordaux *et al.* 2004, Verne *et al.* 2007).

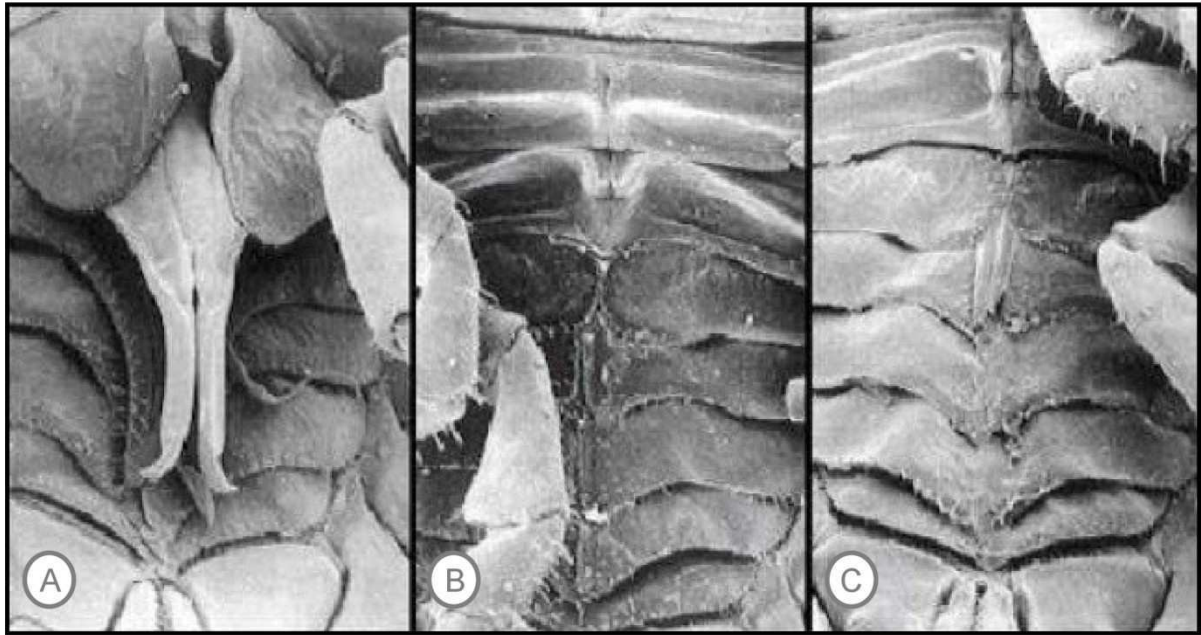


Figure 6: Détermination du sexe chez les isopodes.

Photographies de la face ventrale du pléon de trois phénotypes sexuels d'*Armadillidium vulgare*. De gauche en droite, (A) un mâle (deux stylets copulateurs ou endopodytes), (B) une femelle (pas de stylets copulateurs, mais des ouvertures génitales non visibles à la base des pattes) et (C) un intersexué femelle (femelle fonctionnelle possédant deux stylets copulateurs très courts, asymétriques et non fonctionnels). Il existe également des intersexués stériles dits intersexués mâles car ils possèdent deux endopodytes incomplets.

Chez *A. vulgare*, une première souche *wVulC* est connue comme féminisante (Rigaud *et al.* 2001) et son génome est en cours de séquençage (projet Européen Euwol). Une seconde souche (*wVulM*) a elle aussi un effet féminisant (Cordaux *et al.* 2004) tandis que la troisième (*wVulP*), découverte récemment, est une souche recombinante des deux premières (Verne *et al.* 2007). Ces trois souches peuvent coexister au sein d'une même population de cloportes mais, contrairement à ce qui est observé chez les insectes, aucun cas d'infection multiple d'un même individu n'a été répertorié dans la nature. Ces trois souches sont susceptibles d'avoir des stratégies évolutives, des prévalences et impacts différents dans les populations d'*A. vulgare* (Verne 2007, Lachat 2009).

La féminisation des mâles génétiques par *Wolbachia* a également été démontrée chez un lépidoptère (Hiroki *et al.* 2002, Hiroki *et al.* 2004, Narita *et al.* 2007) et chez une espèce de cicadelle chez laquelle on observe de nombreux intersexués femelles (Negri *et al.* 2006). Néanmoins, le modèle isopode se différencie des autres dans la mesure où la détermination du sexe est presque entièrement contrôlée par la bactérie *Wolbachia* (Bouchon *et al.* 2008). En effet, chez *A. vulgare*, le système de détermination du sexe est une hétérogamétie femelle : les mâles sont homogamétiques ZZ alors que les femelles sont hétérogamétiques WZ (Juchault et Legrand 1972, Legrand *et al.* 1987) (Figure 6). Les mâles possèdent néanmoins l'ensemble du matériel génétique nécessaire pour devenir des femelles puisqu'en présence de *Wolbachia*, ils se développent en femelles fonctionnelles de génotype ZZ. Ils sont capables de transmettre les bactéries à leurs descendants qui se développeront à leur tour en femelles fonctionnelles. Comme ces femelles ZZ produisent beaucoup plus de filles comparées aux femelles génétiques WZ, ces dernières finissent par disparaître complètement de la population (Caubet *et al.* 2000, Bouchon *et al.* 2008). Pour autant, l'action de *Wolbachia* peut s'avérer parfois inefficace ou incomplète. Tel peut être le cas quand, dans les populations naturellement infectées (chez *P. pruinosus* ou *O. asellus* par exemple), sont observés des individus de phénotype intersexué (Figure 6) ou des mâles infectés à l'âge adulte (Marcadé *et al.* 1999, Rigaud *et al.* 1999).

Bien que le mécanisme de la féminisation ne soit pas encore totalement identifié, *Wolbachia* inhibe la différenciation des glandes androgènes, lieu de synthèse de l'hormone androgène (HA) (Martin *et al.* 1999). Cette hormone, de nature protéique (cas exceptionnel dans le règne animal), est responsable du développement des caractères sexuels mâles (Martin *et al.* 1990). Libérée dans la circulation chez l'adulte, cette hormone assure la différenciation des caractères sexuels mâles à chaque mue de l'animal. L'injection d'HA à de très jeunes femelles entraîne leur développement en mâles fonctionnels, démontrant ainsi que les femelles possèdent l'ensemble du matériel génétique nécessaire pour devenir des mâles. La différenciation mâle est donc totalement sous le contrôle de cette hormone.

Dans le cas des intersexués, il est supposé que la charge bactérienne est trop faible durant le développement embryonnaire pour empêcher la différenciation de la glande androgène mais suffisamment élevée à l'âge adulte, pour que les récepteurs à l'AH soient inactivés menant à une féminisation partielle ou à la persistance de mâles infectés (Rigaud et Juchault 1995, Bandi *et al.* 2001).

La féminisation d'embryons mâles infectés par *Wolbachia* pourrait donc résulter de l'inhibition précoce au cours du développement de l'expression et/ou de l'action de l'AH ou de son récepteur, les gonades s'engageant alors par défaut vers la voie femelle. Autrement dit, il est supposé que des effecteurs libérés par *Wolbachia* interagissent avec les mécanismes moléculaires de la différenciation de la glande androgène, que ce soit au niveau du récepteur de l'AH par compétition, par blocage de son expression ou au niveau de la transduction du signal généré par la fixation du ligand sur son récepteur. Pour compléter notre compréhension des processus moléculaires sous-jacents à l'endosymbiose à *Wolbachia* et à ses différents phénotypes étendus, il est donc nécessaire de mieux connaître les voies de sécrétion possibles de *Wolbachia*.

e) *La transition vers le mutualisme.*

Chez les arthropodes, un hyménoptère parasitoïde, *Asobara tabida*, est dépendant de la présence de *wAtab3*, l'une des trois lignées de *Wolbachia* présentes chez cet hôte : *wAtab3* est nécessaire à la réalisation de l'ovogenèse (Dedeine *et al.* 2001, Dedeine *et al.* 2005). Bien que l'on retrouve encore, dans la nature, des populations d'*A. tabida* non infectées, ce cas illustre une transition d'une symbiose facultative vers une symbiose obligatoire. Des résultats récents chez *A. tabida* et des drosophiles révèlent que *Wolbachia* permet également de réduire la concentration de fer circulant protégeant ainsi les cellules hôtes du stress oxydatif et de l'apoptose (Brownlie *et al.* 2009, Kremer *et al.* 2009b). Ces effets étendent alors les possibles conséquences de la présence de *Wolbachia* sur l'évolution des hôtes.

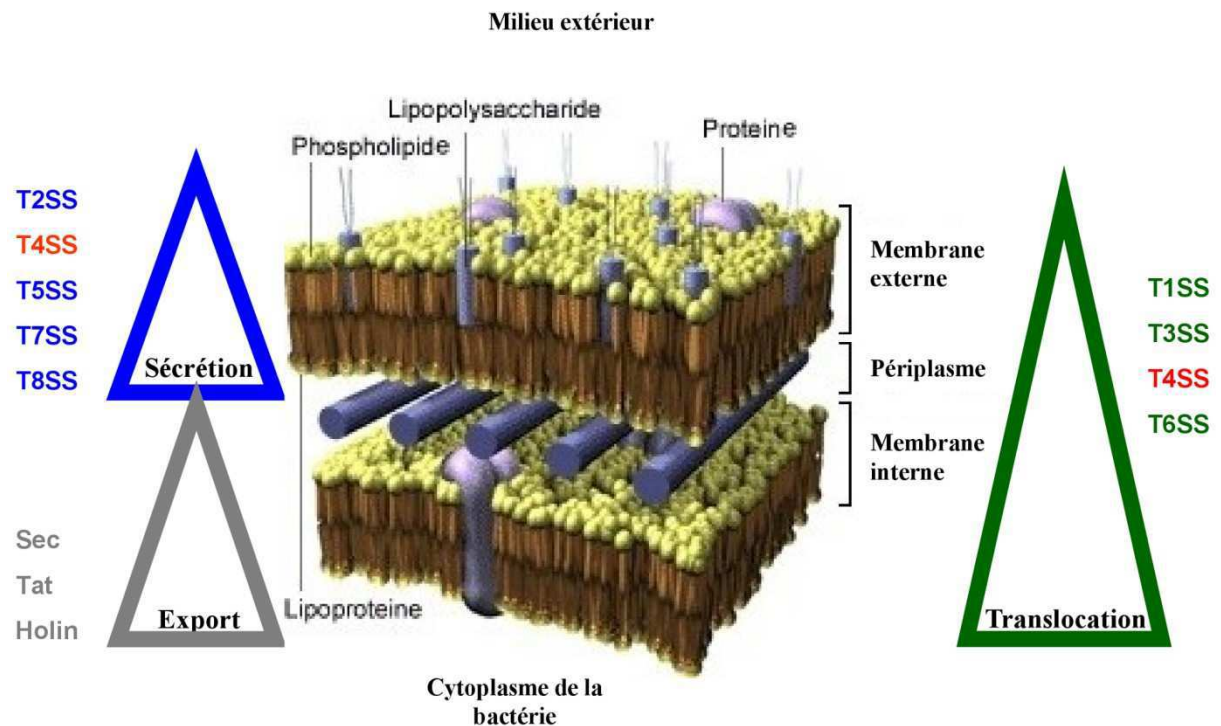


Figure 7: La sécrétion chez les bactéries Gram-négatives.

Représentation schématique des deux voies principales de sécrétion d'effecteurs chez les bactéries Gram-négatives, soit en deux étapes (export puis sécrétion) ou bien en une seule étape (translocation). Crédit pour l'enveloppe bactérienne (http://stl_bjb.ac-dijon.fr/microbio/image/grammoins.jpg).

B. La communication entre *Wolbachia* et ses hôtes.

Au même titre que d'autres bactéries exclusivement intracellulaires, *Wolbachia* doit faire face à certaines contraintes : supprimer ou éviter les défenses de l'hôte et modifier localement l'environnement cellulaire pour satisfaire à ses besoins nutritionnels, sa reproduction et sa transmission. Or l'hôte, qui doit contrôler toute intrusion *via* son système immunitaire, est en fait incapable de se débarrasser de *Wolbachia*. Il est alors supposé que *Wolbachia* n'est pas reconnue et échappe aux défenses de l'hôte. La capacité des *Wolbachia* à échapper au système immunitaire, à se maintenir et à coloniser différents types cellulaires ou différents hôtes suppose que les bactéries sont capables d'interagir avec les molécules impliquées dans différents processus physiologiques de l'hôte.

B.1 - Les barrières membranaires.

Chez les bactéries à Gram négatif, les effecteurs modifiant la biologie de l'hôte doivent franchir au moins deux obstacles que sont les membranes interne et externe de la paroi bactérienne, toutes deux hydrophobes et donc imperméables aux composés hydrophiles. La sécrétion de substances est alors réalisée au moyen de différentes stratégies : la translocation en deux étapes comprenant d'abord l'export à travers la membrane cytoplasmique puis une sécrétion à travers la membrane externe de la bactérie ou bien la translocation d'effecteurs à travers les deux membranes en une seule étape (Figure 7). Les systèmes de sécrétion permettent donc aux bactéries de transférer, dans le cytoplasme de la cellule hôte, des protéines ou complexes ADN-protéines.

B.2 - Les différentes stratégies de sécrétion possibles.

Historiquement, la sécrétion de protéines a été étudiée de manière plus extensive chez des bactéries à Gram négatif pour la plupart pathogènes. A ce jour, six principaux systèmes de sécrétion de protéines ont été mis en évidence et classés du Système de Sécrétion de Type I (T1SS) au Type VI (T6SS) (Henderson *et al.* 2000, Economou *et al.* 2006).

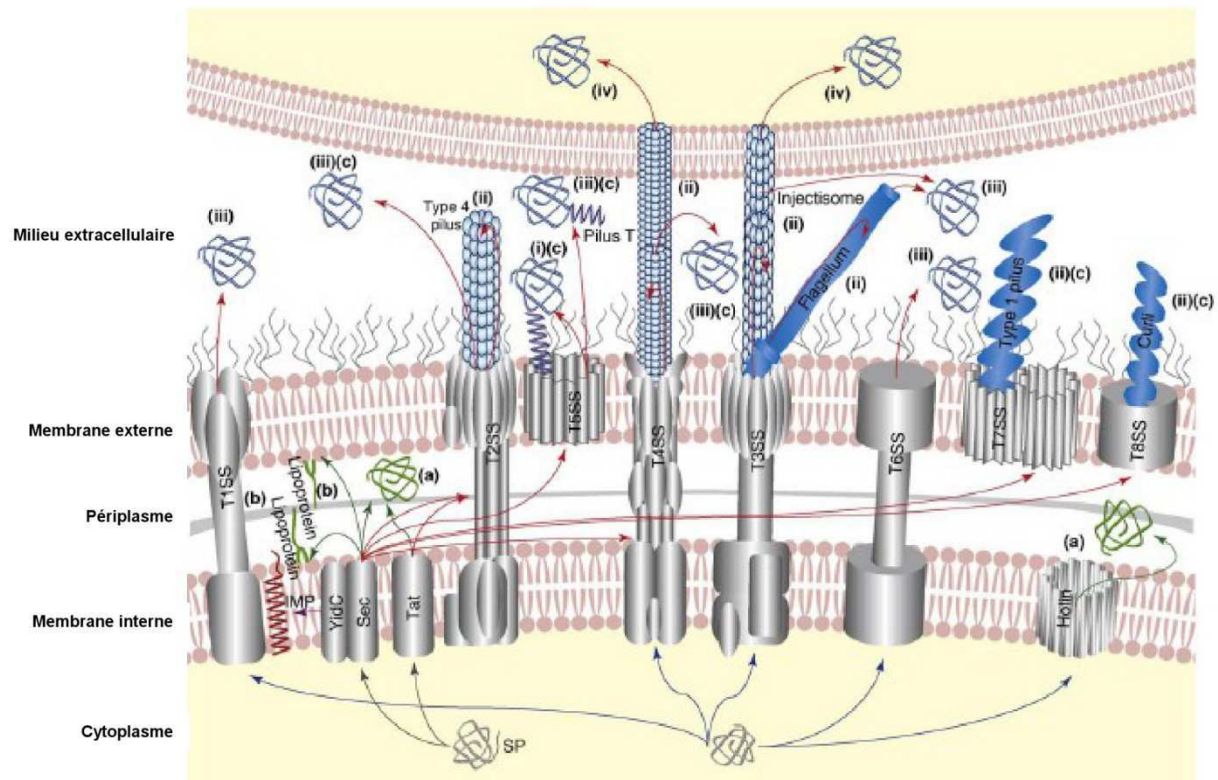


Figure 8: Les systèmes de sécrétion chez les bactéries Gram-négatives.

Représentation schématique des différentes structures de sécrétion d'effecteurs et de leur localisation cellulaire chez les bactéries Gram-négatives (Desvaux *et al.* 2009). Les flèches noires indiquent les différentes voies suivies par les protéines dirigées vers la membrane interne alors que les bleues représentent le devenir des protéines ne possédant pas de peptide signal dans leur région N-terminale. Les effecteurs sont dans certains cas ancrés au niveau de la membrane externe (i) comme tel est le cas des protéines du T5SS; (ii) dans d'autres impliqués dans la structure de pili (T3SS, T2SS, pilus T *via* un T4SS, pilus de type I *via* le T7SS (chaperone–usher pathway) ou curli *via* le T8SS (extracellular nucleation-precipitation pathway)); (iii) relargués dans le milieu extracellulaire ou (iv) directement injecté dans un autre organisme (bactérien ou eucaryote) *via* un T3SS ou un T4SS. Les flèches rouges sont associées aux voies de sécrétion, les violettes liées à l'intégration des protéines membranaires (IMP). Les vertes enfin se limitent aux voies d'export. Les protéines sécrétées sont représentées en bleu et celles exportées en vert.

Un consensus scientifique est attendu pour classer des systèmes auxiliaires dans deux nouvelles catégories (T7SS et T8SS) (Desvaux *et al.* 2009, Tseng *et al.* 2009). Néanmoins, le terme sécrétion doit être réservé au transfert d'effecteurs à travers la membrane externe des bactéries Gram-négatives vers le milieu environnant (Figure 8).

a) L'export à travers la membrane interne.

La sécrétion d'effecteurs peut se décomposer en deux temps, (1) l'export des protéines à travers la membrane interne vers l'espace périplasmique par les systèmes Sec (Secretion), Tat (Twin-arginine translocation) ou holin (hole forming) et (2) le transport à travers la membrane externe par des systèmes de sécrétion (Figure 8). Dans un premier temps, les protéines sont exportées du cytoplasme vers le périplasme de la bactérie par des mécanismes impliquant la reconnaissance, par l'intermédiaire de chaperones, d'une séquence signal spécifique des préprotéines et le transport actif de ces molécules vers et à travers la membrane interne, grâce à des translocases (Economou 1999).

b) La sécrétion via des systèmes adaptés.

Une fois dans l'espace périplasmique, les protéines exportées peuvent y rester, s'ancrer dans la membrane externe ou bien, dans un second temps, être sécrétées par des systèmes de sécrétion adaptés de type II, IV, V, VII et VIII (Hueck 1998). Ces derniers représentent autant de mécanismes permettant la sécrétion des déterminants bactériens à travers la membrane externe (Kostakioti *et al.* 2005) (Figure 8). Le Système de Sécrétion de Type II (T2SS) est responsable du transport extracellulaire d'une variété d'enzymes et de toxines (Sandkvist 2001a, 2001b). L'exemple le mieux étudié reste celui des 14 gènes contigus composant le T2SS de *Klebsiella oxytoca* (Enterobacteriaceae) responsable de la sécrétion de la pullulanase PulA (Hueck 1998). Le T5SS comprend lui, le système de sécrétion à deux partenaires (TPS) et les autotransporteurs (AT) qui assurent leur propre transport à travers la membrane externe des bactéries (Kostakioti *et al.* 2005). Il contient trois domaines : une séquence signal amino-terminale pour son export *via* le système Sec, un domaine interne dit α -domaine passager, un β -domaine carboxy-terminal (Kostakioti *et al.* 2005). Ce dernier s'incorpore dans la membrane externe et forme un pore pour l' α -domaine qui reste soit à la surface de la bactérie ou est relâchée dans l'environnement externe (Kostakioti *et al.* 2005). Les exemples sont nombreux mais l'on peut citer la sécrétion de VacA chez *Helicobacter pylori* (Schmitt et Haas 1994). Longtemps considéré comme système de type V (Thanassi *et al.* 2005), le système de sécrétion chaperon/guide (CU, « Chaperon-Usher ») a été récemment proposé comme appartenant à un nouveau Type VII (T7SS) (Desvaux *et al.* 2009, Tseng *et al.* 2009).

Ce système requière la participation de deux composants, une protéine chaperone dans le périplasme et un guide ancré dans la membrane externe (Kostakioti *et al.* 2005). L'exemple le mieux connu est celui permettant l'assemblage des pili de type P et ceux de type I chez la souche *Escherichia coli* enteropathogène (Thanassi *et al.* 2005, Remaut *et al.* 2008). L'appartenance à un T8SS est suggérée pour la sécrétion « extracellular nucleation-precipitation pathway » (ENP) offrant la possibilité aux bactéries pathogènes comme *E. coli* d'interagir spécifiquement avec son hôte et de former un biofilm (Stathopoulos *et al.* 2000, Kostakioti *et al.* 2005, Desvaux *et al.* 2009).

Les voies de sécrétion permettent également la sécrétion directe (en une étape) d'agents bactériens du cytoplasme vers l'environnement extérieur. Celles-ci comprennent les Systèmes de Sécrétion de Type I, III, IV et VI (T1SS, T3SS, T4SS et T6SS) (Figure 8). Le T1SS comprend des protéines transporteur ABC (« ATP-binding cassette »), des protéines de fusion membranaire (MFP) et des protéines s'ancrant dans la membrane externe (OMP). Ce système est utilisé par de nombreuses bactéries Gram-négatives pour la sécrétion de toxines, protéases ou lipidases (Binet *et al.* 1997) telles l'hémolysine α (HlyA) chez *E. coli* (Gentshev *et al.* 2002). Le type III est certainement le mécanisme de sécrétion le plus étudié (Cornelis 2006, Economou *et al.* 2006). Il met en jeu une vingtaine de protéines, dont beaucoup sont des homologues des composants de la biosynthèse des flagelles (Medini *et al.* 2006). Après l'adhérence de la bactérie à la surface cellulaire hôte, ce système permet la sécrétion de protéines capables de détourner les défenses immunitaires, d'induire l'apoptose, ou encore de réorganiser l'actine du cytosquelette chez l'hôte (Coburn *et al.* 2007). Le T6SS est également une structure tubulaire nécessitant l'action d'une ATPase hexamérique (ClpV) pour le transport à travers l'enveloppe bactérienne d'*Agrobacterium tumefaciens* ou de *Vibrio cholerae* (Filloux *et al.* 2008, Wu *et al.* 2008).

La sécrétion des molécules peut enfin s'effectuer par un Système de Sécrétion de Type IV (T4SS) en une seule étape ou bien après l'export préalable par le système Sec (Figure 8). De plus, le T4SS possède l'originalité de pouvoir réaliser le transport de protéines seules ou de complexes ADN-protéines (Christie 1997). Un grand nombre de molécules peuvent donc transiter *via* ce système vers le milieu extérieur ou vers la surface de la bactérie (Aly *et al.* 2008). La superfamille des T4SS a été subdivisée en deux classes (Christie et Vogel 2000). Le premier groupe (T4aSS) réunit les types dont les gènes sont homologues des gènes *virB* et *virD* d'*A. tumefaciens* et des gènes *lvh* (*Legionella* Vir Homolog) de *Legionella pneumophila* (Segal *et al.* 1999). Le second groupe (T4bSS) contient les gènes proches des gènes *dot/icm* (pour Defective in Organelle Trafficking / Intra-Cellular Multiplication) de *L. pneumophila* et les gènes du plasmide ColIB-P9 (Incl) de *Shigella flexnerii* (Segal *et al.* 1999, Christie 2001, Sexton et Vogel 2002, Cascales et Christie 2003).

Les deux catégories partagent néanmoins une origine commune avec la machinerie de conjugaison bactérienne des bactéries à Gram négatif et positif (Christie 2001, Medini *et al.* 2006). Initialement, différents T4SS ont été caractérisés d’abord chez de nombreux pathogènes puis découverts chez de simples bactéries parasites conférant alors au T4SS un rôle plus large que celui d’être impliqué dans les effets pathogènes. Bien que la plupart de ces T4SS injectent leurs substrats directement dans le cytosol de la cellule eucaryote comme cela a été démontré pour les T3SS (Blöcker *et al.* 2008), les fonctions de la famille du T4SS sont regroupées en trois catégories (Backert et Meyer 2006):

- La conjugaison assurant le transfert d’ADN vers une cellule réceptrice cible. Le système *lov* de *L. pneumophila* fonctionne ainsi.
- La « compétence ou libération d’ADN » permettant l’échange d’ADN avec le milieu extracellulaire. Il a été montré que le système Dot/Icm de *L. pneumophila* remplit cette fonction et est nécessaire pour la virulence des bactéries.
- Le transport d’effecteurs qui consiste en la sécrétion de molécules effectrices bactériennes vers les cellules eucaryotes durant l’infection des plantes ou d’animaux. Par exemple, le système *ptl* permet le transport de la toxine pertussique de *Bordetella pertussis* vers l’extérieur de la bactérie.

En 2000, Masui et ses collègues furent les premiers à mettre en évidence l’existence de gènes codant un T4aSS homologue à celui d’*A. tumefaciens* chez les souches de *Wolbachia* *wTai* (cricket de Taiwan) et *wKueYo* (pyrale de la farine de Méditerranée) (Masui *et al.* 2000). Par la suite, le séquençage des génomes de différentes souches de *Wolbachia* (Figure 9) : *wMel* infectant le diptère *D. melanogaster* (Wu *et al.* 2004), *wBm* hébergée par le nématode *B. malayi* (Foster *et al.* 2005), *wPip-Pel* chez le moustique *Culex pipiens* (Klasson *et al.* 2008) et *wRi* souche de *D. simulans* (Klasson *et al.* 2009) ainsi que le séquençage partiel de trois autres souches de drosophiles (Salzberg *et al.* 2005), ont permis d’identifier des gènes homologues codant un T4SS. Dès 2004, les premières comparaisons bioinformatiques n’ayant pas permis de mettre en évidence de gènes codant d’autres systèmes de sécrétion chez *Wolbachia*, l’attention s’est tournée naturellement vers ceux codant un T4aSS (Wu *et al.* 2004, Foster *et al.* 2005, Salzberg *et al.* 2005). Aujourd’hui, des études ont montré la présence et l’expression des gènes codant un T4SS chez d’autres souches de *Wolbachia* (Félix 2004, Rancès *et al.* 2008) renforçant la nécessité d’approfondir la relation entre T4SS et «virulence» de *wVulC*.

Type de relation hôte/bactérie:

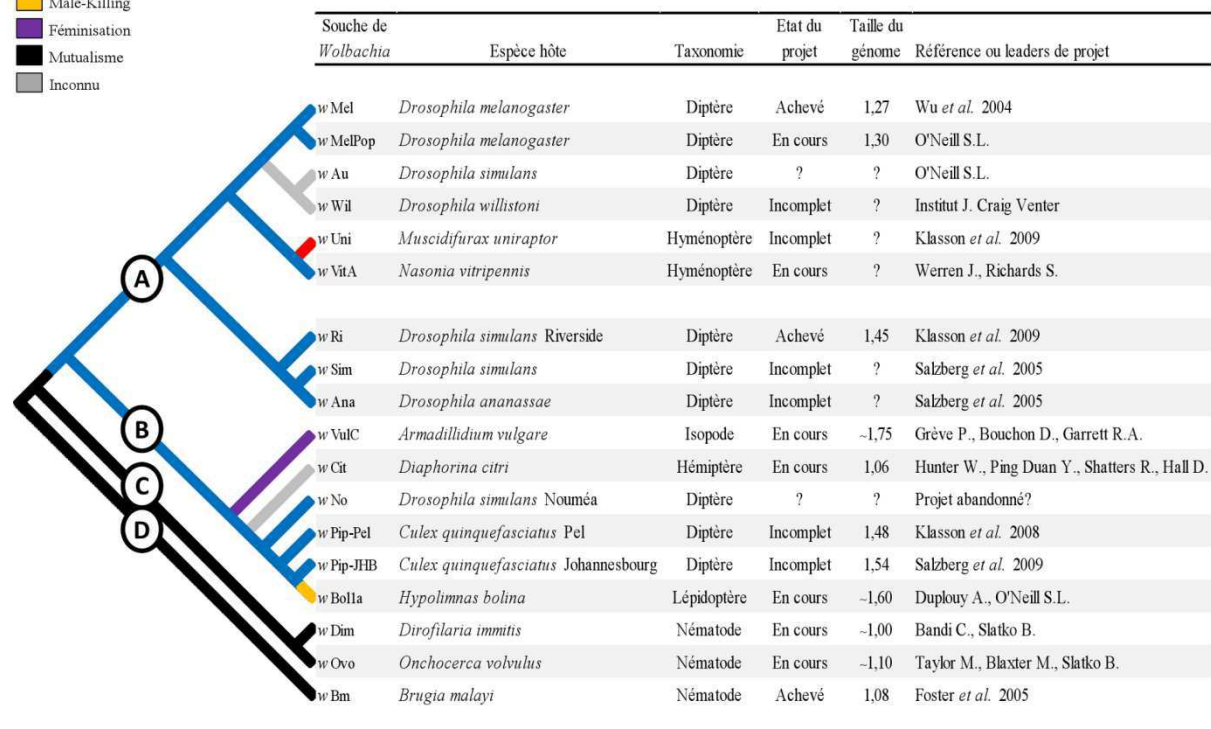
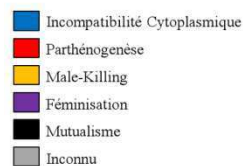


Figure 9: Les projets de génomes de *Wolbachia*.

Liste des génomes de *Wolbachia* actuellement publiés ou en cours de séquençage. Seuls trois génomes sont à ce jour complets (wRi, wMel et wBm) alors que d'autres publiés ne sont pas encore totalement assemblés (wPip-Pel, wAna, wSim, wUni). D'autres génomes sont en cours de séquençage ou d'assemblage. L'arbre phylognénétique a été construit à partir de l'alignement du gène *wsp* (~500 bp) et par la méthode de maximum de parcimonie dans Mega4. Les quatre supergroupes de *Wolbachia* (A–D) impliqués sont représentés sur l'arbre. Les couleurs de branches correspondent au type de relation entretenue entre *Wolbachia* et son hôte.

Les génomes de plusieurs souches de *Wolbachia* sont actuellement séquencés dans le cadre d'un projet européen Euwol coordonné par K. Bourtzis (Université de Ioannina, Grèce) (Figure 9) (Oehler et Bourtzis 2000). Celui de *wVulC*, seul génome de souche féminisante, est lui en cours d'assemblage et d'annotation par des collègues partenaires du projet Euwol (R.A. Garrett, Danish Archea Centre, Université de Copenhague, Danemark). Nous avons donc à ce jour la possibilité d'explorer l'information contenue dans 51 contigs. Ceux-ci représentent un total d'environ 1,6 Mb de séquences ADN alors que la taille du génome avait été estimée à 1,75 Mb après électrophorèse en champ pulsé (Félix 2004). Les premières analyses ont permis de confirmer la présence des opérons *vir* du T4SS chez cette souche. Parallèlement, l'analyse des génomes a aussi conduit à identifier des effecteurs bactériens potentiels dont les protéines à domaines ankyrines qui pourraient être sécrétés par le T4SS et être impliqués dans le dialogue moléculaire entre *Wolbachia* et son hôte (Fenn et Blaxter 2006).

B.3 - Les protéines à domaines ankyrines : des candidats impliqués dans la communication entre l'hôte et la bactérie.

Dans la bibliographie, une grande variété de déterminants bactériens, transitant notamment *via* un T4SS, a été mise en évidence. L'identification de gènes *ank* codant des protéines à motifs ankyrines (ANK) suscite l'intérêt de la communauté à *Wolbachia* pour différentes raisons. Les motifs ANK longs de 33 acides aminés respectent d'abord une structure particulière dite "helix-loop-helix", succession de feuillets β et de boucles déterminant des angles de 90°. De tels motifs sont reconnus pour interagir entre eux et être à la base d'interactions protéine-protéine chez les eucaryotes (Bork 1993, Sedgwick et Smerdon 1999, Mosavi *et al.* 2004). Derrière cette quasi-uniformité de structure se cachent une incroyable diversité de fonctions dont la régulation du développement, du cycle cellulaire, de la transcription ou encore l'interaction avec le cytosquelette et la membrane des cellules (Bork 1993, Sedgwick et Smerdon 1999, Mosavi *et al.* 2004).

De manière générale, virus et bactéries sont dépourvus de gènes *ank* et leur origine est supposée eucaryote. Or, les génomes de *Wolbachia* contiennent une quantité exceptionnellement importante de gènes *ank*. Il est à noter que le nombre de ces gènes est plus faible chez *wBm*, *Wolbachia* de filaire, car on ne retrouve chez cette souche ni prophage ni gènes *ank* associés à ces régions (Foster *et al.* 2005). Par ailleurs, il est supposé que de telles protéines peuvent être sécrétées par le T4SS de *Wolbachia* et interagir avec les molécules de l'hôte *via* leurs motifs ANK.

Les démonstrations récentes du transfert de la protéine AnkA par un T4aSS chez *Anaplasma phagocytophilum*, proche parent de *Wolbachia*, renforcent également l'intérêt pour cette famille de gènes (Caturegli *et al.* 2000, Park *et al.* 2004, Lin *et al.* 2007). D'autres cas viennent étayer notre hypothèse : les protéines ANK de *L. pneumophila* and *Coxiella burnetii* transitent par un T4bSS pour perturber le fonctionnement de cellules de protozoaires ou d'humains (Seshadri *et al.* 2003, Al-Khodori *et al.* 2008, Habyarimana *et al.* 2008, Pan *et al.* 2008, Voth *et al.* 2009). Enfin, des hypothèses formulées par Sinkins et collègues (2005) proposent une implication de ces gènes *ank*, notamment le gène *pk2*, dans le phénotype d'IC induit par *Wolbachia* chez *Culex pipiens*. Les analyses génomique et transcriptionnelle chez différentes souches ont montré notamment que ce même gène, appartenant à un prophage de *Wolbachia*, possède un codon stop prématuré chez une souche *wAu* ne réalisant pas d'IC en comparaison avec *wMel* dont le phénotype est l'IC (Iturbe-Ormaetxe *et al.* 2005). Le gène homologue chez plusieurs variants *wPip* capables d'induire différents degrés d'IC chez *C. pipiens* possède également des profils d'expression différentielle selon le sexe et les populations d'origine (Sinkins *et al.* 2005). Ces premiers résultats montrent que les ankyrines pourraient être exportées vers le cytoplasme de l'hôte et être impliqués dans l'expression des phénotypes de *Wolbachia*.

C.Problématique et objectifs.

Wolbachia n'est, pour l'instant, ni cultivable ni génétiquement manipulable et l'utilisation de mutants pour la caractérisation d'effecteurs bactériens n'est donc pas envisageable. L'analyse comparative des génomes de *Wolbachia* constitue donc un point de départ à l'identification de déterminants impliqués dans les dialogues hôte/symbiote. Chez *Wolbachia*, trois classes de gènes présentent un intérêt tout particulier car les protéines qu'ils codent pourraient être impliquées dans l'expression de certains phénotypes (McGraw et O'Neill 2004, Fenn et Blaxter 2006) :

- les gènes du bactériophage intégrés au génome bactérien représentent une source d'innovation dans leur capacité à introduire ou transférer des gènes d'intérêt entre souches différentes.
- certains gènes codent un T4SS pouvant sécréter, à travers la paroi bactérienne, des macromolécules dans le cytoplasme de la cellule hôte.
- les gènes *ank* codent des protéines à motifs ankyrines (ANK) pouvant potentiellement être sécrétées *via* un T4SS, interagir avec des protéines de l'hôte et être impliquées dans l'altération de la reproduction de l'hôte.

Au cours de ma thèse, j'ai pu mettre à profit les premières données de séquençage de la souche *wVulC* et bénéficier des données préliminaires sur le T4SS constituées par la thèse de Christine Félix (2004). Ainsi, mes objectifs de thèse afin d'explicitier les dialogues moléculaires entre *Wolbachia* et son hôte *A. vulgare* se sont articulés en trois volets:

- compléter la caractérisation du Système de Sécrétion de Type IV chez *wVulC* et le comparer au T4SS d'autres modèles.
- identifier et caractériser les protéines à motifs Ankyrine de *wVulC*.
- étudier les interactions potentielles entre les protéines ankyrines et celles du T4SS chez *wVulC*.

CHAPITRE II :

ETUDE DU T4SS CHEZ

WOLBACHIA.

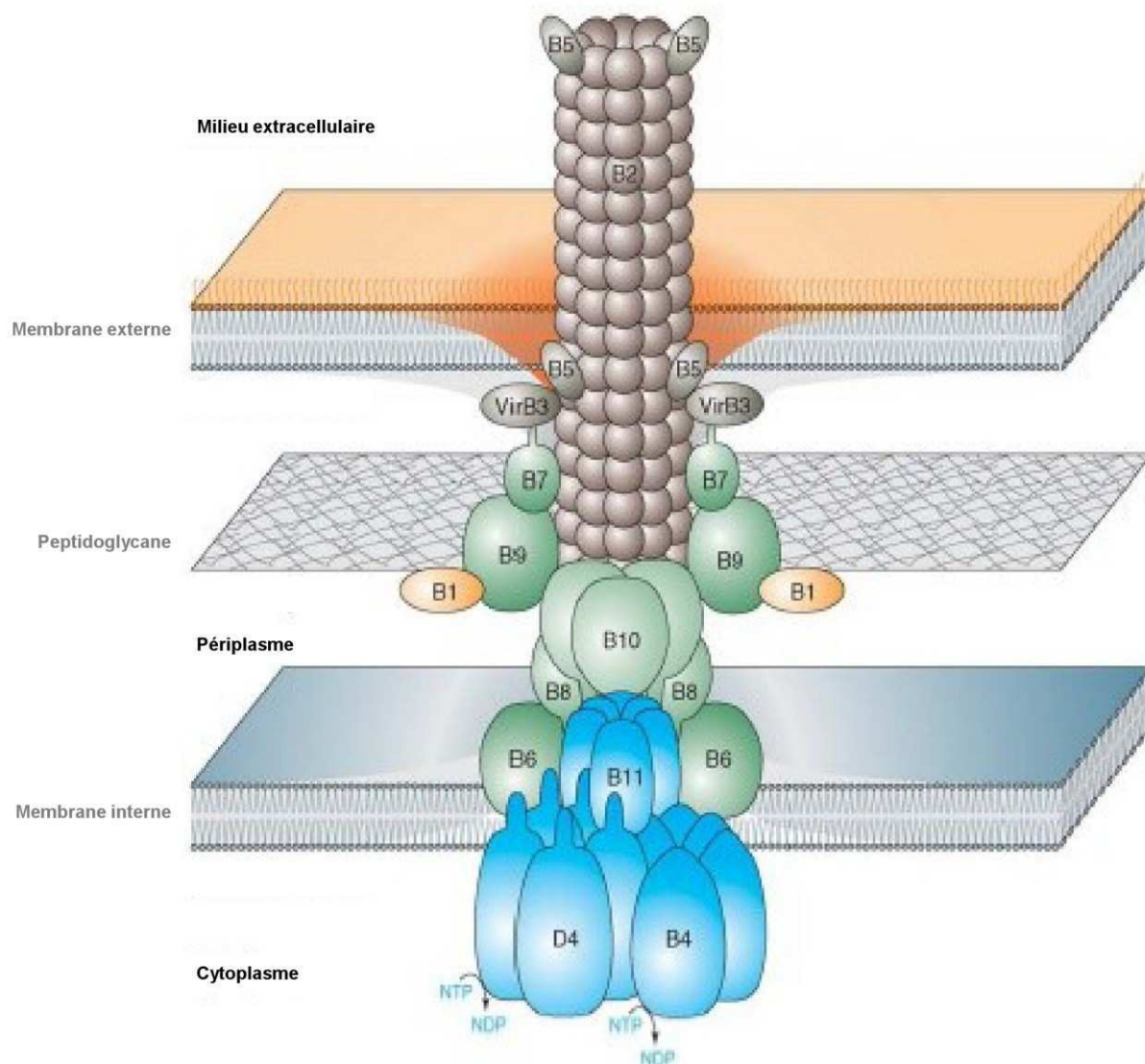


Figure 10: La structure du T4SS chez *Agrobacterium tumefaciens*.

Représentation schématique de l'assemblage supposé des constituants du T4SS entre eux au niveau de l'enveloppe d'*Agrobacterium tumefaciens*. D'après Backert et Meyer 2006.

II. Etude du T4SS chez *Wolbachia*.

A. Objectifs.

De nombreuses bactéries Gram-négatives utilisent un T4SS pour sécréter à travers l'enveloppe bactérienne des facteurs de virulence dans le cytoplasme de la cellule hôte. Les composants de cette machinerie de transport forment un complexe multiprotéique codé par l'opéron *virB/D* comportant jusqu'à 12 gènes (*virB1-11* et *virD4*). Les analyses structurales et biochimiques de ce système de sécrétion ont permis notamment chez l'archétype *A. tumefaciens* d'identifier trois groupes fonctionnels de protéines Vir (Beijersbergen *et al.* 1994, Backert et Meyer 2006, Fronzes *et al.* 2009) (Figure 10) :

- VirB4, VirB11 et VirD4 : trois ATPases localisées sur la face cytoplasmique de la membrane interne, assurant l'apport énergétique nécessaire au transport de substrats à travers le pore, ainsi qu'à l'assemblage du complexe de sécrétion.
- VirB1, VirB3 et VirB6 à B10 ou protéines « core », formant le canal transmembranaire qui traverse les membranes interne et externe, permettant le transfert des effecteurs sécrétés vers le milieu extérieur.
- VirB2 et VirB5 constituant un pilus sur la surface bactérienne qui permet le contact entre la bactérie et la cellule hôte.

Dans le groupe des Rickettsiales, les gènes du T4SS sont organisés en deux clusters de gènes distincts (au total 8 gènes) tous cotranscrits (Brayton *et al.* 2005, Hotopp *et al.* 2006, Gillespie *et al.* 2009). Le premier opéron, *virB3-B6* (gènes *virB3*, *virB4* et *virB6*) est distinct du second nommé *virB8-D4* (gènes *virB8*, *virB9*, *virB10*, *virB11* et *virD4*). *Wolbachia* pourrait donc utiliser un tel système pour sécréter des macromolécules ciblant les protéines de l'hôte et pouvant être impliquées dans la survie, la nutrition, la réplication, ou l'expression du phénotype étendu de la bactérie.

La séquence de ces deux opérons, chez *wVulC*, a été en quasi-totalité obtenue par Christine Félix au cours de sa thèse (Félix 2004). Elle a également montré que ces gènes s'exprimaient dans les ovaires d'*A. vulgare* (Félix 2004). J'ai poursuivi cette caractérisation en deux étapes successives et complémentaires.

La première a permis d'obtenir la totalité de la séquence des deux opérons, notamment des extrémités 5' et 3'. Elle a surtout consisté en l'étude de l'expression de ces gènes dans la totalité des tissus d'*A. vulgare*. La deuxième étape a concerné l'histoire évolutive du T4SS des *Wolbachia* de crustacés Isopodes terrestres et d'autres arthropodes. En comparant les séquences et l'organisation de gènes chez différentes souches de *Wolbachia*, il a été possible de connaître le degré de conservation de ce système et de juger de son importance dans la biologie des bactéries.

B. Caractérisation et étude transcriptionnelle du T4SS chez *wVulC*.

Les résultats obtenus ont fait l'objet d'une publication portant sur la première caractérisation et étude transcriptionnelle d'un T4SS d'une souche féminisante de *Wolbachia* [cf. Article 1 : (Félix *et al.* 2008), p. 33]. Ce travail a été en partie réalisé dans le cadre de la thèse de Christine Félix qui a mis en évidence 8 gènes orthologues codant un T4SS chez *wVulC* et chez trois autres souches infectant des isopodes terrestres (Fig. 1a, Article 1). J'ai ensuite complété les séquences 5' et 3' manquantes de certains gènes ce qui a permis de comparer l'ensemble des séquences du T4SS de *wVulC* aux autres séquences disponibles dans GenBank. L'organisation génétique (Figure 2, Article 1) ainsi que la séquence des gènes ont permis de suggérer une forte conservation du T4SS entre différents génomes de *Wolbachia*. De plus, cet ensemble de gènes représente plus que le minimum nécessaire au fonctionnement d'un T4aSS chez les bactéries Gram-négatives (den Hartigh *et al.* 2008). Il a été également montré par Southern blot que ces gènes *vir* s'organisent en deux clusters localisés en deux loci distincts sur le chromosome bactérien (Fig. 1b, Article 1). Ce découpage en deux clusters est aussi fidèle à l'organisation génomique observée chez de nombreuses Rickettsiales (Brayton *et al.* 2005, Hotopp *et al.* 2006, Gillespie *et al.* 2009). La localisation exacte des deux opérons sur le génome de *wVulC* n'est pas encore connue. Néanmoins, chez les autres souches dont le génome complet est connu, ils sont séparés d'au moins 379 kilobases (Wu *et al.* 2004, Foster *et al.* 2005, Klasson *et al.* 2008, Klasson *et al.* 2009). De plus, l'étude de l'expression des gènes a permis de montrer que chaque cluster de gènes est co-transcrit en opéron (Fig. 3a, Article 1).

Les analyses par RT-PCR montrent que les gènes *vir* s'expriment dans le tissu d'*A. vulgare* le plus riche en *Wolbachia*, les ovaires en fin de vitellogenèse (juste avant la ponte) (Fig. 3, Article 1). Une particularité du modèle isopode est que des *Wolbachia* sont présentes dans de nombreux tissus que l'on peut individualiser par dissection.

Ainsi, j'ai pu montrer l'expression de ces gènes dans l'ensemble des tissus hôtes que colonise la bactérie (Fig. 3b, Article 1). Une PCR emboîtée a été nécessaire pour démontrer cette expression dans le cerveau et les chaînes nerveuses (Fig. 3b, Article 1). Ce dernier résultat peut s'expliquer par une trop faible charge bactérienne dans ces tissus, rendant la détection difficile, ou bien à cause d'une modulation de l'expression des opérons du T4SS.

Une étude par RT-PCR quantitative ou par hybridation *in situ* devra être envisagée pour déterminer s'il existe un lien direct entre le degré de transcription des gènes et la charge bactérienne par cellule. Un autre enjeu sera d'appréhender les mécanismes qui contrôlent ou déclenchent l'expression des gènes vir. Chez *A. tumefaciens*, leur transcription est activée au moyen d'un système régulateur à deux composants VirA/VirG, en réaction à des molécules signal de la plante (Winans *et al.* 1994). Chez d'autres protéobactéries, la présence de régulateurs spécifiques de la famille LuxR/LuxI ajoutée au Quorum Sensing déterminent le contrôle de l'expression des gènes du T4SS (Taminiau *et al.* 2002, Delrue *et al.* 2005, de Jong *et al.* 2008). Chez les Rickettsiales, la régulation de la transcription du T4SS a été très peu étudiée à ce jour. Un facteur de régulation a toutefois été mis en évidence chez *Ehrlichia chaffensis* (Cheng *et al.* 2006, Wang *et al.* 2007). Ce facteur EcxR régule l'expression des deux opérons vir, mais aussi et de manière synchrone les paralogues des gènes vir. Les souches de *Wolbachia* dont le génome est connu possèdent un homologue de ce gène ce qui offre l'opportunité d'étudier plus spécifiquement la régulation de l'expression du T4SS chez *Wolbachia*.

C. Localisation génomique et conservation du T4SS chez *Wolbachia*.

La détermination complète de la séquence du T4SS de *wVulC* nous a permis dans une seconde publication de prospecter de manière extensive l'histoire évolutive du T4SS chez *Wolbachia* [cf. Article 2 : (Pichon *et al.* 2009), p. 35]. L'alignement de chaque gène composant le T4SS a été réalisé en intégrant les séquences disponibles dans GenBank. Nous avons ainsi déterminé des couples d'amorces consensuelles pour amplifier par PCR l'intégralité des opérons et déterminer l'organisation des gènes du T4SS chez 27 souches de *Wolbachia*. Cette étude a été réalisée chez 18 espèces d'isopodes terrestres appartenant à sept des huit familles connues pour héberger *Wolbachia* (Tableau S1, p. 36). L'analyse a été également étendue à 9 autres souches de *Wolbachia* infectant des insectes et des araignées (Tableau S1, p. 36).

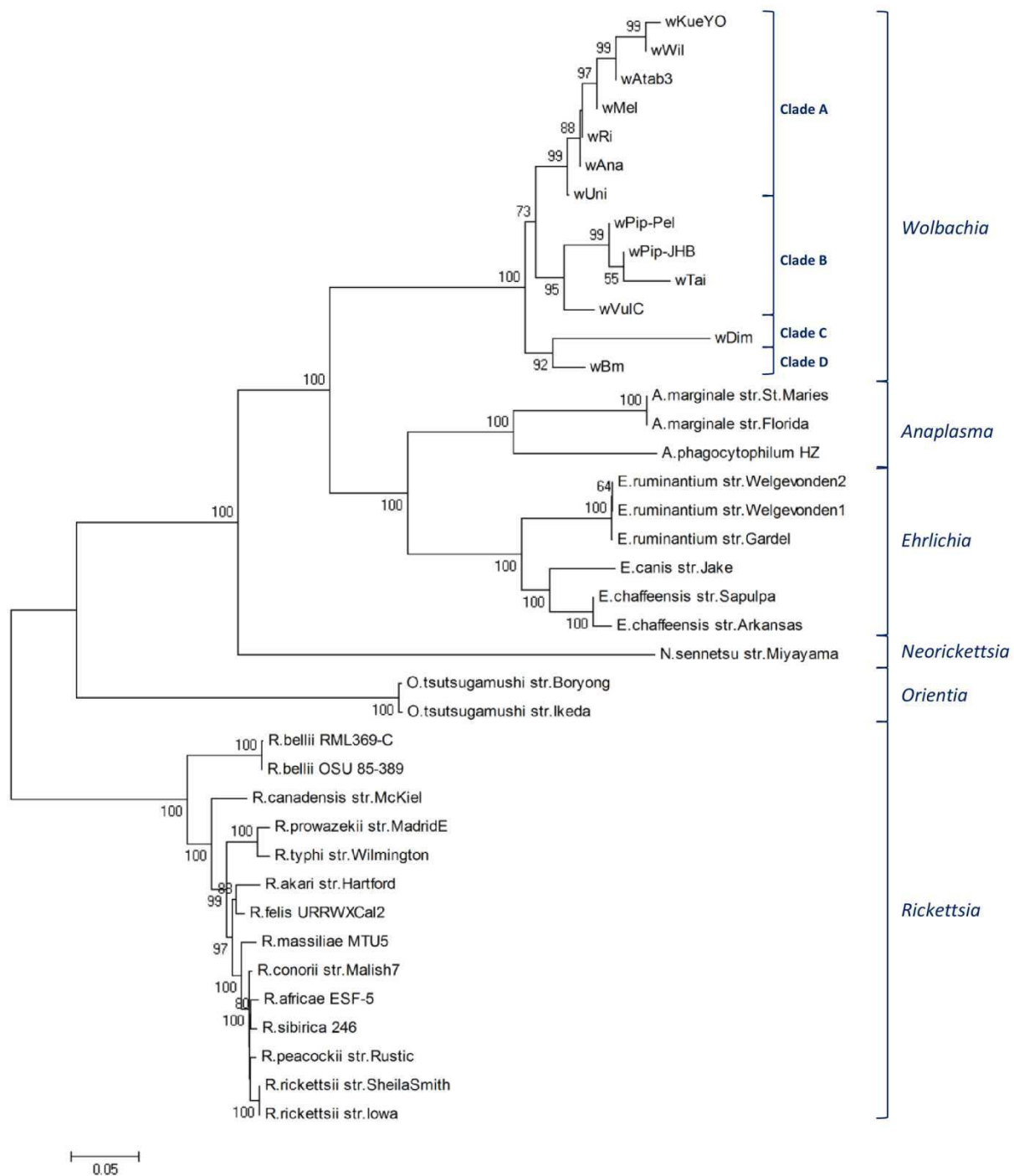


Figure 11: Phylogénie des Rickettsiales.

Arbre phylogénétique des Rickettsiales (39 souches) généré par Neighbor-Joining (distance) avec Mega4 à partir de l'alignement des séquences protéiques concaténées de cinq constituants du T4SS (VirB3,-B4,-B8,-B9,-B11). Les valeurs de bootstrap obtenues (1000 replicats) sont indiquées au niveau des nœuds au dessus de chaque branche. Les supergroupes ou clades de *Wolbachia* sont également indiqués.

Les résultats ont montré que les gènes du T4SS sont présents dans l'ensemble des souches de *Wolbachia* étudiées dans ce projet, quelque soit le phénotype induit (Fig. 1, Article 2). L'analyse des séquences indique que *virB3* et *virB11*, comme tous les autres gènes du T4SS sont des gènes extrêmement conservés au sein des *Wolbachia* et soumis à une sélection conservatrice (Table 1, Article 2). Ainsi, les pressions de sélection s'exerçant sur ce système de sécrétion restreignent la variabilité de séquences à l'échelle nucléique et donc limitent les changements de conformation des protéines. Ce résultat suggère le rôle majeur du T4SS dans la physiologie de *Wolbachia*. Cette notion prend d'autant plus de poids lorsque l'on sait que le génome de ces bactéries symbiotiques est soumis à des recombinaisons. Les opérons du T4SS représentent donc des portions de génomes d'au minimum 10 kb dont la structure et les séquences ont été fortement conservées depuis la divergence des Rickettsiales. A ce propos, la phylogénie réalisée avec les protéines les mieux conservées du T4SS (*VirB3*, *-B4*, *-B8*, *-9* et *-B11*) révèle une topologie similaire à celle obtenue avec d'autres marqueurs (Figure 11).

Si le nombre, l'ordre et la taille des gènes composant les opérons du T4SS sont conservés chez *Wolbachia*, les régions flanquantes en 5' et 3' de l'opéron *virB8-D4* sont variables (Fig. 1 et 2, Article 2). En effet, l'analyse des régions situées en aval de l'opéron *virB8-D4* comprend la première caractérisation de séquences d'insertion (IS) dans le génome de *wVulC* (Fig. 1 et 2, Article 2). Bien que réduits, les génomes de *Wolbachia* sont connus pour être exceptionnellement riches en éléments transposables qui sont tenus pour responsables de réarrangements occasionnés tout au long du génome notamment dans la région en 3' de *virD4* (Sanogo *et al.* 2007). Ce travail m'a aussi permis de mettre à profit le large spectre de souches de *Wolbachia* pour aider à caractériser une famille d'IS, ISWpi1, spécifique et très répandue dans les génomes de *Wolbachia* [cf. Article 3 : (Cordaux *et al.* 2008), p. 38]. Cette étude a montré que ces génomes sont soumis à des épisodes récurrents de prolifération et de pertes de séquences d'insertion (Fig. 1, Article 3).

De façon intéressante, certains gènes du T4SS sont dupliqués dans le génome de *Wolbachia*. Les génomes publiés de *Wolbachia* ont révélé l'existence des paralogues des gènes *virB4*, *virB6*, *virB8* et *virB9* (Wu *et al.* 2004, Foster *et al.* 2005, Klasson *et al.* 2008, Klasson *et al.* 2009) (Figure 12). Bien que détectés dans les contigs issus du séquençage de *wVulC* (Figure 12), les analyses par Southern Blot (Félix 2004) n'avaient pas permis d'identifier ces paralogues certainement à cause de leur faible pourcentage d'identité (34 à 57%). De plus, trois gènes *virB6* suivent directement en 3' le premier gène *virB6* de l'opéron *virB3-B6* (Fig. 2, Article 2). Nous ne savons pas encore si ces trois paralogues de *virB6* sont cotranscrits chez *wVulC* avec les autres gènes de l'opéron *virB3-B6* comme cela a été montré dans d'autres modèles (Ohashi *et al.* 2002, Wu *et al.* 2004, Rancès *et al.* 2008).

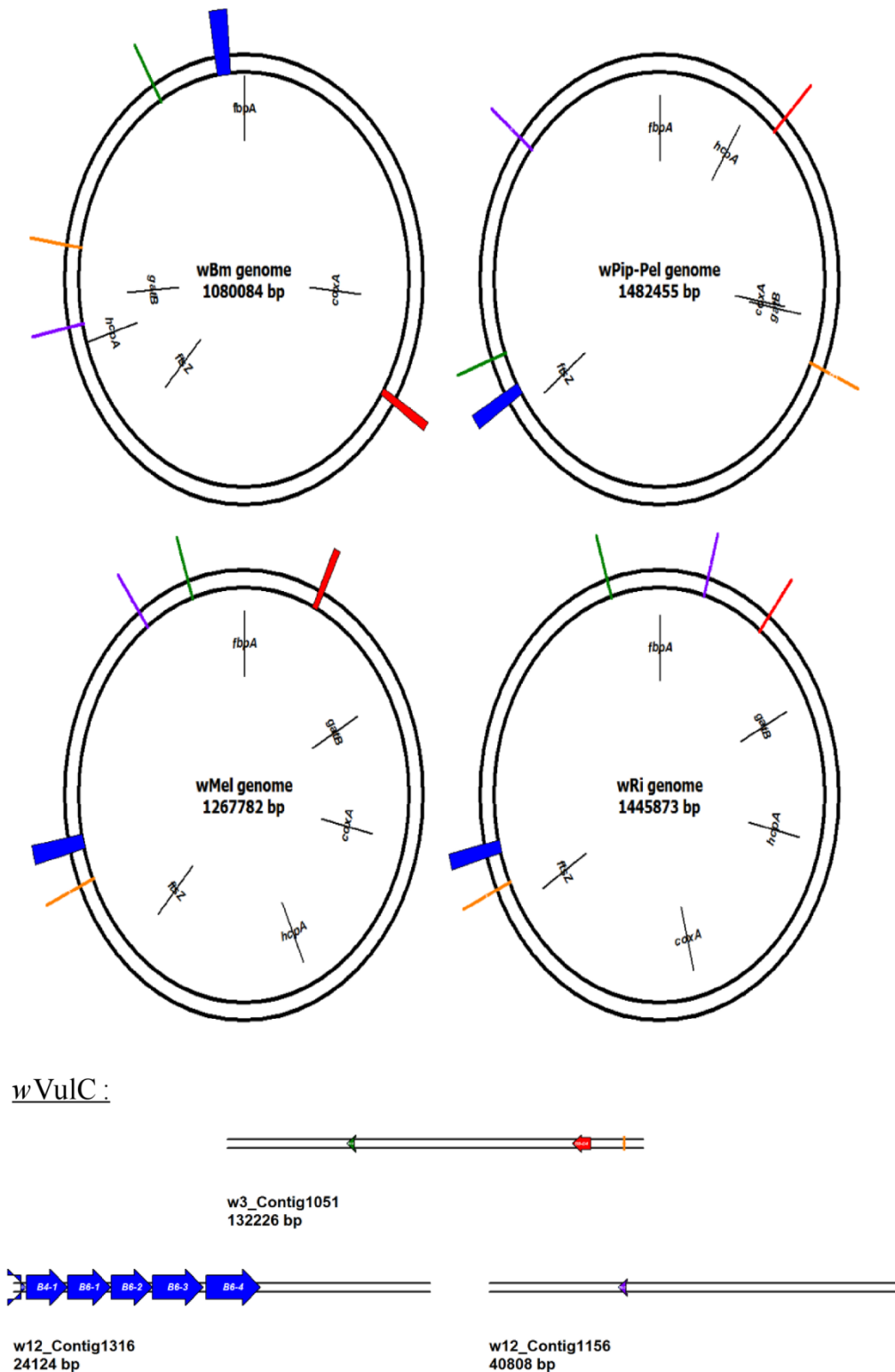


Figure 12: Cartographie génomique des gènes *vir* de *Wolbachia*.

A l'aide du logiciel pDRAW 3.2 (www.acaclone.com), sont représentées les localisations génomiques des opérons *vir* (*vir*B3-B6 en bleu et *vir*B8-D4 en rouge) ainsi que celles des paralogues des gènes *vir*B4-2 (vert), *vir*B8-2 (orange) et *vir*B9-2 (violet). Les régions *vir* identifiées chez wVulC sont représentées sous forme de contigs linéaires et sont dessinées à l'échelle. La localisation génomique de cinq gènes marqueurs (*fbpA*, *hcpA*, *coxA*, *gatB* et *ftsZ*) pour les quatre génomes publiés (wMel, wBm, wPip-Pel et wRi) est également indiquée. La base 1 des génomes circulaires est arbitrairement la première base du gène *fbpA*.

Cependant, les alignements que j'ai réalisés indiquent que les 4 paralogues de *virB6* sont ni alignables entre eux au sein d'une même souche de *Wolbachia* ni avec aucun *virB6* d'autres Rickettsiales. Néanmoins ces gènes ont été identifiés comme des paralogues de *virB6* car ils codaient des protéines possédant toutes un domaine TrbL caractéristique et étaient capables de s'enchâsser dans la membrane interne d'après les analyses de prédiction de localisation cellulaire. Une analyse plus poussée permettrait de déterminer si ces paralogues jouent un rôle dans la formation du T4SS des α -protéobactéries en augmentant la variabilité des constituants du pore ou bien s'ils ne sont que le produit de recombinaisons ou transpositions.

Différents mécanismes peuvent néanmoins être avancés quant à la présence de paralogues de gènes du T4SS :

- La sélection sur ces gènes est neutre et leur maintien dans le génome n'engendre aucune conséquence pour la bactérie.
- Les gènes sont maintenus par conversion génique comme cela a été supposé chez *Anaplasma* spp. (Palmer et Brayton 2007, Noh *et al.* 2008). Les produits de gènes sont alors interchangeables entre eux et ce grâce à la conservation de leur structure et de leur fonction (Seubert *et al.* 2003, de Paz *et al.* 2005).
- La divergence des gènes a permis de créer de nouveaux rôles utiles par exemple pour l'adaptabilité à de nouveaux hôtes ou conjointement à la nature des effecteurs sécrétés. Des études récentes montrent également dans le genre *Ehrlichia* l'implication des quatre paralogues de VirB6 dans la formation d'un complexe avec VirB9 et des protéines de la vacuole contenant la bactérie (Niu *et al.* 2006, Bao *et al.* 2009).

Notre approche de mesure des pressions de sélection sur les gènes du T4SS n'a pas permis de mettre en évidence une diversification des séquences au sein de différentes souches de *Wolbachia*. Au contraire, les paralogues sont soumis à des pressions de conservation tout comme les autres gènes du T4SS. Aucune marque de sélection positive n'a pu être mise en évidence sur les régions flanquantes des opérons. Tout cela suggère que les réarrangements génomiques autour des opérons sont certainement anciens et indépendants de la nature des régions flanquantes des opérons du T4SS chez *Wolbachia*.

D. Conclusions et perspectives.

Dans tous les tissus d'*A. vulgare* où *Wolbachia* est présente, les gènes du T4SS sont exprimés, cotranscrits et codent des protéines conservées. Chez différentes souches hébergées par des arthropodes, ces gènes ont la même organisation en opérons que ceux décrits chez d'autres *Wolbachia* (Masui *et al.* 2000, Wu *et al.* 2004, Foster *et al.* 2005, Klasson *et al.* 2008, Rancès *et al.* 2008, Klasson *et al.* 2009). Pour montrer le caractère fonctionnel du T4SS de *wVulC*, d'autres approches sont nécessaires. Il faudra par exemple localiser au niveau membranaire ou périplasmique les produits Vir bactériens (Christie 1997, Judd *et al.* 2005). Cela permettra de démontrer que les protéines sont bien traduites et effectivement conduites aux membranes où se réalise leur assemblage. Or, un anticorps polyclonal spécifique de VirB6 a été utilisé par nos collègues de Lyon (Drs P. Mavingui et E. Rancès) (Rancès *et al.* 2008). Nous pourrions donc envisager d'« immunolocaliser » les protéines VirB6 au niveau des membranes de la bactérie *wVulC*. Puisqu'à ce jour il n'est pas possible d'obtenir de mutants chez *Wolbachia*, une des clés permettant de comprendre le rôle de ce système de sécrétion est l'identification des molécules effectrices passant à travers le canal. Nous avons donc mis à profit les résultats du séquençage de *wVulC*, ainsi que les données génomiques disponibles, pour tenter d'identifier des toxines potentielles.

E. Publications.

E.1 - Félix *et al.* 2008.

Article 1: Félix C., Pichon S., Braquart-Varnier C., Braig H. Chen L., Garrett R.A., Martin G. and P. Grève. 2008. Characterisation and transcriptional analysis of two gene clusters for type IV secretion machinery in *Wolbachia* of *Armadillidium vulgare*. *Research in Microbiology* 159: 481-485.



ELSEVIER



Research in Microbiology 159 (2008) 481–485



www.elsevier.com/locate/resmic

Characterization and transcriptional analysis of two gene clusters for type IV secretion machinery in *Wolbachia* of *Armadillidium vulgare*

Christine Félix^a, Samuel Pichon^a, Christine Braquart-Varnier^a, Henk Braig^b, Lanming Chen^c,
Roger A. Garrett^c, Gilbert Martin^a, Pierre Grève^{a,*}

^a Université de Poitiers, Laboratoire Ecologie, Evolution, Symbiose, UMR CNRS 6556, 40 avenue du recteur Pineau, F-86022 Poitiers Cedex, France

^b University of Wales Bangor, School of Biological Sciences, Deiniol Road, LL57 2UW Bangor, UK

^c University of Copenhagen, Institute of Molecular Biology, Biocenter, Ole Maaløes Vej 5, 2200 N Copenhagen, Denmark

Received 29 February 2008; accepted 15 May 2008

Available online 6 June 2008

Abstract

Wolbachia are maternally inherited α -proteobacteria that induce feminization of genetic males in most terrestrial crustacean isopods. Two clusters of *vir* genes for a type IV secretion machinery have been identified at two separate loci and characterized for the first time in a feminizing *Wolbachia*. Furthermore, we demonstrated that these operons are transcriptionally active in ovaries and in all other tissues tested, suggesting that T4SS has a significant role in *Wolbachia* biology. These observations and the identification of homologous *vir* genes in *Wolbachia* strains infecting insects or nematodes show that *vir* genes are conserved among *Wolbachia* strains whatever the phenotype induced by the bacteria. © 2008 Elsevier Masson SAS. All rights reserved.

Keywords: Crustacean; Isopod; *Wolbachia*; Feminization; Cytoplasmic incompatibility; T4SS; *vir* Genes

1. Introduction

Bacteria belonging to the *Wolbachia* genus have been recognized to infect a wide range of arthropods and some nematodes [16]. *Wolbachia* are maternally inherited intracellular bacteria that remain difficult to culture. These Gram-negative bacteria belong to the group of α -proteobacteria and its closest known relative groups are *Rickettsia*-like bacteria.

Wolbachia are able to invade and maintain themselves in numerous arthropod host species by inducing a variety of reproductive alterations including male feminization in isopod crustacean species [17], parthenogenetic development in wasps or mites, male-killing in insect species and cytoplasmic incompatibility (CI) in isopod or insect species [5]. To date, nothing is known about bacterial and host genes involved in induction of these reproductive alterations, and molecular mechanisms have not yet been elucidated.

Feminization is a phenomenon whereby infected genetic males are converted into functional females by inhibition of androgenic gland development [17]. Since *Wolbachia* are exclusively localized in the host cell cytoplasm, it is believed that the bacterium secretes macromolecules in order to reverse sexual differentiation. Many Gram-negative bacteria such as *Agrobacterium tumefaciens*, *Helicobacter pylori*, *Bordetella pertussis*, *Legionella pneumophila*, *Brucella* spp. or *Bartonella* spp. use a type IV secretion system (T4SS) to deliver virulence factors within the cytoplasm of host cells [8]. Amongst obligate intracellular bacteria, T4SS has been found previously in *Rickettsia prowazekii* [1] and characterized in *Rickettsia*-like bacteria such as *Anaplasma phagocytophila* and *Ehrlichia chaffeensis*, where eight *virB* and *virD* genes were clustered at two separate loci and polycistronically transcribed [15]. Furthermore, five genes coding for the T4SS (*virB8*, *virB9*, *virB10*, *virB11* and *virD4*) were identified in *wTai* and in *wKueYO*, which are *Wolbachia* symbionts of the Taiwan cricket and the Mediterranean flour moth, respectively [13]. More recently, the complete genome sequence of different *Wolbachia* strains infecting *Drosophila melanogaster* (*wMel*)

* Corresponding author. Tel.: +33 5 49 45 39 81; fax: +33 5 49 45 40 15.
E-mail address: pierre.greve@univ-poitiers.fr (P. Grève).

and *Drosophila ananassae* (wAna), or the nematode *Brugia malayi* (wBm) [10], also revealed complete T4SS as two operons as in *Anaplasma* and *Ehrlichia* [15].

Amongst *Wolbachia*, T4SS has only been identified in the CI *Wolbachia* of insects and in the mutualistic *Wolbachia* of nematodes. Therefore, in this study, we checked for the presence of the T4SS in isopod *Wolbachia* strains that induce another phenotype, the feminizing wVul, wAse and wPru harbored by *Armadillidium vulgare*, *Oniscus asellus* and *Porcellionides pruinosus*, respectively [17]. The presence of T4SS in a CI *Wolbachia* of isopod, wPet harbored by *Porcellio dilatatus petiti*, has also been tested.

2. Materials and methods

2.1. *Wolbachia* strains

Uninfected and *Wolbachia*-infected terrestrial isopods *A. vulgare*, *O. asellus*, *P. dilatatus petiti*, *P. pruinosus* were reared in the laboratory at 20 °C, on moistened soil with dead leaves and carrots as food. *Wolbachia* infection was checked by PCR using specific *wsp* primers (Supplementary data 1).

2.2. Amplification and sequencing of *vir* genes

DNA extractions were performed from ovaries or testes as described elsewhere [4]. Several degenerated primer pairs were designed based on the conserved sequences of *vir* genes among *A. phagocytophila*, *E. chaffeensis*, wTai, wKueYO, wMel, wAna or wBm (Supplementary data 1). The DNA fragments amplified by PCR (40 cycles at 95 °C for 1 min, 50 °C for 1 min, 72 °C for 1 min per kb) using these primers were sequenced by a dideoxy termination method using the kit “ABI PRISM™” (Applied Biosystems), according to the manufacturer’s instruction, and using the Walking chromosome method for long PCR fragments (ESGS or QIAGEN Companies). The sequence data from this study has been submitted to GenBank under accession numbers AY967767 (*phosphatase-virD4*) and AY967766 (*virB3–virB6*). The missing 3'-ends of *virB6* and *virD4* genes were obtained from shotgun library clones (whole-genome shotgun-sequencing project: European *Wolbachia* project EUWOL). A database search was carried out with the BLAST program (<http://www.ncbi.nlm.nih.gov/BLAST>).

2.3. RNA isolation and transcriptional analysis

Total RNA was isolated from different tissues (ovaries, hematopoietic organs, hemocytes, nerve chains, brains, gut and adipose tissue) using the RNeasy Mini-kit according to the manufacturer’s instructions (Qiagen) and reverse-transcribed using a Thermotranscript kit (Invitrogen-Life Technologies) with random hexamer primers at 25 °C for 10 min and at 50 °C for 1 h. PCR was performed under conditions described above using primers as indicated (Fig. 3, Supplementary data 1). Nested

PCRs were also realized using internal primers to reveal *vir* gene expression in nerve tissues.

2.4. Genomic Southern blot analysis

Genomic DNA extracted from infected ovaries (20 µg) was digested with restriction endonucleases, electrophoresed and subjected to Southern blotting. The 4.67 kb *virB8–virD4* (*virB8*-1F/*virD4*R) and the 2.83 kb *virB3–virB6* (*virB3*-F/5'*virB6*-R) PCR fragments were labelled with [α -³²P]dCTP by the random primer method and hybridized to nylon membranes. Hybridized blots were imaged and analyzed using a phosphorimager (Molecular Dynamics, Sunnyvale, CA, U.S.A.).

3. Results and discussion

Initially, we successfully amplified the *virD4* gene only in infected ovaries or testes of different isopod species (Supplementary data 2). Hence, based on the conserved sequences of several *vir* genes of different *Wolbachia* strains, several primer pairs were designed and used to amplify DNA fragments corresponding to two *vir* gene clusters (Fig. 1a). The first one, named *virB8–virD4* operon, was composed of five

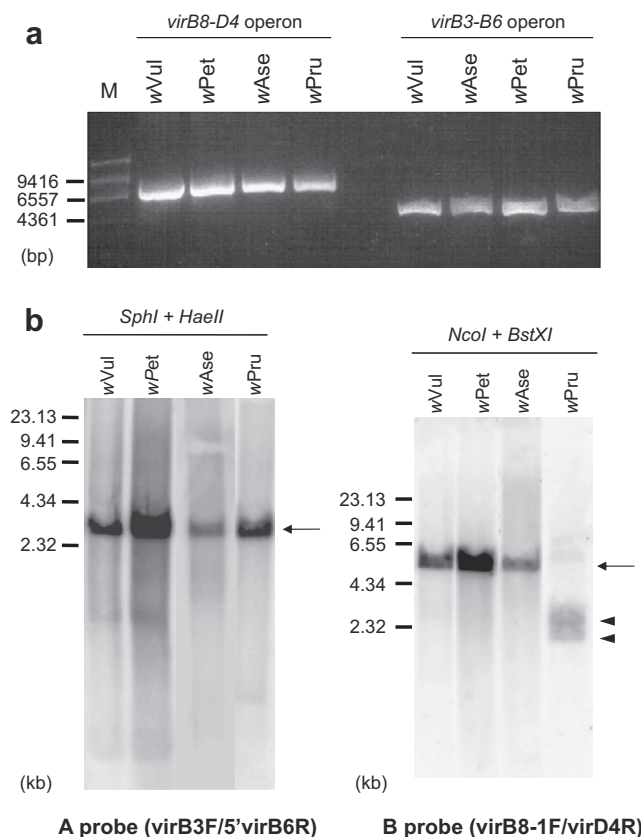


Fig. 1. Detection of *virB* and *virD* genes by PCR and Southern blotting. Detection of *virB3–virB6* and *virB8–virD4* operons by PCR (a) using *virB3F/virB6R* and *phosph-1F/virD4R* primer pairs or by Southern blots (b) using A and B probe, respectively, in different *Wolbachia* strains harbored by *A. vulgare* (wVul), *O. asellus* (wAse), *P. pruinosus* (wPru), *P. dilatatus petiti* (wPet). M: DNA size markers. Arrows and arrowheads indicate hybridized fragments.

tandem genes (*virB8*, *virB9*, *virB10*, *virB11* and *virD4*), and the second one, named the *virB3*–*virB6* operon, was composed of three tandem genes (*virB3*, *virB4* and *virB6*) (Fig. 2). Southern blot analysis indicated that these two loci were not contiguous on the *wVul* genome (Fig. 1b). Identical PCR and Southern blot results were found with two other feminizing *Wolbachia* strains, *wAse* and *wPru*, and with the CI *Wolbachia*, *wPet* (Fig. 1a and b). In *wPru*, the *virB8*–*virD4* operon should have an additional *NcoI* or a *BstXI* restriction site, since two bands of about 2.15 and 2.5 kb have been observed instead of a single 4.67 kb band as determined for other strains (Fig. 2b). The gene organization of the *virB* and *virD* regions in two separate loci was similar to the corresponding region of the T4SS in *wMel*, *wAna* and *wBm* [10]. Furthermore, this gene organization in several loci seems to be conserved among obligate intracellular bacteria: two loci have also been observed in *Anaplasma* sp. and *Ehrlichia* sp. [6,9,15], and three loci in *Rickettsia* sp. [3]. The localization of the *virB* and *virD* genes in two separate loci suggested that two conditions may be fulfilled to trigger the complete T4SS assembly [15].

VirB and VirD orthologous proteins of *wVul* had highest amino acid identities with those of *Wolbachia* sp. (69.8–96.9%), and to a lesser extent with those of *E. chaffeensis*, *A. phagocytophila* and *R. prowazekii* (17.9–77.4%), all these

bacteria being obligate intracellular α -proteobacteria (Supplementary data 4). Among *Wolbachia* sp., Vir proteins of *wVul* were closer to those of *wTai* than to those of *wKueYO*, *wMel*, *wAna* or *wMel*, except for VirB10. Of all these Gram-negative intracellular bacteria, the more conserved proteins were VirD4, VirB4 and VirB11, which shared, respectively, from 28.3, 30.9 and 28.5% identities with *A. tumefaciens* to 94.3, 94.4 and 93% identities with *wMel* or *wTai*, whereas the more variable protein was VirB8, with identities ranging from 21.2% with *R. prowazekii* to 93.8% with *wTai* (Supplementary data 4). These proteins are essential components of the T4SS involved in pore structures or in channel assembly [8]. All these rickettsial strains, including *Wolbachia*, seem to lack genes orthologous to *virB1*, *virB2* and *virB5*, and most of them lack orthologs of the *virB7* genes of *A. tumefaciens*. VirB1 of *A. tumefaciens* has a motif common to lytic transglycosylases and it may be involved in the local lysis of peptidoglycan during transporter assembly [11]. In *B. pertussis*, it has been shown that this process may be facilitated by VirB8 [18], but in *wVul*, VirB8 does not carry any motifs characteristic of lytic transglycosylases. Whereas VirB7 is found in only a subset of T4SS, VirB5 and especially VirB2 are components of the pilus. Their absence therefore indicates that the *wVul* T4SS is not involved in attachment to the host cell, as observed for many free-living or facultative intracellular bacteria. These observations are consistent with the strict intracellular lifestyle of *Wolbachia* and we can thus hypothesize that the T4SS of *wVul* is more likely implicated in intracellular survival or in the localization of the bacteria in an intracellular niche.

The Southern blot result also revealed the absence of conserved *vir* paralogs in all four *Wolbachia* strains tested. The number of *virB4* paralogs seems to be highly variable, since isopod *Wolbachia* strains and *E. chaffeensis* have a single *virB4* gene, whereas *wMel*, *wBm*, *Rickettsia* sp., *Anaplasma* sp. and *E. ruminantium* have an additional *virB4* paralog [1,3,6,9,10,15]. The gene *virB6* seems to have been duplicated several times in *wAna*, *wBm* and the *R. prowazekii* genomes, where a cluster of four or five *virB6* paralogs is observed downstream from the *virB3*–*virB6* operon. A second *virB8* gene has also been identified in each of these genomes. In any case, all these paralogs are quite divergent; while, in *wMel* the *virB8* gene is not transcribed, the three *virB6* paralogs are expressed as part of the *virB3*–*virB6* operon. It has been hypothesized that these three membrane-spanning proteins may constitute putative additional components of the *Wolbachia* T4SS (Ref. in [10]).

Two genes upstream from the *virB8*–*virD4* cluster were also identified: one, in the opposite orientation, encodes a Ser/Thr protein phosphatase and the other, *ribA*, encodes a bi-functional enzyme with the activities of both 3,4-dihydroxy-2-butanone-4-phosphate synthase and GTP cyclohydrolase II, which catalyze two essential steps in riboflavin biosynthesis (Fig. 2). A *ribA* gene is also localized upstream of the *virB8*–*virD4* operon of *Ehrlichia* sp., *Anaplasma* sp., *wMel*, *wAna* and *wBm* [10,15]. The intergenic space between *ribA* and *virB8* (72 bp) genes in *wVul* was shorter than that in

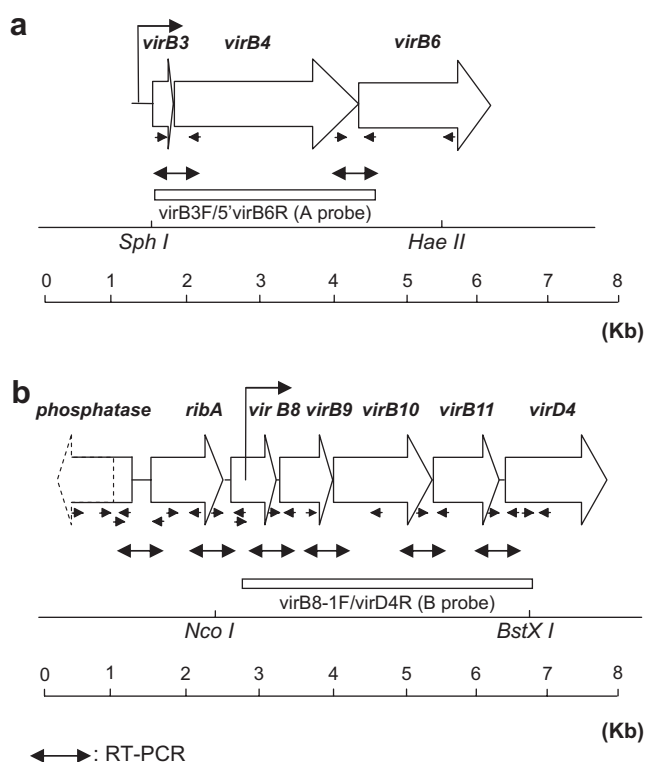


Fig. 2. Schematic representation of the gene organization of the *virB* and *virD* regions in *wVul*. ORFs are represented as open boxes, with arrowheads indicating their orientation. Lines with (non-facing) arrows at both ends indicate the RT-PCR regions in the transcriptional analysis. Bent arrows located upstream of *virB3* or *virB8* gene represent putative transcriptional promoter regions. Open boxes indicate DNA probes used for genomic Southern blot analysis. The dotted line represents an unsequenced region.

E. chaffeensis (134 bp), where putative sequences of promoter regions for *virB8* were found upstream of the transcriptional initiation sites. One of them was likely to be located within the 3' coding sequence of the *ribA* gene, suggesting that the expression of *ribA* may influence the transcription of clustered *virB8–virD4* genes [14]. However, the transcriptional regulation mechanisms of the *wVul* *vir* genes remain to be elucidated. A transcriptional regulator named EcxR has recently been identified in *E. chaffeensis* [7]. This transcription factor was shown to activate the expression of all *vir* operons during intracellular development. EcxR homologs are present in the

genomes of *wBm* (58% identity) and *wMel* (50% identity). Further works will be needed to identify EcxR homologs in *wVul* and to investigate its role in the regulation of *vir* gene expression in *Wolbachia*.

We examined the expression of *vir* genes in *A. vulgare* ovaries by RT-PCR using overlapping primers (Fig. 3a and Supplementary data 2). A PCR fragment corresponding to each *vir* gene pair was detected and showed that *virB8*, *virB9*, *virB10*, *virB11* and *virD4* genes were co-transcribed as an operon as well as the *virB3*, *virB4* and *virB6* genes. By the same approach, we showed that the *phosphatase* and *ribA* genes

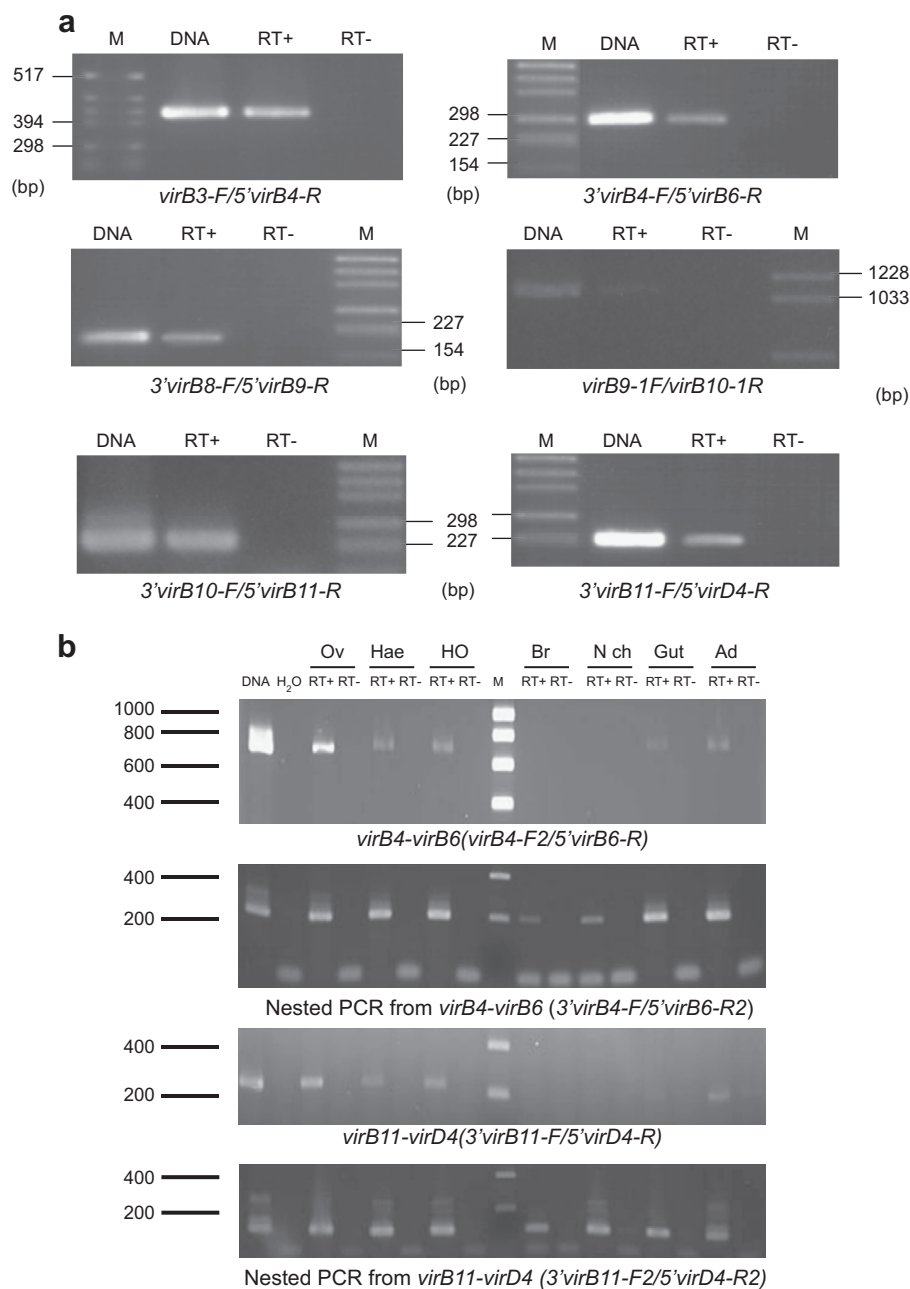


Fig. 3. Transcriptional analysis of clustered *virB* and *virD* genes in ovaries (a) and seven different host tissues (b). RT-PCR was used for transcriptional analysis of the genes in the *virB* and *virD* regions. Primers used and their positions are shown in Supplementary data 1 and Fig. 2. The DNA template control shows the intensity and specificity of the band detected with each pair of primers. RT+ and RT- indicate, respectively, the presence and absence of reverse transcriptase in the reaction. Ov: ovaries; Hae: hemocytes; HO: hematopoietic organ; Br: brain; N ch: nerve chain; Gut: gut; Ad: adipose tissue; and M: DNA size markers.

were independently transcribed from the *virB8*–*virD4* operon (Supplementary data 3). Similar results were obtained for *vir* genes of *wTai*, *wKueYO* or *wMel*, suggesting that T4SS may be functional in *Wolbachia* [10,13], as well as in those of *A. phagocytophila* and *E. chaffeensis* [15]. We further examined *vir* gene expression in other female tissues known to be infected by *Wolbachia* at a very low density. In all tissues tested, *vir* gene expression was revealed either by RT-PCR or by nested PCR using internal primers (Fig. 3b). This is the first study in *Wolbachia* showing the expression of *vir* genes in all infected tissues. Further experiments, including immunolocalization of Vir proteins, will be needed to check the functionality of T4SS. Taken together, these results, along with *vir* gene expression in hemocytes and hematopoietic organs, suggest that T4SS may play an important role in the biology of *Wolbachia*.

In conclusion, *wVul* possesses the minimum components of typical T4SS, consisting of *virB4*, *virB7* (or *virB8*), *virB9*, *virB10*, *virB11* and *virD4* homologs [19] which are co-transcribed in oocytes. It shows that feminizing and CI *Wolbachia* of isopods, like CI *Wolbachia* of insects, or the mutualistic *Wolbachia* of nematodes, do carry the necessary genetic components for a functional T4SS machinery. It is conceivable that T4SS may be involved in maintenance of *Wolbachia* infection and *Wolbachia*-induced reproductive alterations such as feminization, cytoplasmic incompatibility, male-killing or parthenogenesis, by secreting virulence factors, as has been demonstrated for several other Gram-negative bacteria [2]. New approaches such as protein interaction maps that have been applied to the genome of *Rickettsia sibirica* [12] or yeast two hybrid experiments similar to those realized in *L. pneumophila* [14] are needed to identify these factors used by *Wolbachia* to manipulate eukaryotic cellular processes in order to traffic and replicate in host cells.

Acknowledgments

CF was the recipient of a European postgraduate fellowship. This work was supported by an EEC grant (Euwol, QLRT-2000-01079) and a CNRS grant (ANR-06-BLAN-0316). We are grateful to C. Debenest, C. Delaunay and M. Frelon-Raimond for valuable technical aid at different levels of this study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.resmic.2008.05.007

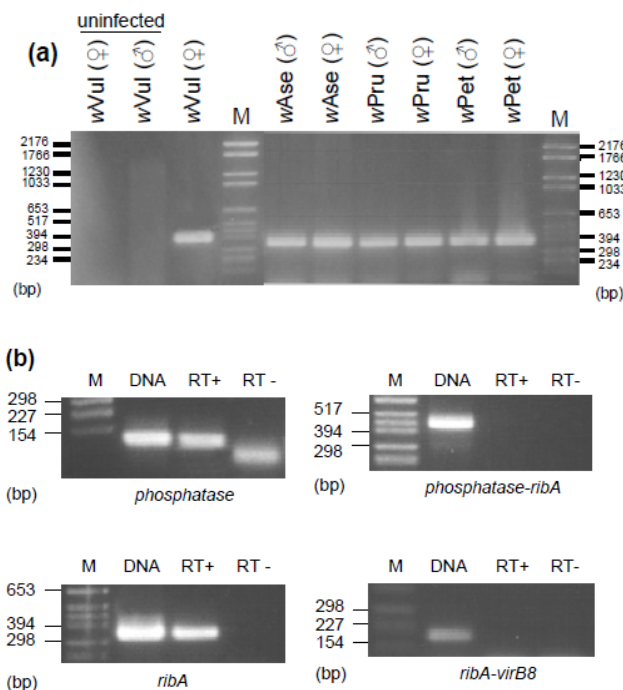
References

- [1] Andersson, S.G., Zomorodipour, A., Andersson, J.O., Sicheritz-Ponten, T., Alsmark, U.C., Podowski, R.M., et al. (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396, 133–140.
- [2] Backert, S., Meyer, T.F. (2006) Type IV secretion systems and their effectors in bacterial pathogenesis. *Curr Opin Microbiol* 9(2), 207–217.
- [3] Blanc, G., Ogata, H., Robert, C., Audic, S., Claverie, J.M., Raoult, D. (2007) Lateral gene transfer between obligate intracellular bacteria: evidence from the *Rickettsia massiliae* genome. *Genome Res* 17(11), 1657–1664.
- [4] Bouchon, D., Rigaud, T., Juchault, P. (1998) Evidence for widespread *Wolbachia* infection in isopod crustaceans: molecular identification and host feminization. *Proc R Soc Lond B Biol Sci* 265, 1081–1090.
- [5] Bourtzis, K., Miller, T.A. (2003) *Insect Symbiosis*. Boca Raton: CRC Press.
- [6] Brayton, K.A., Kappmeyer, L.S., Herndon, D.R., Dark, M.J., Tibbals, D.L., Palmer, G.H., et al. (2005) Complete genome sequencing of *Anaplasma marginale* reveals that the surface is skewed to two super-families of outer membrane proteins. *Proc Natl Acad Sci USA* 102(3), 844–849.
- [7] Cheng, Z., Wang, X., Rikihisa, Y. (2008) Regulation of type IV secretion apparatus genes during *Ehrlichia chaffeensis* intracellular development by a previously unidentified protein. *J Bacteriol* 190(6), 2096–2105.
- [8] Christie, P.J., Atmakuri, K., Krishnamoorthy, V., Jakubowski, S.J., Cascales, E. (2005) Biogenesis, architecture and function of bacterial type IV secretion systems. *Annu Rev Microbiol* 59, 451–485.
- [9] Collins, N.E., Liebenberg, J., de Villiers, E.P., et al. (2005) The genome of the heartwater agent *Ehrlichia ruminantium* contains multiple tandem repeats of actively variable copy number. *Proc Natl Acad Sci USA* 102(3), 838–843.
- [10] Fenn, K., Blaxter, M. (2006) *Wolbachia* genomes: revealing the biology of parasitism and mutualism. *Trends Parasitol* 22(2), 60–65.
- [11] Llosa, M., Zupan, J., Baron, C., Zambryski, P. (2000) The N- and C-terminal portions of the *Agrobacterium* VirB1 protein independently enhance tumorigenesis. *J Bacteriol* 182, 3437–3445.
- [12] Malek, J.A., Wierzbowski, J.M., Tao, W., Bosak, S.A., Saranga, D.J., Doucette-Stamm, L., et al. (2004) Protein interaction mapping on a functional shotgun sequence of *Rickettsia sibirica*. *Nucleic Acids Res* 32, 1059–1064.
- [13] Masui, S., Sasaki, T., Ishikawa, H. (2000) Genes for the type IV secretion system in an intracellular symbiont, *Wolbachia*, a causative agent of various sexual alterations in arthropods. *J Bacteriol* 182, 6529–6531.
- [14] Ninio, S., Roy, C.R. (2007) Effector proteins translocated by *Legionella pneumophila*: strength in numbers. *Trends Microbiol* 15(8), 372–380.
- [15] Ohashi, N., Zhi, N., Lin, Q., Rikihisa, Y. (2002) Characterization and transcriptional analysis of gene clusters for a type IV secretion machinery in human granulocytic and monocytic ehrlichiosis agents. *Infect Immun* 70, 2128–2138.
- [16] O'Neill, S.L., Hoffmann, A.A., Werren, J.H. (1997) *Influential Passengers: Inherited Microorganisms and Arthropod Reproduction*. New York: Oxford University Press.
- [17] Rigaud, T. (1997). In: S.L. O'Neill, J.H. Werren, & A.A. Hoffmann (Eds.), *Influential Passengers: Inherited Microorganisms and Arthropod Reproduction* (pp. 81–101). Oxford: Oxford University Press.
- [18] Ward, D.V., Draper, O., Zupan, J.R., Zambryski, P.C. (2002) Peptide linkage mapping of the *Agrobacterium tumefaciens* vir-encoded type IV secretion system reveals protein subassemblies. *Proc Natl Acad Sci USA* 99, 11493–11500.
- [19] Winans, S.C., Burns, D.L., Christie, P.J. (1996) Adaptation of a conjugal transfer system for the export of pathogenic macromolecules. *Trends Microbiol* 4, 64–68.

Target genes	Amplicon sizes (bp)	5' primers		3' primers	
		Nucleotide sequences	code	Nucleotide sequences	code
<i>virB3-virB6</i>	4658	5'-GACCTMCVATGCTBTTTGGTG-3'	virB3-F*	5'-AAASGCAAAAGKATWACMGGRTA-3'	virB6-R*
	2827	5'-GACCTMCVATGCTBTTTGGTG-3'	virB3-F*	5'-CAATCAGGATTAGAGGCACGAC-3'	5'virB6-R
<i>virB3-virB4</i>	530	5'-GATATAGTATTAGTGTAGAGC-3'	virB3-43F*	5'-CCTCRTCATCKGCAGTTTCAA-3'	5'virB4-R*
	440	5'-GACCTMCVATGCTBTTTGGTG-3'	virB3-F*	5'-CCTCRTCATCKGCAGTTTCAA-3'	5'virB4-R*
<i>virB4-virB6</i>	269	5'-CTGGACGTGCAGAAAGTGTCT-3'	3'virB4-F	5'-CAATCAGGATTAGAGGCACGAC-3'	5'virB6-R
<i>Phosphatase</i>	693	5'-AAAAATTCATCGGTCCAGCATCTAAG-3'	phosph-F	5'-CTAGTTTTTCATTTTCGTCATGCAC-3'	phosph-R
	108	5'-CAATAGGACACATGTCATTGTC-3'	phosph-1F	5'-CTAGTTTTTCATTTTCGTCATGCAC-3'	phosph-R
<i>ribA</i>	337	5'-GTTATTGCACTGGACACGTCATTC-3'	ribA-1F	5'-CAATGCCTCTTCCATCTTGATC-3'	ribA-1R
<i>virD4</i>	382	5'-CAGAGAAAGACATACGAAAAGCAGG-3'	virD4-F	5'-CTTGTTTCAGGCATGATTAGGTTAGC-3'	virD4-R
<i>ribA-virD4</i>	6357	5'-CAATAGGACACATGTCATTGTC-3'	phosph-1F	5'-CTTGTTTCAGGCATGATTAGGTTAGC-3'	virD4-R
<i>virB8-virD4</i>	4675	5'-CGCTACAGCACAGTTATTGCTC-3'	virB8-1F	5'-CTTGTTTCAGGCATGATTAGGTTAGC-3'	virD4-R
<i>phosphatase-ribA</i>	431	5'-TTTGTGCATGGACGAAATGAAAC-3'	3'phosph-F	5'-GGCCACTTCTAATTCGCTGATAGC-3'	5'ribA-R
<i>ribA-virB8</i>	190	5'-CTAGCCATGGATTAAGGGTTTTTAG-3'	3'ribA-F	5'-GAGCAATAACTGTGCTGTAGCG-3'	5'virB8-R
<i>virB8-virB9</i>	219	5'-GCATCGCTTGAATGAATGATCAG-3'	3'virB8-F	5'-CCGTAATACCTCATTAGGCT-3'	5'virB9-R
<i>virB9-virB10</i>	1089	5'-GAAGTCAGTAGTCGCACTAACATG-3'	virB9-1F	5'-GCCTTGTCTCTCCACTACTACC-3'	virB10-1R
<i>virB10-virB11</i>	230	5'-GATATAGTATTTCTCCACAGGC-3'	3'virB10-F	5'-CTTGAGCAATCAGCCTGCC-3'	5'virB11-R
<i>virB11-virD4</i>	224	5'-GAAGTGTAGGGGCAGAGAATC-3'	3'virB11-F	5'-CACACTATCTGGACCGTCAAC-3'	5'virD4-R

Table S1: Oligonucleotide primers used in PCR and RT-PCR.

wsp specific primers (Braig *et al.* 1998) were used to check *Wolbachia* infection status of samples. *: Sequences of degenerated primers according to IUB codes [International Union of Biochemistry].

**Figure S1: Other PCR results.**

Specific PCR detection of the *virD4* gene using *virD4F* and *virD4R* primers in different *Wolbachia* strains harbored by *A. vulgare* (*wVulC*), *O. asellus* (*wAse*), *P. prunosus* (*wPru*), *P. dilatatus petiti* (*wPet*). The sequence data for the *virD4* genes of *wAse*, *wPru* and *wPet* has been submitted to GenBank under accession numbers AY967768, AY967770 and AY967769, respectively. M: DNA size marker. Transcriptional analysis of the *ribA* and *phosphatase* genes in ovaries. RT-PCR was used for transcriptional analysis of the genes in the *virB* and *virD* regions. Primers used and their positions are shown in Table S1 and Fig. 2. The DNA template control shows the intensity and specificity of the band detected with each pair of primers. RT+ and RT- indicate respectively the presence and absence of reverse transcriptase in the reaction. M: DNA size marker.

E.2 - Pichon *et al.* 2009.

Article 2: Pichon S., Bouchon D., Cordaux R., Chen L., Garrett R.A. and P. Grève. Conservation of the Type IV Secretion System throughout *Wolbachia* evolution. *Biochemical and Biophysical Research Communications* 385: 557-562.



Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc



Conservation of the Type IV Secretion System throughout *Wolbachia* evolution

Samuel Pichon^a, Didier Bouchon^a, Richard Cordaux^a, Lanming Chen^b, Roger A. Garrett^b, Pierre Grève^{a,*}

^a Université de Poitiers, UMR CNRS 6556 Ecologie, Evolution, Symbiose, 40 avenue du Recteur Pineau, F-86022 Poitiers, France

^b Department of Molecular Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 N Copenhagen, Denmark

ARTICLE INFO

Article history:

Received 18 May 2009

Available online 30 May 2009

Keywords:

Type IV Secretion System
Structure conservation
Wolbachia pipientis
Endosymbiosis
Genomic plasticity
Insertion sequence
Alphaproteobacteria
Rickettsiales

ABSTRACT

The Type IV Secretion System (T4SS) is an efficient pathway with which bacteria can mediate the transfer of DNA and/or proteins to eukaryotic cells. In *Wolbachia pipientis*, a maternally inherited obligate endosymbiont of arthropods and nematodes, two operons of *vir* genes, *virB3–B6* and *virB8–D4*, encoding a T4SS were previously identified and characterized at two separate genomic loci. Using the largest data set of *Wolbachia* strains studied so far, we show that *vir* gene sequence and organization are strictly conserved among 37 *Wolbachia* strains inducing various phenotypes such as cytoplasmic incompatibility, feminization, or oogenesis in their arthropod hosts. In sharp contrast, extensive variation of genomic sequences flanking the *virB8–D4* operon suggested its distinct location among *Wolbachia* genomes. Long term conservation of the T4SS may imply maintenance of a functional effector translocation system in *Wolbachia*, thereby suggesting the importance for the T4SS in *Wolbachia* biology and survival inside host cells.

© 2009 Elsevier Inc. All rights reserved.

Introduction

Wolbachia pipientis is an unculturable obligate endosymbiont belonging to the α -Proteobacteria (Rickettsiales). They are estimated to be present in ~60% of arthropod species [1,2] and in filarial nematodes [3]. The interactions between *Wolbachia* and their hosts range from pathogenesis to parasitism and mutualism. In many arthropod species *Wolbachia* parasite host reproduction by inducing cytoplasmic incompatibility, feminization, male killing and parthenogenesis to enhance their transmission and spread in hosts [1,4]. Thus, intracellular lifestyle implies communication between *Wolbachia* and their host cells.

Many symbiotic and pathogenic bacteria are able to export DNA and/or proteins into eukaryotic host cells. Such bacteria use a wide range of mechanisms for the transport of proteins and/or DNA across cell membranes, including Type IV Secretion Systems (T4SSs) [5]. T4SSs were considered as effector translocators in several α -Proteobacteria, including *Agrobacterium tumefaciens*, *Bartonella* spp. and *Brucella suis* (Ref. [6]).

In the Rickettsiales order, which mainly comprise obligatory endosymbiotic bacteria of *Rickettsia*, *Ehrlichia*, *Anaplasma* and *Wolbachia* genera, orthologs of T4SS genes are related to those of the *A. tumefaciens virB* cluster [6]. In contrast to *A. tumefaciens*, Rickettsiales genomes have lost non-essential genes such as those encoding

pilus-associated proteins (VirB1, -B2, -B5 and VirB7) involved in host cell attachment, presumably due to intracellular lifestyle [7–11]. However, essential T4SS genes for bacteria biology and survival in host cellular environment are expected to be conserved. Hence, the minimum components of typical and functional T4SS are present in Rickettsiales and consist of *virB4*, -B8, -B9, -B10, -B11 and *virD4* homologs [12]. Unlike *A. tumefaciens*, these genes are clustered in two separate loci in Rickettsiales genomes [8,13,14], including *Wolbachia* (Refs. [7,9,10]). One cluster is composed of five tandem genes (*virB8*, -B9, -B10, -B11 and *virD4*) and the other one contains three tandem genes (*virB3*, -B4 and *virB6*). Both operons were found to be transcriptionally active in ovaries of different arthropod hosts [11,15–17] but also in all *Wolbachia*-infected tissues from the isopod *Armadillidium vulgare* [15].

The *Wolbachia* T4SS has been suggested to be a potential pathway to export effectors into host cytoplasm and therefore it may be involved in *Wolbachia*-induced host phenotypes [7]. If so, it is expected that the *Wolbachia* T4SS should be conserved at an evolutionary timescale. In this context, we assessed the genomic structure, location and evolution of T4SS loci in a large data set of 37 *Wolbachia* strains belonging to 5 supergroups [18] and encompassing a wide range of known phenotypes.

Materials and methods

Wolbachia strains. Thirty-seven *Wolbachia* strains were used (Table S1). Sequence data available in databases were used for 10 *Wolbachia* strains (Table S1). In addition, total DNA from 27 arthro-

* Corresponding author. Fax: +33 (0) 5 49 45 40 15.

E-mail addresses: pierre.greve@univ-poitiers.fr, samuel_pichon@hotmail.com (P. Grève).

pod hosts was extracted as previously described [19] (Table S1). Infection status of each individual was checked or tested using a PCR assay based on the *Wolbachia*-specific *wsp* gene, as previously described [20].

DNA amplification and sequencing. T4SS loci structure (i.e. gene content and order) was analyzed in the 27 *Wolbachia* strains shown in bold in Table S1 by a PCR strategy amplifying overlapping fragments. Based on available T4SS sequence data from 7 *Wolbachia* strains [7,9,11,15,16], gene orthologs were aligned and specific oligonucleotide primers were designed in conserved regions of the two operons (Table S2). Primers were also designed to amplify *virB8*–*D4* operon flanking regions including the *dnaA*, *ribA* and *phos* genes at the 5' end of *virB8* and *wspB* at the 3' end of *virD4* (Table S2). Strain-specific primers surrounding the *virB8*–*D4* operon of the wVulC *Wolbachia* strain infecting the isopod *A. vulgare* were designed using sequences from shotgun library clones of the ongoing EuWol sequencing project (contract #QLK3-CT2000-01079) (Table S2). PCR amplification and DNA sequencing were performed as previously described [20].

Sequence analyses. Full-length *virB3*, *virB11* and partial *wsp* sequences from the 37 *Wolbachia* strains were used to construct phylogenetic trees as previously described [20] (Table S1). To assess the quality of phylogenetic signals, we compared tree topologies based on *virB3*, *virB11* and *wsp* taken individually or concatenated in different combinations.

To test for selected sites, analyses of the *vir* sequences from the studied strains were performed using a local codon model of the web-server <http://www.datamonkey.org> [21]. Pairwise estimates of the number of non-synonymous (dN) and synonymous (dS) substitutions per site were compared to the observed mean ω ratio (dN/dS) using the codeml program in the PAML package v3.14 [22]. The random effects likelihood (REL) analysis [21] was used to identify individual sites of codon alignments under selective pressure based on a Bayes factor of 95.

Results and discussion

Structural organization of T4SS operons is conserved among *Wolbachia* strains

To gain insight into the role of T4SS in the biology of the unculturable *Wolbachia*, we assessed diversity, genomic location and evolution of T4SS loci in 37 *Wolbachia* strains. Both *vir* gene operons were successfully amplified in our 27 *Wolbachia* strains. Operon size and *virB* gene order (Fig. 1) were conserved in all *Wolbachia* strains examined [7,9–11,15–17]. *Wolbachia* T4SS loci do not encode all the proteins involved in bacterial conjugation systems from which they may have evolved [6]. Genes homologous to *virB1*, -B2, -B5 and *virB7* of *A. tumefaciens* are not found in *Wolbachia* genomes and the present *virB* loci may encode the minimum T4SS components of a functional translocator system [12]. The same organization is recorded in *Rickettsia*, *Ehrlichia* and *Anaplasma* relatives of *Wolbachia* (Fig. 2). This observation suggests that T4SS operons structure may have been shared by the common ancestor of all the Rickettsiales [7–11,13,14]. This may be a consequence of genome reduction induced by an intracellular lifestyle of these obligatory endosymbiotic bacteria, as observed in the streamlined genomes of many symbionts such as *Wolbachia* [8].

We also found that all *Wolbachia* T4SS components are highly conserved at the nucleotide level. Nucleotide sequences of *virB3* (297 bp) and *virB11* (993 bp) were obtained from 27 *Wolbachia* strains and manually aligned with available sequences from 10 additional *Wolbachia* strains (Table S1). *virB3* and -B11 gene lengths were conserved apart from three *Wolbachia* strains infecting the nematodes *Brugia malayi* (wBm), *Dirofilaria immitis* (wDm)

and the spider *Segestria florentina*, in which *virB11* stop codon was observed 9–12 bp downstream relative to other strains. Average nucleotide divergence over all sequence pairs was 0.052 ± 0.008 for *virB3* and 0.092 ± 0.006 for *virB11*. By contrast, *wsp* sequences exhibited two- to fourfold higher average nucleotide divergence (0.192 ± 0.013 hypervariable regions excluded) than *virB3* and *virB11* among the same *Wolbachia* strains. Phylogenetic reconstructions based on *virB3* or *virB11* sequences revealed overall congruent information with those based on *wsp* sequences whatever the reconstruction method chosen (data not shown). *Wolbachia* strains were clustered in different supergroups as previously described [18]. No correlation between tree topology and *Wolbachia* extended phenotype could be inferred from our results.

When comparing sequence evolutionary rates independently among *Wolbachia* strains, *vir* genes seem to evolve under strong negative selection. Mean $\omega = \text{dN/dS}$ ranged from 0.11 to 0.65 (Table 1). No gene with $\text{dN} > \text{dS}$ was inferred from this dataset, suggesting that any site essential for the function of these proteins would evolve under drastic evolutionary pressures. The highest ω ratio corresponds to *virB6* which is also the most divergent T4SS gene exhibiting size heterogeneity and ~28% divergence between *Wolbachia* strains. On the contrary, the lowest ω ratio was obtained for *virB4*. Additionally, *virB4*, -B11 and *virD4* genes exhibited significant negative selection acting on ~20–50% of their sites (Table 1). These three proteins are known to possess ATPase activity and are involved in effector recruitment in other bacteria [5]. Although secretory systems are mostly conserved among bacteria, substrates of the individual systems may differ from each other depending on the specificity of delivered effector(s) [5]. To date, no functional evidence has shown involvement of T4SS in effector transfer in *Wolbachia*. Nevertheless, T4SS has been shown to be responsible for the secretion of an ankyrin repeat-containing protein (AnkA) in *Anaplasma phagocytophilum*, which is closely related to *Wolbachia* [23]. Together with transcriptional analyses demonstrating coexpression of *vir* genes within each cluster [11,15–17], these results suggest a major role for the T4SS in *Wolbachia* biology.

Flanking genomic regions of the *virB8*–*D4* operon are highly variable

Analyses of flanking regions surrounding the *virB8*–*D4* operon revealed different patterns (Figs. 1 and 2). The first gene identified upstream of the *virB8*–*D4* operon, *ribA*, encoded a bifunctional enzyme (3,4-dihydroxy-2-butanone-4-phosphate synthase and GTP cyclohydrolase II), which catalyzes two essential steps in riboflavin biosynthesis [24]. This gene was found upstream of the *virB8*–*D4* operon in all the *Wolbachia* strains, as well as in 5 *Ehrlichia* genomes, 2 *Anaplasma* genomes, and in the *Neorickettsia sennetsu* genome (Ref. [14]). This feature suggests that a *ribA*–*virD4* locus may have been present in their common ancestor. In *Anaplasma* and *Ehrlichia*, transcriptional analyses revealed that promoters of the *virB8*–*D4* operon are located within the intergenic spacer between *ribA* and *virB8* as well as in the 3' end of *ribA* [24]. Therefore, conservation of at least *ribA* upstream of *virB8* may be important for the integrity and transcriptional activity of the *virB8*–*D4* operon in *Anaplasma*, *Ehrlichia* and *Wolbachia*. Additionally, a Ser/Thr protein phosphatase encoding gene, *phos*, belonging to the calcineurin-phosphoesterase superfamily [25], is also located upstream of *ribA* in all the *Wolbachia* studied, constituting a *phos*–*virD4* locus which may have been present since the *Wolbachia* ancestor.

Diverse genes were localized upstream of these two genes (Fig. 1). *dnaA*, which is essential in *Escherichia coli* for initiation of bidirectional replication at the chromosomal origin of replication [26], was detected in 15 *Wolbachia* strains, including all A- and G-supergroup *Wolbachia* strains tested, as well as the B-supergroup wSn and the D-supergroup wBm strains. This result is congruent with previous observations [7,9,11,17,26]. Interestingly,

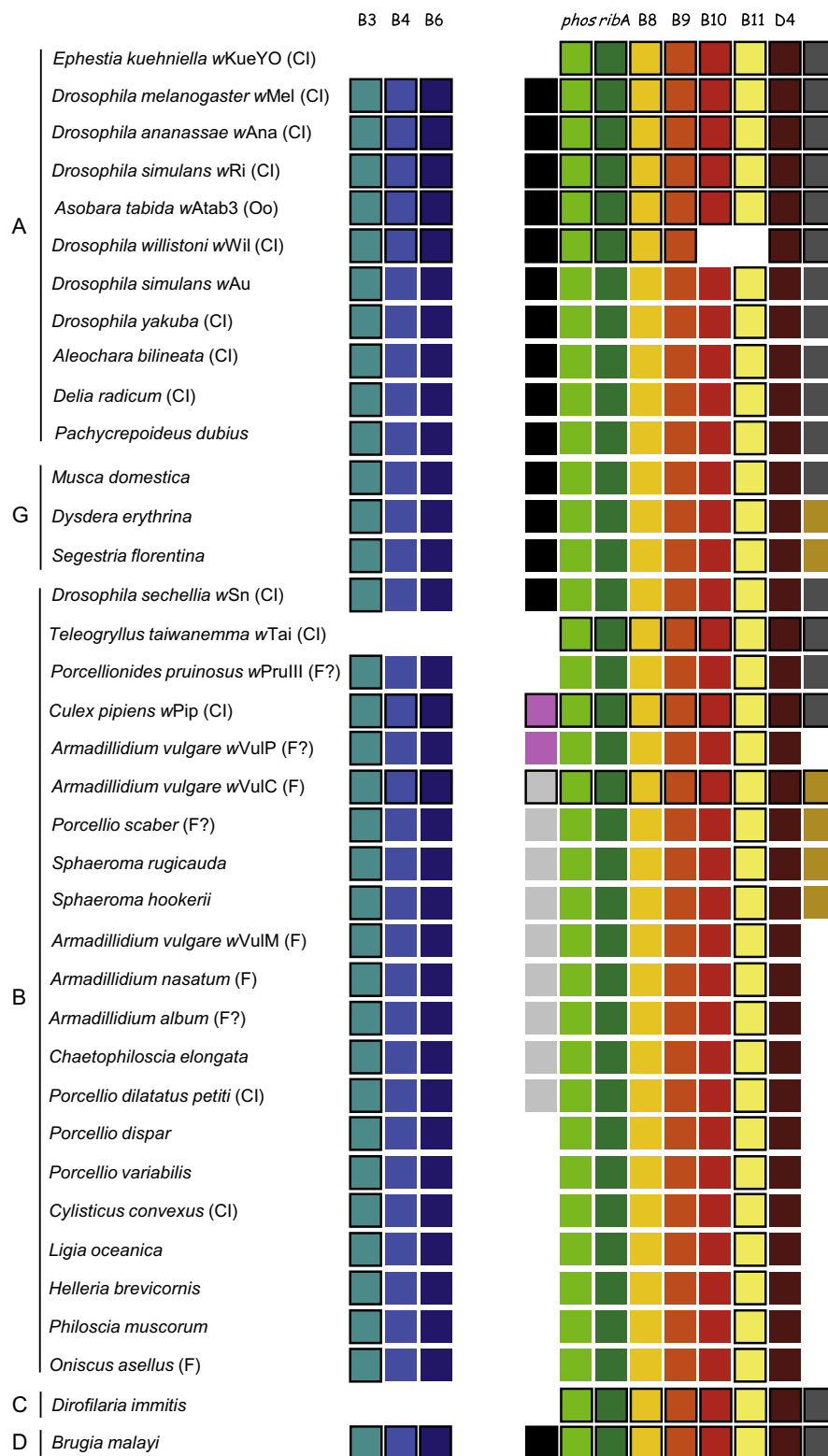


Fig. 1. T4SS operons and flanking region structure among 37 *Wolbachia* strains. Strains are identified by their host species. Bordered frame squares correspond to orthologous genes for which sequences are available in databases or that we have sequenced (see Table S1). No frame indicates that orthologous genes have been amplified by our PCR strategy. Absence of a square at orthologous sites indicates that PCR detection failed or that the gene organization is not available in databases. Colour motifs also identify the nature of genes upstream and downstream of the *virB8*–*D4* operon that differ amongst strains (black: *dnaA*; pink: *wf-2*; light-grey: *recG*; dark-grey: *wspB*; light-brown: insertion sequences). Proven or putative *Wolbachia* extended phenotype is indicated: CI, cytoplasmic incompatibility; F, feminization; F?, putative feminization; Oo, necessary for oogenesis. Supergroups identified in the present phylogenetic framework are identified by the designations A, B, C, D and G. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

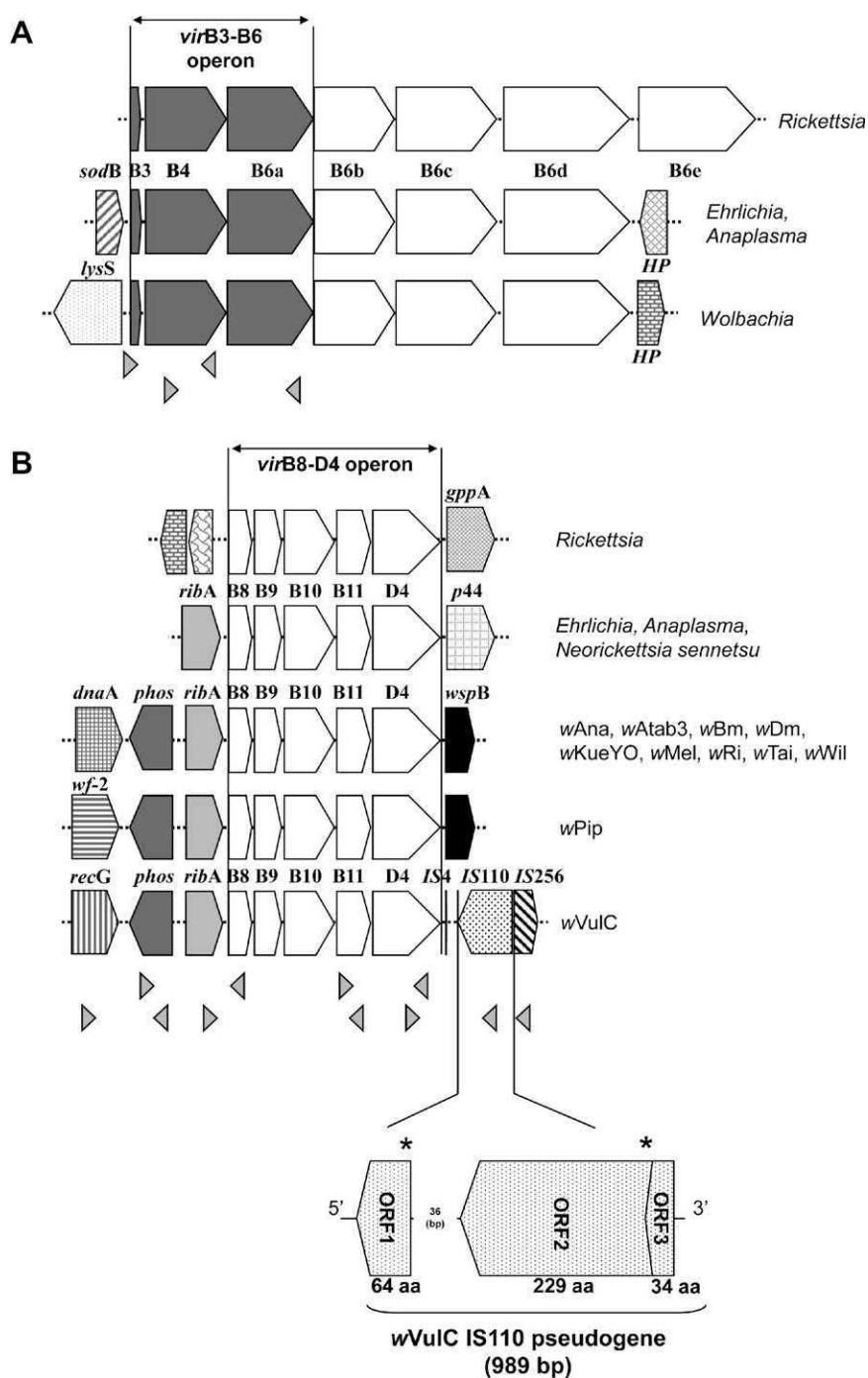


Fig. 2. Schematic physical and genetic map of the two T4SS operons in *Wolbachia* and comparison with other Rickettsiales. Open reading frames (ORFs) are represented as open boxes drawn to scale, with arrowheads indicating the direction of their transcription. Gene names are given at the top (*vir* genes are identified as B3 to D4). Grey orientated triangles refer to relative primer positions (Table S2). Dotted lines represent intergenic spacers. (A) *virB3-B6* operon structure. B6b–e are paralogues of *virB6a*. Flanking genes are indicated by shadings: *sodB*, iron superoxide dismutase gene; *lysS*, lysyl-tRNA synthetase gene; *HP*, Hypothetical Protein encoding gene. (B) *virB8-D4* operon structure. Bacterial strains are indicated at the right; name of host species is available in Table S1. Flanking genes are indicated by shadings: *gppA*, guanosine pentaphosphate phosphohydrolase gene; *p44*, polymorphic major outer membrane protein encoding gene; *phos*, serine/threonine phosphatase protein encoding gene; *ribA*, riboflavin synthase; *wspB*, *Wolbachia* surface protein encoding gene. Bottom part: magnification of the putative degraded IS110 region downstream of the *virD4* gene in *Wolbachia* wVulC strain infecting the isopod *A. vulgare*. Nucleotide insertions causing frameshifts are designated by an asterisk. aa, Amino acids; bp, base pair.

the *virB8-D4* flanking regions obtained from shotgun library clones of wVulC, one the three strains infecting the isopod *A. vulgare* [27], revealed the presence of a helicase homologous to *recG* upstream of *phos*. *recG-phos* PCR products were then obtained in 9 *Wolbachia* strains infecting isopods. Analysis of the *Wolbachia* genome infecting *Culex pipiens* revealed the presence of a *wf-2* homolog upstream of *phos* (Figs. 1 and 2B) which encodes a membrane protein. A

BLASTP search revealed that *wf-2* of wPip (WP1252) shared 66% identity with that of wBm (Wbm0291) but localized elsewhere in the genome whereas no *wf-2* homolog was found in genomes of *Wolbachia* strains infecting *Drosophila*. *wf-2* was also found upstream of the *virB8-D4* operon in wVulP, a recently described *Wolbachia* strain infecting *A. vulgare* [27] (Fig. 1). These results may reflect genetic drift around the T4SS operon. However, we cannot

Table 1

Non-synonymous (dN) and synonymous (dS) substitution rates for individual *vir* genes among various *Wolbachia* strains. Significance (*p*-value) was determined by a likelihood ratio test (LRT) using the codeml program in the PAML package v3.14 [22]. Proportion of negatively selected sites (NSS) indicates the percentage of sites detected under negative selection (REL analysis, Posterior Probability >0.995).

<i>vir</i> genes	Number of codons	Number of strains	Mean dN/dS	LRT <i>p</i> -value	Proportion of NSS (%)
<i>virB3</i>	98	34	0.128	0.6604	0
<i>virB4</i>	801	8	0.114	0.0272	41.1
<i>virB6</i>	919	8	0.648	<0.0001	0
<i>virB8</i>	227	11	0.238	<0.0001	0
<i>virB9</i>	266	11	0.247	<0.0001	0
<i>virB10</i>	509	11	0.432	<0.0001	0.6
<i>virB11</i>	330	35	0.179	<0.0001	20.6
<i>virD4</i>	708	12	0.265	<0.0001	51.4

exclude the hypothesis that operon flanking regions could alternatively influence genomic plasticity in *Wolbachia* that infect isopod species.

Downstream of the *virB8*–*D4* operon, 19 *Wolbachia* strains exhibited a *wspB* homolog, as previously reported [7,9,11,16,17] (Fig. 1). *wspB* is a paralog of the *wsp* gene encoding an outer surface protein and was found in most of the *Wolbachia* strains. An analogous structure has been reported in *A. phagocytophilum* in which an antigenic outer membrane protein is located downstream of *virD4* [24]. *wspB* has previously been hypothesized to be a component of the T4SS operon in *Wolbachia* as *wspB* is cotranscribed as part of the *virB8*–*D4* operon in 3 *Wolbachia* strains belonging to A- and B-supergroups [11,17]. Strain wPruIII [28] was the only *Wolbachia* infecting an isopod in which we could amplify a *virD4*–*wspB* fragment with our primer set suggesting that the *virB8*–*D4* operon may end at *virD4* in *Wolbachia* of other isopod strains. In good agreement with this hypothesis, mfold analysis [29] predicted a rod-shaped secondary structure for the *virB8*–*D4* mRNA potentially producing a transcription termination signal of the transcript 80 bp downstream of *virD4* in wVulC. Although *wspB* was found downstream of *virD4* in wPruIII, wSn and wTai (B-supergroup) as in A-, C- and D-supergroup strains, flanking regions seem to be more variable in B- than in other *Wolbachia* supergroups, suggesting massive genomic rearrangements. This is in agreement with results based on different phylogenetic markers that B- is more diverse than A-supergroup [4]. It is also possible that genomic rearrangements downstream of *virD4* in *Wolbachia* genomes could be the result of horizontal transfers as it has been shown to drive T4SS evolution [6]. This conclusion is further substantiated by the fact that we identified downstream of the *virB8*–*D4* operon of wVulC three insertion sequence (IS) copies belonging to the IS4, IS110 and IS256 families (Fig. 2B). These data represent the first identification of IS elements in the feminizing wVulC *Wolbachia* strain and they are consistent with the notion that *Wolbachia* genomes possess an unusually high proportion of repetitive and mobile DNA, including IS [11,30,31]. The IS4 copy was only represented by its terminal inverted repeat (TIR) (Fig. 2B) [32]. Next, three ORFs were identified as parts of an IS110 transposase gene which is disrupted by two nucleotide insertions causing frame-shifts. As members of the IS110 family do not possess TIRs [32], element boundaries could not be determined precisely. Finally, the IS256 identified in wVulC was homologous to that disrupting *wspB* in wPip [33]. No TIR could be identified in this IS256 element and its 5' region was apparently disrupted by the IS110 copy. Interestingly, IS were also found downstream of *virD4* in *Wolbachia* strains infecting 3 other isopod species and spiders *S. florentina* and *D. erythrina* (Fig. 1). IS may play a considerable role in shaping the evolution of their *Wolbachia* host genomes, for example by promoting gene inactivation and genome plasticity [33,34]. For example, *wspB* is disrupted by an IS256 copy in wPip infecting *C. p. quinquefasciatus* mosquitoes [33]. Therefore, it is plausible that

the observed genomic plasticity around the *virB8*–*D4* operon may be linked to IS presence and/or mobility.

In conclusion, we showed that genetic loci of the *virB* and *virD* genes are conserved but their flanking regions are highly variable among *Wolbachia* strains. These data, together with previous transcriptional analyses [11,15–17], strongly suggest that T4SS is essential in *Wolbachia* biology and survival inside host cells. Genome plasticity in the *virB8*–*D4* operon flanking sequences may favour a potential link between *Wolbachia* genome innovations and *Wolbachia* diversity.

Acknowledgments

We are grateful to Denis Poinot and Hervé Merçot for providing samples, Julien Crine and Sébastien Verne for host species identification and Daniel Guyonnet for technical assistance. This research was funded by an EEC grant (Euwol, QLRT-2000-01079) and the Centre National de la Recherche Scientifique (CNRS), the French Ministère de l'Éducation Nationale, de l'Enseignement Supérieur et de la Recherche and the Agence Nationale de la Recherche (ANR-06-BLAN-0316). SP was supported by a Ph.D. fellowship from Région Poitou-Charentes. RC was supported by a CNRS Young Investigator ATIP award.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbrc.2009.05.118.

References

- [1] D. Bouchon, R. Cordaux, P. Grève, Feminizing *Wolbachia* and the evolution of sex determination in isopods, in: K. Bourtzis, T.A. Miller (Eds.), *Insect Symbiosis*, Taylor & Francis Group, Boca Raton, FL, 2008, pp. 273–294.
- [2] K. Hilgenboecker, P. Hammerstein, P. Schlattmann, A. Telschow, J.H. Werren, How many species are infected with *Wolbachia*? – A statistical analysis of current data, *FEMS Microbiol. Lett.* 281 (2008) 215–220.
- [3] M.J. Taylor, C. Bandi, A. Hoerauf, *Wolbachia* bacterial endosymbionts of filarial nematodes, *Adv. Parasitol.* 60 (2005) 245–284.
- [4] J.H. Werren, *Biology of Wolbachia*, *Annu. Rev. Entomol.* 42 (1997) 587–609.
- [5] P.J. Christie, K. Atmakuri, V. Krishnamoorthy, S. Jakubowski, E. Cascales, Biogenesis, architecture, and function of bacterial type IV secretion systems, *Annu. Rev. Microbiol.* 59 (2005) 451–485.
- [6] A.C. Frank, C.M. Alsmark, M. Thollessen, S.G. Andersson, Functional divergence and horizontal transfer of type IV secretion systems, *Mol. Biol. Evol.* 22 (2005) 1325–1336.
- [7] K. Fenn, M. Blaxter, *Wolbachia* genomes: revealing the biology of parasitism and mutualism, *Trends Parasitol.* 22 (2006) 60–65.
- [8] J.J. Gillespie, N.C. Ammerman, S.M. Dreher-Lesnick, M.S. Rahman, M.J. Worley, J.C. Setubal, B.S. Sobral, A.F. Azad, An anomalous type IV secretion system in *Rickettsia* is evolutionarily conserved, *PLoS ONE* 4 (2009) e4833.
- [9] L. Klasson, T. Walker, M. Sebahia, M.J. Sanders, M.A. Quail, A. Lord, S. Sanders, J. Earl, S.L. O'Neill, N. Thomson, S.P. Sinkins, J. Parkhill, Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group, *Mol. Biol. Evol.* 25 (2008) 1877–1887.
- [10] L. Klasson, J. Westberg, P. Sapountzis, K. Naslund, Y. Lutnaes, A.C. Darby, Z. Veneti, L. Chen, H.R. Braig, R. Garrett, K. Bourtzis, S.G. Andersson, The mosaic

- genome structure of the *Wolbachia* wRi strain infecting *Drosophila simulans*, Proc. Natl. Acad. Sci. USA 106 (2009) 5725–5730.
- [11] M. Wu, L.V. Sun, J. Vamathevan, M. Riegler, R. Deboy, J.C. Brownlie, E.A. McGraw, W. Martin, C. Esser, N. Ahmadinejad, C. Wiegand, R. Madupu, M.J. Beanan, L.M. Brinkac, S.C. Daugherty, A.S. Durkin, J.F. Kolonay, W.C. Nelson, Y. Mohamoud, P. Lee, K. Berry, M.B. Young, T. Utterback, J. Weidman, W.C. Nierman, I.T. Paulsen, K.E. Nelson, H. Tettelin, S.L. O'Neill, J.A. Eisen, Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements, PLoS Biol. 2 (2004) E69.
 - [12] A.B. den Hartigh, H.G. Rolan, M.F. de Jong, R.M. Tsolis, VirB3 to VirB6 and VirB8 to VirB11, but not VirB7, are essential for mediating persistence of *Brucella* in the reticuloendothelial system, J. Bacteriol. 190 (2008) 4427–4436.
 - [13] K.A. Brayton, L.S. Kappmeyer, D.R. Herndon, M.J. Dark, D.L. Tibbals, G.H. Palmer, T.C. McGuire, D.P. Knowles Jr., Complete genome sequencing of *Anaplasma marginale* reveals that the surface is skewed to two superfamilies of outer membrane proteins, Proc. Natl. Acad. Sci. USA 102 (2005) 844–849.
 - [14] J.C. Hotopp, M. Lin, R. Madupu, J. Crabtree, S.V. Angiuoli, J. Eisen, R. Seshadri, Q. Ren, M. Wu, T.R. Utterback, S. Smith, M. Lewis, H. Khouri, C. Zhang, H. Niu, Q. Lin, N. Ohashi, N. Zhi, W. Nelson, L.M. Brinkac, R.J. Dodson, M.J. Rosovitz, J. Sundaram, S.C. Daugherty, T. Davidsen, A.S. Durkin, M. Gwinn, D.H. Haft, J.D. Selengut, S.A. Sullivan, N. Zafar, L. Zhou, F. Benahmed, H. Forberger, R. Halpin, S. Mulligan, J. Robinson, O. White, Y. Rikihisa, H. Tettelin, Comparative genomics of emerging human ehrlichiosis agents, PLoS Genet. 2 (2006) e21.
 - [15] C. Felix, S. Pichon, C. Braquart-Varnier, H. Braig, L. Chen, R.A. Garrett, G. Martin, P. Greve, Characterization and transcriptional analysis of two gene clusters for type IV secretion machinery in *Wolbachia* of *Armadillidium vulgare*, Res. Microbiol. 159 (2008) 481–485.
 - [16] S. Masui, T. Sasaki, H. Ishikawa, Genes for the type IV secretion system in an intracellular symbiont, *Wolbachia*, a causative agent of various sexual alterations in arthropods, J. Bacteriol. 182 (2000) 6529–6531.
 - [17] E. Rances, D. Voronin, V. Tran-Van, P. Mavingui, Genetic and functional characterization of the type IV secretion system in *Wolbachia*, J. Bacteriol. 190 (2008) 5020–5030.
 - [18] N. Lo, C. Paraskevopoulos, K. Bourtzis, S.L. O'Neill, J.H. Werren, S.R. Bordenstein, C. Bandi, Taxonomic status of the intracellular bacterium *Wolbachia pipientis*, Int. J. Syst. Evol. Microbiol. 57 (2007) 654–657.
 - [19] D. Bouchon, T. Rigaud, P. Juchault, Evidence for widespread *Wolbachia* infection in isopod crustaceans: molecular identification and host feminization, Proc. Biol. Sci. 265 (1998) 1081–1090.
 - [20] R. Cordaux, A. Michel-Salzat, D. Bouchon, *Wolbachia* infection in crustaceans: novel hosts and potential routes for horizontal transmission, J. Evolution. Biol. 14 (2001) 237–243.
 - [21] S.L. Kosakovsky Pond, S.D. Frost, Not so different after all: a comparison of methods for detecting amino acid sites under selection, Mol. Biol. Evol. 22 (2005) 1208–1222.
 - [22] Z. Yang, PAML: a program package for phylogenetic analysis by maximum likelihood, Comput. Appl. Biosci. 13 (1997) 555–556.
 - [23] M. Lin, A. den Dulk-Ras, P.J. Hooykaas, Y. Rikihisa, *Anaplasma phagocytophilum* AnkA secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection, Cell. Microbiol. 9 (2007) 2644–2657.
 - [24] N. Ohashi, N. Zhi, Q. Lin, Y. Rikihisa, Characterization and transcriptional analysis of gene clusters for a type IV secretion machinery in human granulocytic and monocytic ehrlichiosis agents, Infect. Immun. 70 (2002) 2128–2138.
 - [25] A.J. Cozzone, Role of protein phosphorylation on serine/threonine and tyrosine in the virulence of bacterial pathogens, J. Mol. Microbiol. Biotechnol. 9 (2005) 198–213.
 - [26] L.V. Sun, A. Babaratsas, C. Savakis, S.L. O'Neill, K. Bourtzis, Gene organization of the dnaA region of *Wolbachia*, J. Bacteriol. 181 (1999) 4708–4710.
 - [27] S. Verne, M. Johnson, D. Bouchon, F. Grandjean, Evidence for recombination between feminizing *Wolbachia* in the isopod genus *Armadillidium*, Gene 397 (2007) 58–66.
 - [28] A. Michel-Salzat, R. Cordaux, D. Bouchon, *Wolbachia* diversity in the Porcellionides pruinosus complex of species (Crustacea: Oniscidea): evidence for host-dependent patterns of infection, Heredity 87 (2001) 428–434.
 - [29] M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction, Nucleic Acids Res. 31 (2003) 3406–3415.
 - [30] R. Cordaux, ISWpi1 from *Wolbachia pipientis* defines a novel group of insertion sequences within the IS5 family, Gene 409 (2008) 20–27.
 - [31] R. Cordaux, S. Pichon, A. Ling, P. Perez, C. Delaunay, F. Vavre, D. Bouchon, P. Greve, Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont, Mol. Biol. Evol. 25 (2008) 1889–1896.
 - [32] M. Chandler, J. Mahillon, Insertion sequences revisited, in: N.L. Craig, M. Gellert, A.M. Lambowitz (Eds.), Mobile DNA II, ASM Press, Washington, DC, 2002, pp. 305–366.
 - [33] Y.O. Sanogo, S.L. Dobson, S.R. Bordenstein, R.J. Novak, Disruption of the *Wolbachia* surface protein gene wspB by a transposable element in mosquitoes of the *Culex pipiens* complex (Diptera, Culicidae), Insect Mol. Biol. 16 (2007) 143–154.
 - [34] J.C. Brownlie, S.L. O'Neill, *Wolbachia* genomes: insights into an intracellular lifestyle, Curr. Biol. 15 (2005) R507–509.

<i>Wolbachia</i>				Accession numbers		
Supergroup	Strain	Host	Taxon	<i>vir</i> B3	<i>vir</i> B11	<i>wsp</i>
A		<i>Aleochara bilineata</i>	Insecta Coleoptera	FJ223232	FJ223258	EU288004
A	wAtab3	<i>Asobara iabida</i>	Insecta Hymenoptera	DQ887630	DQ887626	AF124859
A		<i>Delia radicum</i>	Insecta Diptera	FJ223233	FJ223259	EU288005
A	wAna	<i>Drosophila ananassae</i>	Insecta Diptera	AAGB01000009	AAGB01000038	DQ235410
A	wMel	<i>Drosophila melanogaster</i>	Insecta Diptera	AE017196	AE017196	AE017196
A	wAu	<i>Drosophila simulans</i>	Insecta Diptera	FJ223234	FJ223260	AF290890
A	wRi	<i>Drosophila simulans</i>	Insecta Diptera	EF423641	EF423636	AF020070
A	wWil	<i>Drosophila willistoni</i>	Insecta Diptera	ZP_01315126	NA	AY620218
A		<i>Drosophila yakuba</i>	Insecta Diptera	FJ223235	FJ223261	AJ620679
A	wKueYO	<i>Ephesia kuehniella</i>	Insecta Lepidoptera	NA	AB045235	DQ842473
A		<i>Pachycrepoides dubius</i>	Insecta Hymenoptera	FJ223236	NA	AF124852
B		<i>Armadillidium album</i>	Crustacea Isopoda	FJ223237	FJ223262	DQ778099
B		<i>Armadillidium nasatum</i>	Crustacea Isopoda	FJ223238	FJ223263	DQ778098
B	wVulC	<i>Armadillidium vulgare</i>	Crustacea Isopoda	AAX86700	FJ223284	DQ778095
B	wVulM	<i>Armadillidium vulgare</i>	Crustacea Isopoda	FJ223239	FJ223264	DQ778097
B	wVulP	<i>Armadillidium vulgare</i>	Crustacea Isopoda	FJ223240	FJ223265	DQ778096
B		<i>Chaetophiloscia elongata</i>	Crustacea Isopoda	FJ223241	FJ223266	AJ276611
B	wPip	<i>Culex pipiens</i>	Insecta Diptera	AM999887	AM999887	AM999887
B		<i>Cylisticus convexus</i>	Crustacea Isopoda	FJ223242	FJ223267	AJ276612
B	wSn	<i>Drosophila sechellia</i>	Crustacea Isopoda	FJ223243	FJ223268	AF468036
B		<i>Helleria brevicornis</i>	Crustacea Isopoda	FJ223244	FJ223269	AJ276609
B		<i>Ligia oceanica</i>	Crustacea Isopoda	FJ223245	FJ223270	AJ276610
B		<i>Oniscus asellus</i>	Crustacea Isopoda	FJ223246	FJ223271	AJ276601
B		<i>Philoscia muscorum</i>	Crustacea Isopoda	FJ223247	FJ223272	EU288009
B		<i>Porcellio dilatatus petiti</i>	Crustacea Isopoda	FJ223248	FJ223273	AJ276606
B		<i>Porcellio dispar</i>	Crustacea Isopoda	FJ223249	FJ223274	FJ223282
B		<i>Porcellio scaber</i>	Crustacea Isopoda	FJ223250	FJ223275	AJ276607
B		<i>Porcellio variabilis</i>	Crustacea Isopoda	FJ223251	FJ223276	FJ223283
B	wPrulII	<i>Porcellionides pruinosus</i>	Crustacea Isopoda	FJ223252	FJ223277	AJ300581
B		<i>Sphaeroma hookeri</i>	Crustacea Isopoda	FJ223253	FJ223278	AJ276602
B		<i>Sphaeroma rugicauda</i>	Crustacea Isopoda	FJ223254	FJ223279	AJ276603
B	wTai	<i>Teleogryllus taiwanemma</i>	Insecta Orthoptera	NA	AB045234	AB035514
C	wDm	<i>Dirofilaria immitis</i>	Nematoda Spirurida	NA	AM117567	AJ252062
D	wBm	<i>Brugia malayi</i>	Nematoda Spirurida	AE017321	AE017321	AE017321
G		<i>Dysdera erythrina</i>	Arachnida Araneae	FJ223255	FJ223280	AJ276615
G		<i>Musca domestica</i>	Insecta Diptera	FJ223256	FJ223282	EU288015
G		<i>Segestria florentina</i>	Arachnida Araneae	FJ223257	FJ223281	EU288008

Table S1: List of the 37 *Wolbachia* strains.

List of the 37 *Wolbachia* strains used in this study and accession numbers of *vir*B3, -B11 and *wsp* deposited in GenBank/EMBL databases. *Wolbachia* strains for which new sequence data (bold accession numbers) were obtained in this study are shown in bold. Entries in normal font correspond to data collected from sequence databases. NA indicates that sequences are not available.

ORF	Forward primer		Reverse primer		PCR product size (bp)
	Name	Sequence (5'-3')	Name	Sequence (5'-3')	
Overlapping fragment amplification and Sequencing					
<i>wsp</i>	81F	TGGTCCAATAAGTGATGAAGAAAC	691R	TGGAGTAGCGTTTAATTTT	588
<i>virB3</i>	B3-*43F	GATATAGTATTAGTGTAGAGC	B4-487Rd	TTTGAAACAGCCGATGATGAAG	530
<i>virB3-virB4</i>	B3-*43F	GATATAGTATTAGTGTAGAGC	B4-2297Rd	GATAATCCAGTTTTTGCACCTAA	2340
<i>virB4-virB6</i>	B4-1205F	AAATGCAGATACAGCAAATCG	B6-5367Rd	GTGTCAACAGAGCAGAAGTGG	4163
<i>ribA-virB11</i>	RibA-982Fd	GTACAAAAGTGCTACCACTTATT	B11-3512R	GCTACTTTGCGTGCTATTCC	3686
<i>virB10-virD4</i>	B10-2901F	GCATTGAAAGTATTTGTTAAAC	D4-5820Rd	AAGATGCAATATACGATGAG	2920
<i>virB11</i>	B10-2901F	GCATTGAAAGTATTTGTTAAAC	B11-3512R	GCTACTTTGCGTGCTATTCC	612
	B11-3438Fd	GAAAAATATTATAATTAGTGGTGG	D4-4030Rd	CGCAATATTCTCATAGGAGGTG	593
virB8-D4 flanking region amplification					
<i>phos-virB8</i>	Phos-154F	CAATAGGACACATGTCATTGTC	B8-67R	GATAAGGATATAAATGGAAITCAA	1670
<i>recG-phos</i>	RecG-1980Fd	GCTTGATATATTCGAATATAGGAG	Phos-1208Rd	GCAGAATACATTTCTTATTATCT	666
<i>dnaA-phos</i>	DnaA-1331F	CCGACTCAAAATTTGCTGATG	Phos-1208Rd	GCAGAATACATTTCTTATTATCT	162 *
<i>wf-2-phos</i>	WF2-F	AGTGCATGTTTTTCTCATCTTC	Phos-1208Rd	GCAGAATACATTTCTTATTATCT	648 **
<i>virD4-wspB</i>	D4-5548Fd	CAAGAAATTATTATGTTGCCACG	wsp-80R	CTTTTGCAAGTGAACAGAAGG	588 *
<i>virD4-IS110</i>	D4-5548Fd	CAAGAAATTATTATGTTGCCACG	IS110R	TCTGCATCAGTCTGATCATTCT	1321
<i>virD4-IS256</i>	D4-5548Fd	CAAGAAATTATTATGTTGCCACG	IS256R	CAGGTATATTATGCTATAAARAAG	2261

Table S2: List of primers used.

List of primers used in this study for sequencing or for T4SS PCR amplifications. ORF, open reading frame. Expected PCR product size in base pair (bp) were calculated relative to the wVulC, wMel (*, AE017196) or wPip-Pel (*, AM999887) reference sequences.

E.3 - Cordaux *et al.* 2008.

Article 3: Cordaux R., Pichon S., Ling A., Pérez P., Delaunay C., Vavre F., Bouchon D. and P. Grève. Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. *Molecular Biology and Evolution* 25(9): 1889-1896.

Intense Transpositional Activity of Insertion Sequences in an Ancient Obligate Endosymbiont

Richard Cordaux,* Samuel Pichon,* Alison Ling,* Philippe Pérez,* Carine Delaunay,* Fabrice Vavre,† Didier Bouchon,* and Pierre Grève*

*CNRS UMR 6556 Ecologie, Evolution, Symbiose, Université de Poitiers, Poitiers, France; and †Laboratoire de Biométrie et Biologie Evolutive (UMR 5558), CNRS, Université de Lyon, Lyon; Université Lyon 1, Villeurbanne, France

The streamlined genomes of ancient obligate endosymbionts generally lack transposable elements, such as insertion sequences (IS). Yet, the genome of *Wolbachia*, one of the most abundant bacterial endosymbionts on Earth, is littered with IS. Such a paradox raises the question as to why there are so many ISs in the genome of this ancient endosymbiont. To address this question, we investigated IS transpositional activity in the unculturable *Wolbachia* by tracking the evolutionary dynamics and history of ISWpi1 elements. We show that 1) ISWpi1 is widespread in *Wolbachia*, being present in at least 55% of the 40 sampled strains, 2) ISWpi1 copies exhibit virtually identical nucleotide sequences both within and among *Wolbachia* genomes and possess an intact transposase gene, 3) individual ISWpi1 copies are differentially inserted among *Wolbachia* genomes, and 4) ISWpi1 occurs at variable copy numbers among *Wolbachia* genomes. Collectively, our results provide compelling evidence for intense ISWpi1 transpositional activity and frequent ISWpi1 horizontal transmission among strains during recent *Wolbachia* evolution. Thus, the genomes of ancient obligate endosymbionts can carry high loads of functional and transpositionally active transposable elements. Our results also indicate that *Wolbachia* genomes have experienced multiple and temporally distinct ISWpi1 invasions during their evolutionary history. Such recurrent exposition to new IS invasions may explain, at least partly, the unusually high density of transposable elements found in the genomes of *Wolbachia* endosymbionts.

Introduction

Insertion sequences (ISs) are prokaryotic autonomous transposable elements that encode a transposase gene mediating their transposition (i.e., their ability to move to another locus in a genome) (Chandler and Mahillon 2002). ISs are widespread among prokaryotic genomes (e.g., present in >75% of 262 representative genomes surveyed; Touchon and Rocha 2007), in which they can represent substantial proportions (Chandler and Mahillon 2002; Siguier et al. 2006; Filee et al. 2007). However, when host lifestyle is considered, it is notable that ISs are largely missing from the genomes of obligate endosymbionts, that is, intracellular bacteria that replicate exclusively in the cells of other organisms and typically have no extracellular state (Moran and Plague 2004; Bordenstein and Reznikoff 2005; Touchon and Rocha 2007). This is generally ascribed to the confined and isolated intracellular environment in which these bacteria reside, which reduces opportunities for acquisition of genetic material. This view is supported by the strikingly stable genomes of various obligate endosymbionts of insects such as *Buchnera*, which lack IS and have experienced no genomic rearrangement and gene acquisition for the past 50–70 Myr (Tamas et al. 2002). Yet, comparative genomic analyses of various Rickettsiales, a diverse group of intracellular alpha-Proteobacteria, have demonstrated striking exceptions to this pattern in that these genomes exhibit extensive variability in their mobile element content, including IS (Darby et al. 2007). However, the within-species IS dynamics has not been studied for this group of bacteria, making difficult the analysis of the microevolutionary events responsible for this variability.

Within Rickettsiales, *Wolbachia* bacteria are ancient obligate endosymbionts that have been associated with arthropod and nematode hosts for >100 Myr (Rousset et al. 1992; O'Neill et al. 1997; Bandi et al. 1998; Bourtzis and Miller 2003) and possibly represent one of the most abundant bacterial endosymbionts on Earth (Werren, Windsor, and Guo 1995). These maternally inherited bacteria are often referred to as reproductive parasites because they are able to manipulate the reproduction of their arthropod hosts to increase their own transmission (O'Neill et al. 1997; Bourtzis and Miller 2003; Cordaux et al. 2004). In addition to vertical transmission, *Wolbachia* from arthropods are occasionally transmitted horizontally (Werren, Zhang, and Guo 1995; Vavre et al. 1999; Cordaux et al. 2001). Contrary to expectations, genome sequencing of the *Wolbachia* strain harbored by the fruit fly *Drosophila melanogaster* (wMel) revealed an unusually high proportion of repetitive and mobile DNA, including IS (Moran and Plague 2004; Wu et al. 2004; Bordenstein and Reznikoff 2005). This result is particularly significant given that wMel otherwise exhibits many typical features of a long-term symbiotic lifestyle, such as reduced genome size and A + T nucleotide composition richness (Wernegreen 2002; Wu et al. 2004). Such a paradox raises the question as to why there are so many ISs in the genome of this endosymbiont.

To address this question, we investigated IS transpositional activity in the unculturable *Wolbachia* by tracking the evolutionary dynamics and history of ISWpi1, a group of IS related to the IS5 family, the distribution of which is so far exclusively restricted to *Wolbachia* bacteria (Cordaux 2008). Previous results suggest that ISWpi1 transposase may potentially be functional because 1) the 2 overlapping open reading frames constituting ISWpi1 transposase are intact in many copies (Cordaux 2008) and 2) several ISWpi1 copies are differentially inserted in various *Wolbachia* strains (Duron et al. 2005; Iturbe-Ormaetxe et al. 2005; Riegler et al. 2005). Here, we show that *Wolbachia* endosymbionts have recently experienced, and probably continue to experience, high levels of ISWpi1 transpositional activity

Key words: transposable element, insertion sequence, transpositional activity, horizontal transmission, obligate endosymbiont, *Wolbachia*.

E-mail: richard.cordaux@univ-poitiers.fr.

Mol. Biol. Evol. 25(9):1889–1896. 2008

doi:10.1093/molbev/msn134

Advance Access publication June 17, 2008

© 2008 The Authors

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

Table 1
Distribution of ISWpi1 in 40 *Wolbachia* Strains

Host Species (Strain)	Taxonomic Group	Geographic Origin	<i>Wolbachia</i> Supergroup	ISWpi1 Presence
<i>Aleochara bilineata</i>	Insecta, Coleoptera	Canada	A	Yes
<i>Delia radicum</i>	Insecta, Diptera	Brittany, France	A	Yes
<i>Drosophila ananassae</i>	Insecta, Diptera	Rio de Janeiro, Brazil	A	Yes
<i>Drosophila auraria</i>	Insecta, Diptera	Tokyo, Japan	A	Yes
<i>Drosophila melanogaster</i> (wMel) ^a	Insecta, Diptera	Antibes, France	A	Yes
<i>Drosophila simulans</i> (wAu)	Insecta, Diptera	Yaounde, Cameroon	A	Yes
<i>D. simulans</i> (wRi)	Insecta, Diptera	Antibes, France	A	Yes
<i>Drosophila suzukii</i>	Insecta, Diptera	Tokyo, Japan	A	Yes
<i>Drosophila triauraria</i>	Insecta, Diptera	Tokyo, Japan	A	Yes
<i>Drosophila yakuba</i>	Insecta, Diptera	Ogoue River, Gabon	A	Yes
<i>Zaprionus sepsoides</i>	Insecta, Diptera	Sao Tomé	A	Yes
<i>Asobara tabida</i> (wAtab3)	Insecta, Hymenoptera	Antibes, France	A	Yes
<i>Asobara japonica</i>	Insecta, Hymenoptera	Sapporo, Japan	A	Yes
<i>Leptopilina heterotoma</i> (wLhet1)	Insecta, Hymenoptera	Antibes, France	A	Yes
<i>Pachycrepoides dubius</i>	Insecta, Hymenoptera	France	A	Yes
<i>Amaurobius ferox</i>	Arachnida, Araneae	Poitiers, France	B	Yes
<i>Segestria florentina</i>	Arachnida, Araneae	Chizé, France	B	No
<i>Talitrus saltator</i>	Crustacea, Amphipoda	La Rochelle, France	B	Yes
<i>Lepas anatifera</i>	Crustacea, Cirripedia	La Rochelle, France	B	No
<i>Armadillidium vulgare</i> (wVulC)	Crustacea, Isopoda	Saint Cyr, France	B	Relic only
<i>A. vulgare</i> (wVulM)	Crustacea, Isopoda	Méry sur Cher, France	B	Relic only
<i>Cylisticus convexus</i>	Crustacea, Isopoda	Avanton, France	B	Relic only
<i>Helleria brevicornis</i>	Crustacea, Isopoda	Bastia, France	B	No
<i>Oniscus asellus</i>	Crustacea, Isopoda	Golbey, France	B	Relic only
<i>Philoscia muscorum</i>	Crustacea, Isopoda	Poitiers, France	B	No
<i>Platyarthus hoffmannseggii</i>	Crustacea, Isopoda	Liniers, France	B	Yes
<i>Porcellio dilatatus petiti</i>	Crustacea, Isopoda	Saint Honorat, France	B	No
<i>Porcellionides pruinosus</i> (wPruIII)	Crustacea, Isopoda	Nevers, France	B	Relic only
<i>Sphaeroma hookeri</i>	Crustacea, Isopoda	Graye sur Mer, France	B	No
<i>Sphaeroma rugicauda</i>	Crustacea, Isopoda	Alresford Creek, United Kingdom	B	No
<i>Drosophila sechellia</i> (wSn)	Insecta, Diptera	Seychelles Archipelago	B	Yes
<i>Reticulitermes santoniensis</i>	Insecta, Isoptera	Charente, France	B	Yes
<i>Charanyca trigrammica</i>	Insecta, Lepidoptera	Pinail, France	B	No
<i>Lomaspilis marginata</i>	Insecta, Lepidoptera	Poitiers, France	B	Yes
<i>Maniola jurtina</i>	Insecta, Lepidoptera	Poitiers, France	B	No
<i>Peribatodes rhomboidaria</i>	Insecta, Lepidoptera	Poitiers, France	B	Yes
<i>Spilosoma lubricipeda</i>	Insecta, Lepidoptera	Poitiers, France	B	No
<i>Dysdera crocata</i>	Arachnida, Araneae	Chizé, France	G	No
<i>Dysdera erythrina</i>	Arachnida, Araneae	Saint Benoit, France	G	No
<i>Musca domestica</i>	Insecta, Diptera	Poitiers, France	G	No
Water control			—	No

^a Used as a positive control because ISWpi1 presence is confirmed by in silico analyses (Wu et al. 2004; Cordaux 2008).

within genomes and horizontal transfers among genomes. Our results thus provide compelling evidence that ancient obligate endosymbionts can carry high loads of functional and transpositionally active transposable elements. This may explain, at least partly, why the genomes of *Wolbachia* endosymbionts are littered with IS.

Materials and Methods

Wolbachia Strains

Forty *Wolbachia* strains identified from 23 insect (5 different orders), 13 crustacean (3 different orders), and 4 arachnid individual hosts were used (table 1). Some animals originated from laboratory strains, whereas others were caught in the wild. Total DNA was extracted as previously described (Bouchon et al. 1998). To confirm the presence of *Wolbachia* DNA of suitable quality in the samples, 2–3 loci from *Wolbachia* chromosomal DNA (*wsp*, *16S rRNA*, and *GroE*) were amplified by polymerase chain reaction (PCR), as previously described (Bouchon et al. 1998; Cordaux et al. 2001; Verne et al. 2007). Purified

wsp PCR products were directly sequenced as previously described (Cordaux et al. 2001). Each of the 40 samples was infected by a single *Wolbachia* strain, as indicated by the lack of ambiguity in the electrophoregrams. Sequences generated in this study were deposited in GenBank under accession numbers EU288004–EU288015.

ISWpi1 Detection Assay

To investigate the distribution of ISWpi1 among the 40 *Wolbachia* strains, we designed an intra-ISWpi1 PCR assay, using primers internal to the ISWpi1 consensus sequence. A 681-bp long region internal to ISWpi1 was amplified using specific oligonucleotide primers ISWpi1-F (5'-GATCTAAGCGAAAGGGAATGG) and ISWpi1-R (5'-CAACCCATCTTCTTGGCTGT). PCR amplification was performed using a standard protocol, with an annealing temperature of 60 °C (Cordaux et al. 2006). Resulting PCR products were separated on 1.5% agarose gels, stained with ethidium bromide, and visualized using UV fluorescence. To confirm the results, PCR amplifications were performed

at least twice for each sample, and purified PCR products were directly sequenced as above. ISWpi1 sequences were deposited in GenBank under accession numbers EU288016–EU288038 and EU684314–EU684317. To further confirm the results, *Wolbachia* strains inferred to lack ISWpi1 based on the above PCR assay were subjected to a second PCR assay amplifying 197 bp of ISWpi1 internal sequence, using specific oligonucleotide primers ISWpi1-F1 (5'-CGAAAGGGAATGGTCAAGAA) and ISWpi1-R1 (5'-GCTTCTTCCATTGCCTGAAC) and an annealing temperature of 54 °C.

ISWpi1 Locus Genotyping

To evaluate the timing of ISWpi1 transpositional activity during *Wolbachia* evolution, we assessed the “presence” or “absence” of 24 ISWpi1 copies at orthologous genomic sites in 16 A-supergroup *Wolbachia* strains. Nucleotide sequences of 24 different ISWpi1 copies identified from the *wMel*, *wAna*, *wSim*, and *wWil* *Wolbachia* genomes (Wu et al. 2004; Salzberg, Dunning Hotopp, et al. 2005; Salzberg, Hotopp, et al. 2005; Cordaux 2008) were downloaded from GenBank along with 500 bp of genomic sequence flanking each element on both sides (when available). Specific oligonucleotide primers were designed in the flanking sequences of each ISWpi1 copy, using the program Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). The presence or absence of the 24 ISWpi1 copies was investigated in 12 A-supergroup *Wolbachia* strains from table 1 (strains from *Delia radicum*, *Drosophila suzukii*, and *Pachycrepoideus dubius* were excluded because of insufficient amounts of DNA) using locus-specific PCR assays and confirmed by sequencing of the resulting PCR products, as described above. PCR conditions for each locus, including primer sequences and expected PCR product sizes, are shown in supplementary table S1 (Supplementary Material online). Two loci (*wMel*#4 and *wMel*#9 in supplementary table S1, Supplementary Material online) had to be discarded for further analyses because PCR amplification was successful only in the *wMel* sample. No case of double amplification of expected PCR products for both presence and absence alleles was observed, suggesting homogeneity of the *Wolbachia* population within individual hosts. Sequences were deposited in GenBank under accession numbers EU714507–EU714683. In addition, we performed in silico PCR for 4 A-supergroup *Wolbachia* strains for which genome sequence is available: *wMel*, *wAna*, *wSim*, and *wWil* (Wu et al. 2004; Salzberg, Dunning Hotopp, et al. 2005; Salzberg, Hotopp, et al. 2005; Cordaux 2008).

Southern Blotting

To assess ISWpi1 copy number variation among *Wolbachia* strains, approximately 5 µg of total DNA from various samples were digested with *Hind*III at 37 °C overnight. *Hind*III was chosen because in silico digestion of the *wMel* genome predicted the 13 *wMel* ISWpi1 copies to be located on different digested genomic fragments of relevant sizes. Digested DNA was size fractionated on 1% agarose gels

and Southern blotted to nylon membranes. Probes were prepared as internal portions of ISWpi1 amplified by PCR using the aforementioned primers ISWpi1-F and ISWpi1-R. PCR products were labeled using [α - 32 P]-deoxycytidine triphosphate by the random primer method and hybridized overnight to membranes. The final wash was at 52 °C in 0.1 × standard saline citrate. Hybridized blots were imaged and analyzed using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA).

Sequence Analyses

Sequences were aligned using ClustalW as implemented in the software Bioedit version 7.0 (Hall 1999), followed by manual adjustments. MEGA version 4 (Tamura et al. 2007) was used to calculate nucleotide sequence divergence and build A- and B-supergroup *Wolbachia* phylogenetic trees using distance-based (Neighbor-Joining [NJ] and unweighted pair group method with arithmetic mean) and character-based (maximum parsimony [MP]) methods. The different methods yielded largely congruent phylogenies, and we show in the paper the trees that displayed the highest confidence levels in branching patterns, as detailed below.

Due to low genetic differentiation among strains (Werren, Zhang, and Guo 1995), distance-based methods yielded A-supergroup *Wolbachia* trees with mostly short branches and low confidence in the branching patterns (i.e., low bootstrap scores). By contrast, MP yielded only 5 equally most parsimonious trees (tree length: 875 steps) that differed only in the branching patterns of the 4 highly closely related *Wolbachia* strains from *Drosophila simulans* (*wRi* and *wSim* variants) and *Drosophila ananassae* (2 *wAna* variants). Overall, this suggested high support for the branching patterns of the MP inference. Based on prior knowledge on strain origins, the most parsimonious tree linking the 2 *D. simulans* *Wolbachia* variants, on the one hand, and the 2 *D. ananassae* *Wolbachia* variants, on the other hand, was considered as the most biologically relevant tree. The high consistency index (0.875) provided further support for the MP tree shown in figure 1.

Regarding B-supergroup *Wolbachia* strains, distance-based and MP trees essentially differed on the position of the *Reticulitermes santonensis* *Wolbachia* strain. However, the MP analysis yielded as many as 190 equally parsimonious trees (tree length: 440 steps), with a consistency index of only 0.745. By contrast, the 2 distance-based methods (which agreed on the branching pattern of the *R. santonensis* *Wolbachia* strain) were characterized by high bootstrap scores. Hence, 10 out of 14 nodes displayed bootstrap values >95% in the NJ tree, thus providing strong support for the NJ topology.

Results and Discussion

Widespread Distribution of ISWpi1 among *Wolbachia* Strains

The taxonomic distribution of ISWpi1 is apparently restricted to *Wolbachia* bacteria, as found earlier by Blast

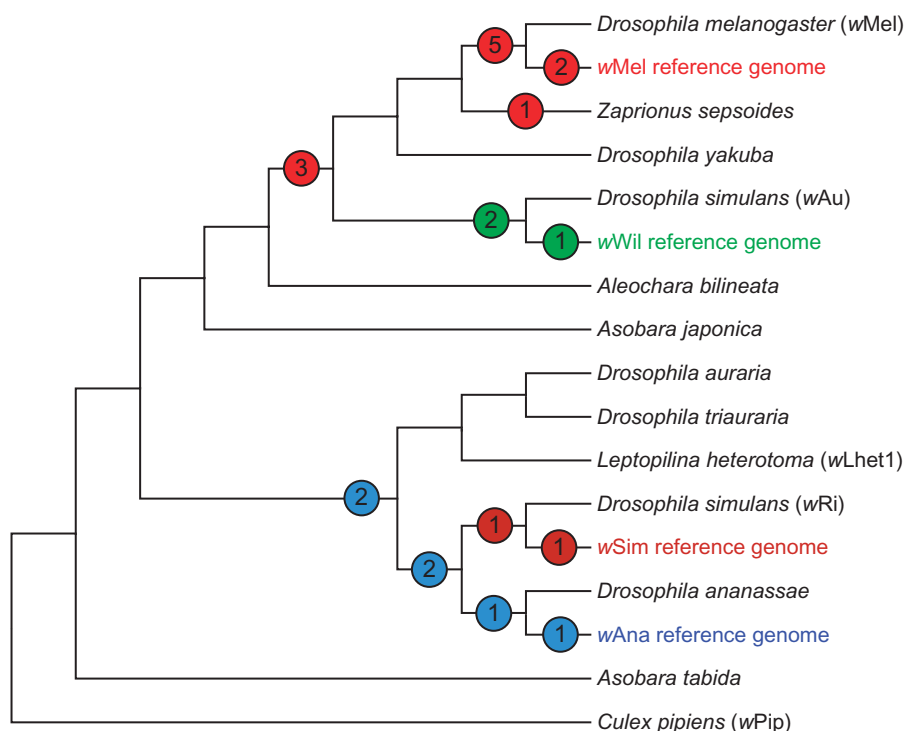


FIG. 1.—Distribution of 22 ISWpi1 copies isolated from the wMel (red), wWil (green), wSim (brown), and wAna (blue) reference genome sequences. Colored circles highlight the numbers of inferred absence/presence transitions of ISWpi1 copies in different branches of the phylogenetic tree of 16 A-supergroup *Wolbachia* strains. The tree was reconstructed by MP (based on 9,782 bp of sequence flanking the 22 ISWpi1 loci and the *wsp* gene) and rooted using the B-supergroup *Wolbachia* strain from *Culex pipiens* (Sanger Institute, http://www.sanger.ac.uk/Projects/W_pipientis/). Branch length is arbitrary. *Wolbachia* strains are identified by the host species from which they were isolated.

searches against the entire GenBank database and all prokaryote genomes listed in the “microbial genomes” section of GenBank (Cordaux 2008). In this study, we confirm ISWpi1 restricted distribution even though new sequence data have been added to GenBank since previous searches. Using a PCR-based ISWpi1 detection assay, we screened a panel of 40 diverse *Wolbachia* strains belonging to the A, B, and G *Wolbachia* supergroups (table 1). A PCR fragment of the expected size (681 bp) was obtained in 22 out of the 40 tested *Wolbachia* strains. Absence of the expected 681-bp long PCR fragment in some strains is unlikely to be caused by systematic PCR failure due to primer mismatches because average ISWpi1 sequence divergence across 22 *Wolbachia* strains is only 0.22% (see below), indicating that 2 full-length ISWpi1 sequences are expected to differ by only 2 substitutions on average. Moreover, *Wolbachia* strains inferred to lack ISWpi1 based on the first PCR assay were subjected to a second ISWpi1 detection assay, which confirmed the initial results.

ISWpi1 was not uniformly distributed among *Wolbachia* supergroups ($P < 10^{-5}$, Fisher’s exact test). It was present in all 15 A-supergroup *Wolbachia* strains screened (table 1), in agreement with its presence in all A-supergroup *Wolbachia* strains for which genomic information is available (Cordaux 2008). By contrast, ISWpi1 was found in only 32% (7/22) of B-supergroup and none (0/3) of the G-supergroup *Wolbachia* strains tested (table 1). Overall, these results indicate that ISWpi1 is widespread among

Wolbachia endosymbionts because it is present in the genomes of 55% of all *Wolbachia* strains tested.

Extreme ISWpi1 Sequence Homogeneity within and among *Wolbachia* Strains

To investigate ISWpi1 nucleotide variation, we compared the ISWpi1 sequences obtained from the 22 *Wolbachia* strains identified above as possessing ISWpi1. PCR products were directly sequenced to simultaneously sequence homologous regions from multiple ISWpi1 copies possibly occurring within a single *Wolbachia* genome. Lack of ambiguous sites in the sequence trace files suggested extremely low to no nucleotide divergence among the different ISWpi1 copies occurring within each *Wolbachia* genome. This result is consistent with the virtual lack of nucleotide variation previously recorded among the ISWpi1 copies present within various sequenced *Wolbachia* genomes (Cordaux 2008). However, some private substitutions might have remained undetected with this sequencing strategy. Thus, the 22 ISWpi1 sequences can actually be viewed as consensus sequences of all individual ISWpi1 copies inserted within each of the analyzed *Wolbachia* genomes, making them useful for comparisons among strains. Overall, nucleotide divergence of the 22 ISWpi1 sequences from the various A- and B-supergroup *Wolbachia* strains was only 0.22%. This virtual lack of ISWpi1 sequence variation among *Wolbachia* genomes is in sharp

contrast with the $\sim 3.7\%$ average nucleotide divergence among *Wolbachia* supergroups A and B recorded for 8 highly conserved housekeeping genes (range: 2.2–4.9%) and even much lower than the divergence ($\sim 0.7\%$) observed for the extremely conserved 16S rRNA gene (Paraskevopoulos et al. 2006).

Purifying selection acting on ISWpi1 transposase genes is unlikely to account for this extreme ISWpi1 sequence homogeneity because it would imply that selection for transposition is stronger than selection constraining housekeeping genes essential for *Wolbachia* metabolism. Maintaining such intense levels of purifying selection on ISWpi1 sequences seems further implausible given the elevated evolutionary rates and relative inefficiency of natural selection in endosymbiotic bacteria with reduced effective population sizes, such as *Wolbachia* (Wu et al. 2004). Gene conversion (i.e., the nonindependent evolution of repetitive DNA sequences) could explain the homogeneity of ISWpi1 copies within *Wolbachia* genomes, but it cannot account for the homogeneity of ISWpi1 among *Wolbachia* genomes. Therefore, the most likely explanation for the presence of highly homogeneous ISWpi1 sequences in *Wolbachia* strains as divergent as those belonging to different supergroups is that ISWpi1 has been transpositionally active and laterally acquired by diverse *Wolbachia* strains during very recent evolutionary times (Wagner 2006).

Recent and Intense ISWpi1 Transpositional Activity

To evaluate the timing of ISWpi1 transpositional activity during *Wolbachia* evolution, we analyzed the phylogenetic distribution of 22 individual ISWpi1 copies in 16 A-supergroup *Wolbachia* strains. This approach allowed us to pinpoint transitions between absence and presence of individual ISWpi1 copies, which are signatures of transpositional activity, during A-supergroup *Wolbachia* evolutionary history. Some transitions might have been overlooked because ISWpi1 status could not be determined for some loci in some taxa. We emphasize, however, that it would not affect our conclusions based on a conservative set of unambiguously determined transitions.

We were able to map presence/absence transitions to the *Wolbachia* phylogeny for 11 *wMel* ISWpi1 copies. Our results indicated that none of the ISWpi1 copies is shared by all A-supergroup *Wolbachia* strains (fig. 1 and supplementary table S2 [Supplementary Material online]). Instead, all copies showed very narrow strain distributions. Hence, 7 ISWpi1 copies identified from the *wMel* genome sequence were apparently specific to *wMel*. The other copies were shared with just a few closely related *Wolbachia* strains that exhibit $>99\%$ nucleotide sequence identity with *wMel* based on the hypervariable *Wolbachia*-specific *wsp* gene (Charlat et al. 2003, 2004). In fact, 2 copies have presumably been transpositionally active so recently in *wMel* that they are polymorphic for insertion presence or absence among different geographic *wMel* variants. Specifically, ISWpi1 copies at loci *wMel*#6 and *wMel*#12 isolated from the sequenced *wMel* genome (Wu et al. 2004; Cordaux 2008) were absent from our *wMel* sample originating from France. Although *wMel*#6 (WD0516-0517 in the original

wMel genome annotation) has previously been shown to be polymorphic (Riegler et al. 2005), we identified here *wMel*#12 as a novel polymorphic marker that may prove useful for studies of *Wolbachia* diversity and evolution in *D. melanogaster*.

To test if the very recent ISWpi1 transpositional activity suggested by the transition patterns of ISWpi1 copies isolated from *wMel* can be generalized to other ISWpi1 copies, we extended our analysis to 11 additional ISWpi1 copies isolated from the partial genome sequences of *wAna* (6 loci), *wSim* (2 loci), and *wWil* (3 loci). Again, all ISWpi1 copies exhibited very narrow strain distributions (fig. 1 and supplementary table S2 [Supplementary Material online]). Even the 2 most widely distributed ISWpi1 copies isolated from *wAna* were found in closely related *Wolbachia* strains that are identical based on the hypervariable *Wolbachia*-specific *wsp* gene (Miller and Riegler 2006).

Next, we assessed ISWpi1 copy number variation among A-supergroup *Wolbachia* strains by Southern blotting. Results indicated that the number of distinct bands (i.e., putative distinct copies) for A-supergroup *Wolbachia* strains varies from 7 to 13 copies (fig. 2). These figures are in line with the copy numbers estimated from genome sequence data for other A-supergroup *Wolbachia* strains (Cordaux 2008). Interestingly, there are approximately twice as many ISWpi1 copies in *wMel* compared with the closely related *wAu*, whereas there are similar copy numbers between *wMel* and the distantly related *wAna* (fig. 2 and Cordaux 2008).

Overall, extensive heterogeneity in ISWpi1 copy numbers among *Wolbachia* strains, along with very narrow distribution of 22 individual ISWpi1 copies identified from 4 different host genomes and extreme ISWpi1 sequence homogeneity, provides compelling evidence for intense ISWpi1 transpositional activity during recent *Wolbachia* evolution. We emphasize that the extensive polymorphism observed, both in terms of overall copy numbers and patterns of presence or absence of individual copies among *Wolbachia* strains, may result from a combination of insertion events and secondary excisions. In any event, this testifies to the intense transpositional activity that *Wolbachia* endosymbionts have recently experienced and may continue to currently experience. ISWpi1 recent transposition in various *Wolbachia* strains is further supported by the fact that the 2 overlapping open reading frames constituting ISWpi1 transposase are intact in all sequenced portions, suggesting that there are sources of functional transposases in all A- and B-supergroup *Wolbachia* genomes containing ISWpi1 we analyzed. If so, our results provide strong evidence that the genomes of ancient obligate endosymbionts can carry high loads of functional and active transposable elements.

Frequent ISWpi1 Horizontal Transfers during Recent *Wolbachia* Evolution

The ubiquitous presence of ISWpi1 in the *Wolbachia* A supergroup, coupled with reduced levels of sharing of individual copies among *Wolbachia* strains, suggests that some *Wolbachia* strains may have independently acquired

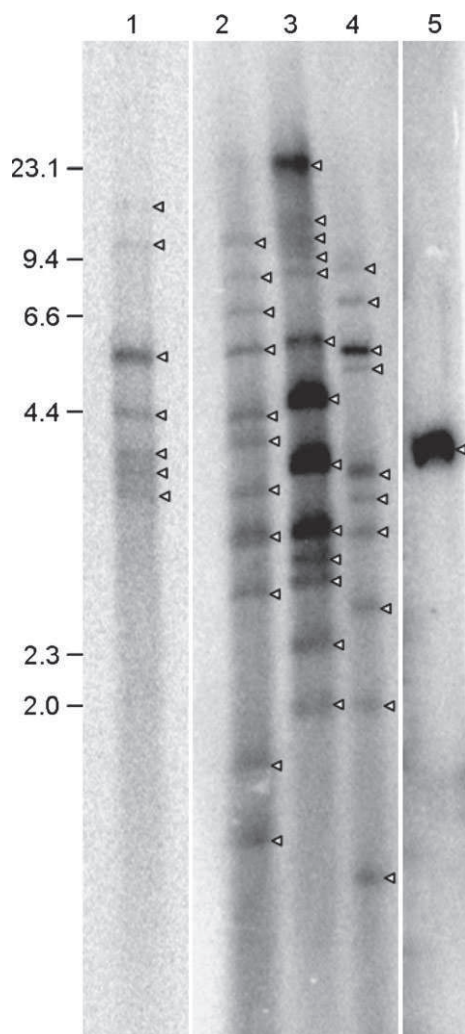


FIG. 2.—Southern blotting of *Hind*III-digested DNA. Lanes: *Wolbachia* strains from *Drosophila simulans* wAu (1), *Drosophila melanogaster* (2), *Asobara japonica* (3), *Drosophila ananassae* (4), and *Armaddillidium vulgare* wVulC (5). Figures on the left indicate fragment sizes (kb). White triangles highlight the positions of the fragments. A single band was detected in wVulC, which presumably corresponds to the ISWpi1 relic identified by PCR and sequencing (see main text). Other *Wolbachia* strains exhibit from 7 to 13 distinct bands.

ISWpi1 via lateral transfers. To estimate the number of independent ISWpi1 acquisitions in the *Wolbachia* B supergroup, we analyzed ISWpi1 distribution according to bacterial strain phylogenetic relationships (fig. 3). At this level of resolution, the presence of ISWpi1 in B-supergroup *Wolbachia* strains putatively results from at least 4 independent acquisitions (fig. 3). This may be an underestimate because 1) a higher phylogenetic resolution in the *Lomaspilis marginata*/*Talitrus saltator*/*Amaurobius ferox* group of closely related *Wolbachia* strains might result in the inference of additional independent ISWpi1 acquisitions, 2) a larger screening of *Wolbachia* strains for ISWpi1 presence might uncover additional acquisition events, and 3) one cannot formally exclude that ISWpi1 has been transferred several times to individual *Wolbachia* strains. In any event, these results suggest that horizontal transmission may be a major determinant of the current ISWpi1 distri-

bution among *Wolbachia* strains. Only limited cases of horizontal transfers of mobile DNA in obligate endosymbiotic bacteria have been reported previously, including a plasmid in *Buchnera* (Van Ham et al. 2000), a bacteriophage in *Wolbachia* (Bordenstein and Wernegreen 2004; Gavotte et al. 2007), and a putative conjugative element in *Rickettsia* (Blanc et al. 2007). ISWpi1 from *Wolbachia* is the first transposable element unambiguously shown to horizontally transfer in obligate endosymbiotic bacteria.

Frequent ISWpi1 transfers among different *Wolbachia* strains could be facilitated by the occasional co-occurrence of divergent *Wolbachia* strains within the same host cells (Vavre et al. 1999; Bordenstein and Wernegreen 2004), as well as the presence of bacteriophage WO in many *Wolbachia* genomes (Bordenstein and Wernegreen 2004; Wu et al. 2004; Braquart-Varnier et al. 2005; Gavotte et al. 2007) that could serve as a shuttle for efficiently transferring genetic material among strains. Consistently, the wBm *Wolbachia* genome from the nematode *Brugia malayi* that lacks bacteriophage WO (Foster et al. 2005) also lacks recent ISWpi1 copies (Cordaux 2008). On the other hand, bacteriophage WO distribution seems restricted to *Wolbachia*, and it has never been found in other bacteria to date (Bordenstein and Wernegreen 2004; Gavotte et al. 2007), which could also contribute to explain why ISWpi1 taxonomic distribution also appears restricted to *Wolbachia* (Cordaux 2008). If so, *Wolbachia* bacteria may constitute a highly dynamic system for genetic exchanges among strains (Bordenstein and Wernegreen 2004), whereas at the same time being less prone to exchanges with other bacterial species, perhaps as a result of the specialization of vectors involved in IS horizontal transfer.

Why So Many ISs in *Wolbachia* Genomes?

While investigating ISWpi1 distribution by PCR in 40 *Wolbachia* strains, we amplified ISWpi1 “relics” from the genomes of 5 B-supergroup *Wolbachia* strains: a 312-bp fragment in 4 *Wolbachia* strains (including wVulC) and a 550-bp fragment in 1 *Wolbachia* strain (table 1). DNA sequencing revealed that the shorter and longer fragments exhibited 12.3% and 10.4% nucleotide divergence with ISWpi1, respectively, and 20.1% with each other. In addition, both fragments were severely truncated compared with ISWpi1 due to multiple internal deletions and both were lacking any significant coding capacity. Southern blotting of wVulC *Wolbachia* strain DNA against an ISWpi1 probe identified a single band (fig. 2), suggesting that the ISWpi1 relic identified above is the only ISWpi1 copy currently inserted in the wVulC genome. Other highly divergent copies have also been reported from the B-supergroup wPip and D-supergroup wBm *Wolbachia* strains (Duron et al. 2005; Cordaux 2008), suggesting an ancient presence of ISWpi1 in *Wolbachia* genomes. Because our PCR-based strategy was designed to preferentially detect ISWpi1 copies closely related to the ISWpi1 consensus sequence (i.e., presumably recent copies), it is possible that some ISWpi1 relics have remained undetected in our screening. Thus, the distribution of ISWpi1 relics among *Wolbachia* genomes may be underestimated.

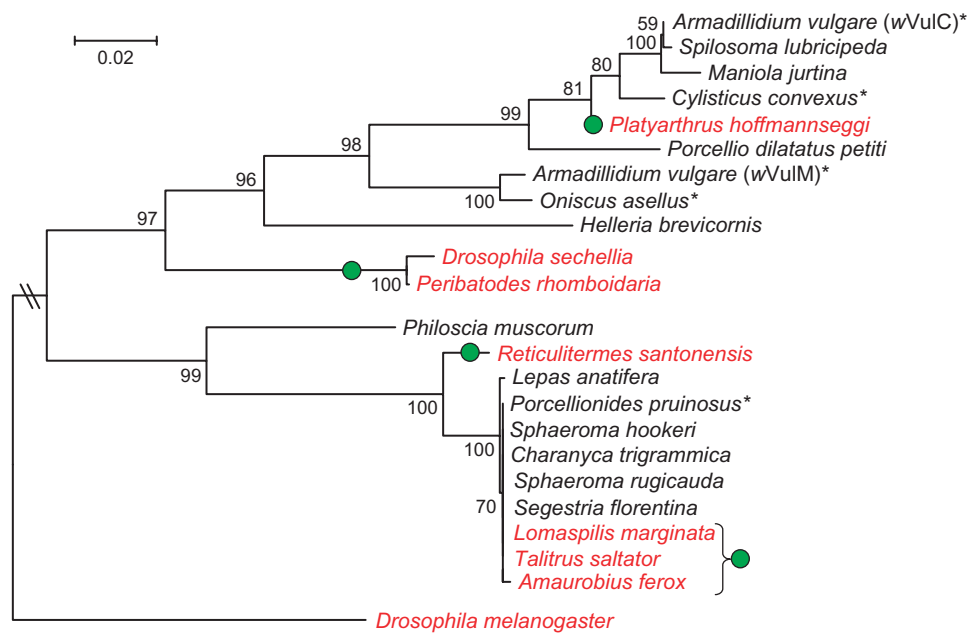


FIG. 3.—NJ tree of 22 B-supergroup *Wolbachia* strains for which ISWpi1 presence or absence is known, based on *wsp* sequences and the Kimura 2-parameter substitution model. Bootstrap values (based on 1,000 replicates) are shown on branches (%). The tree was rooted using the A-supergroup *Wolbachia* strain from *Drosophila melanogaster*. The broken line indicates that the branch is not drawn to scale. *Wolbachia* strains are identified by the host species from which they were isolated. Strains possessing (red) or lacking (black) ISWpi1 copies are shown. Stars indicate strains possessing ISWpi1 relics. Green circles indicate the putative independent ISWpi1 acquisitions.

Overall, our results are consistent with a scenario in which ISs recurrently invade and then go extinct in bacterial genomes (Wagner 2006) so that ancient relics and recent ISWpi1 copies represent temporally distinct ISWpi1 invasions of *Wolbachia* genomes. It has been proposed that IS could be maintained in *Wolbachia* genomes because they confer a selective advantage to their bacterial hosts (Brownlie and O'Neill 2005; Foster et al. 2005). Alternatively, it is possible that ISs are maintained in *Wolbachia* simply as a consequence of the inefficiency of host genomes to eliminate them (Wu et al. 2004). The rationale underlying this hypothesis is that symbiotic bacteria tend to have small effective population sizes, thus rendering selection against deleterious mutations and transposable element insertions less efficient (Wu et al. 2004). The evolutionary history and dynamics of ISWpi1 suggest yet another explanation: *Wolbachia* genomes are recurrently exposed to new IS invasions (Bordenstein and Wernegreen 2004).

Conclusion

It is generally considered that IS proliferation characterizes lineages that have recently evolved toward an obligate endosymbiotic lifestyle (Moran and Plague 2004; Plague et al. 2008). By contrast, ancient obligate endosymbionts typically lack IS because of degradation of old insertions and absence of exposure to new transposition events (Moran and Plague 2004). Unexpectedly, our results show that at least a subset of all IS copies of the obligate endosymbiont *Wolbachia* are not remnants of ancient IS proliferation following the shift to endosymbiotic lifestyle at an earlier stage of *Wolbachia* evolution. Instead, *Wolbachia* experi-

ence recurrent invasions by new IS, which may explain, at least partly, the unusually high density of transposable elements found in the genomes of these endosymbionts.

Supplementary Material

Supplementary tables S1 and S2 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We are grateful to Vincent Doublet, Hervé Merçot, Jean Louis Picaud, Denis Poinot, and Sébastien Verne for providing samples. We thank Daniel Guyonnet for technical assistance and Christine Braquart-Varnier and Mathieu Sicard for comments on an earlier version of the manuscript. This research was funded by the Centre National de la Recherche Scientifique (CNRS), the French Ministère de l'Éducation Nationale, de l'Enseignement Supérieur et de la Recherche, and the Agence Nationale de la Recherche (ANR-06-BLAN-0316). R.C. was supported by a CNRS Young Investigator ATIP award. S.P. was supported by a PhD fellowship from Région Poitou-Charentes.

Literature Cited

- Bandi C, Anderson TJ, Genchi C, Blaxter ML. 1998. Phylogeny of *Wolbachia* in filarial nematodes. *Proc Biol Sci.* 265:2407–2413.
- Blanc G, Ogata H, Robert C, Audic S, Claverie JM, Raoult D. 2007. Lateral gene transfer between obligate intracellular

- bacteria: evidence from the *Rickettsia massiliae* genome. *Genome Res.* 17:1657–1664.
- Bordenstein SR, Reznikoff WS. 2005. Mobile DNA in obligate intracellular bacteria. *Nat Rev Microbiol.* 3:688–699.
- Bordenstein SR, Wernegreen JJ. 2004. Bacteriophage flux in endosymbionts (Wolbachia): infection frequency, lateral transfer, and recombination rates. *Mol Biol Evol.* 21:1981–1991.
- Bouchon D, Rigaud T, Juchault P. 1998. Evidence for widespread Wolbachia infection in isopod crustaceans: molecular identification and host feminization. *Proc Biol Sci.* 265:1081–1090.
- Bourtzis K, Miller TA. 2003. *Insect symbiosis*. Boca Raton (FL): CRC Press.
- Braquart-Varnier C, Greve P, Felix C, Martin G. 2005. Bacteriophage WO in Wolbachia infecting terrestrial isopods. *Biochem Biophys Res Commun.* 337:580–585.
- Brownlie JC, O'Neill SL. 2005. Wolbachia genomes: insights into an intracellular lifestyle. *Curr Biol.* 15:R507–R509.
- Chandler M, Mahillon J. 2002. Insertion sequences revisited. In: CR, Craig NL, Gellert M, Lambowitz AM, editors. *Mobile DNA II*. Washington (DC): ASM Press. p. 305–366.
- Charlat S, Ballard JW, Mercot H. 2004. What maintains noncytoplasmic incompatibility inducing Wolbachia in their hosts: a case study from a natural *Drosophila yakuba* population. *J Evol Biol.* 17:322–330.
- Charlat S, Le Chat L, Mercot H. 2003. Characterization of noncytoplasmic incompatibility inducing Wolbachia in two continental African populations of *Drosophila simulans*. *Heredity.* 90:49–55.
- Cordaux R. 2008. ISWp1 from Wolbachia pipientis defines a novel group of insertion sequences within the IS5 family. *Gene.* 409:20–27.
- Cordaux R, Lee J, Dinoso L, Batzer MA. 2006. Recently integrated Alu retrotransposons are essentially neutral residents of the human genome. *Gene.* 373:138–144.
- Cordaux R, Michel-Salzat A, Bouchon D. 2001. Wolbachia infection in crustaceans: novel hosts and potential routes for horizontal transmission. *J Evol Biol.* 14:237–243.
- Cordaux R, Michel-Salzat A, Frelon-Raimond M, Rigaud T, Bouchon D. 2004. Evidence for a new feminizing Wolbachia strain in the isopod *Armadillidium vulgare*: evolutionary implications. *Heredity.* 93:78–84.
- Darby AC, Cho NH, Fuxelius HH, Westberg J, Andersson SG. 2007. Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *Trends Genet.* 23:511–520.
- Duron O, Lagnel J, Raymond M, Bourtzis K, Fort P, Weill M. 2005. Transposable element polymorphism of Wolbachia in the mosquito *Culex pipiens*: evidence of genetic diversity, superinfection and recombination. *Mol Ecol.* 14:1561–1573.
- Filee J, Siguier P, Chandler M. 2007. Insertion sequence diversity in archaea. *Microbiol Mol Biol Rev.* 71:121–157.
- Foster J, Ganatra M, Kamal I, et al. (23 co-authors). 2005. The Wolbachia genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol.* 3:e121.
- Gavotte L, Henri H, Stouthamer R, Charif D, Charlat S, Bouletreau M, Vavre F. 2007. A survey of the bacteriophage WO in the endosymbiotic bacteria Wolbachia. *Mol Biol Evol.* 24:427–435.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 41:95–98.
- Iturbe-Ormaetxe I, Burke GR, Riegler M, O'Neill SL. 2005. Distribution, expression, and motif variability of ankyrin domain genes in Wolbachia pipientis. *J Bacteriol.* 187:5136–5145.
- Miller WJ, Riegler M. 2006. Evolutionary dynamics of wAu-like Wolbachia variants in neotropical *Drosophila* spp. *Appl Environ Microbiol.* 72:826–835.
- Moran NA, Plague GR. 2004. Genomic changes following host restriction in bacteria. *Curr Opin Genet Dev.* 14:627–633.
- O'Neill SL, Hoffmann AA, Werren JH. 1997. *Influential passengers: inherited microorganisms and arthropod reproduction*. New York: Oxford University Press.
- Paraskevopoulos C, Bordenstein SR, Wernegreen JJ, Werren JH, Bourtzis K. 2006. Toward a Wolbachia multilocus sequence typing system: discrimination of Wolbachia strains present in *Drosophila* species. *Curr Microbiol.* 53:388–395.
- Plague GR, Dunbar HE, Tran PL, Moran NA. 2008. Extensive proliferation of transposable elements in heritable bacterial symbionts. *J Bacteriol.* 190:777–779.
- Riegler M, Sidhu M, Miller WJ, O'Neill SL. 2005. Evidence for a global Wolbachia replacement in *Drosophila melanogaster*. *Curr Biol.* 15:1428–1433.
- Rousset F, Bouchon D, Pintureau B, Juchault P, Solignac M. 1992. Wolbachia endosymbionts responsible for various alterations of sexuality in arthropods. *Proc Biol Sci.* 250:91–98.
- Salzberg SL, Dunning Hotopp JC, Delcher AL, Pop M, Smith DR, Eisen MB, Nelson WC. 2005. Correction: serendipitous discovery of Wolbachia genomes in multiple *Drosophila* species. *Genome Biol.* 6:402.
- Salzberg SL, Hotopp JC, Delcher AL, Pop M, Smith DR, Eisen MB, Nelson WC. 2005. Serendipitous discovery of Wolbachia genomes in multiple *Drosophila* species. *Genome Biol.* 6:R23.
- Siguier P, Filee J, Chandler M. 2006. Insertion sequences in prokaryotic genomes. *Curr Opin Microbiol.* 9:526–531.
- Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, Wernegreen JJ, Sandstrom JP, Moran NA, Andersson SG. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. *Science.* 296:2376–2379.
- Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol.* 24:1596–1599.
- Touchon M, Rocha EP. 2007. Causes of insertion sequences abundance in prokaryotic genomes. *Mol Biol Evol.* 24:969–981.
- Van Ham RC, Gonzalez-Candela F, Silva FJ, Sabater B, Moya A, Latorre A. 2000. Postsymbiotic plasmid acquisition and evolution of the repA1-replicon in *Buchnera aphidicola*. *Proc Natl Acad Sci USA.* 97:10855–10860.
- Vavre F, Fleury F, Lepetit D, Fouillet P, Bouletreau M. 1999. Phylogenetic evidence for horizontal transmission of Wolbachia in host-parasitoid associations. *Mol Biol Evol.* 16:1711–1723.
- Verne S, Johnson M, Bouchon D, Grandjean F. 2007. Evidence for recombination between feminizing Wolbachia in the isopod genus *Armadillidium*. *Gene.* 397:58–66.
- Wagner A. 2006. Periodic extinctions of transposable elements in bacterial lineages: evidence from intragenomic variation in multiple genomes. *Mol Biol Evol.* 23:723–733.
- Wernegreen JJ. 2002. Genome evolution in bacterial endosymbionts of insects. *Nat Rev Genet.* 3:850–861.
- Werren JH, Windsor D, Guo L. 1995. Distribution of Wolbachia among neotropical arthropods. *Proc Biol Sci.* 262:197–204.
- Werren JH, Zhang W, Guo LR. 1995. Evolution and phylogeny of Wolbachia: reproductive parasites of arthropods. *Proc Biol Sci.* 261:55–63.
- Wu M, Sun LV, Vamathevan J, et al. (30 co-authors). 2004. Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.* 2:E69.

Jennifer Wernegreen, Associate Editor

Accepted June 10, 2008

Locus name	PCR primers (5' > 3')		PCR product size (bp)	
	Forward	Reverse	Present IS	Absent IS
wMel#1	GCATAAAAACGCGCAAAAGT	AAAGACTGCCCACTTCCAGA	1,397	478
wMel#2	GTGACGGTGGATTTTCTTGG	AAAGCAGAGCGCAATCTAT	1,325	406
wMel#3	GAAGACAATCTTCCGCTTGC	TCCAGCATTTTGTGAGCTTG	1,428	509
wMel#4	GAATCGTGAGAGGGAACCTGC	CAAAAGTCGCCTCAGACCTC	1,112	193
wMel#5	TAATCTATCCGGTGGGCAAA	TATCACGGCGGTATGACGTA	1,400	481
wMel#6	ACGTAGGACGGCTAGCTTCA	CTTGAAACCAGCGATTGGT	1,415	496
wMel#7	TTTTTGCCAGCTCAGTCTT	ACGCTTGCTTCCAAGAAAAA	1,202	283
wMel#8	AAGCTACGCAGTGCTGTCAA	AGCACCCTTAGGCCTTGAA	1,030	111
wMel#9	CAAAATCGTCGCAAAACCTT	AAGCGTAAACCTGCCAAAAA	1,231	312
wMel#10	TTCATGCCATCTTCACCGTA	TAGCGCTTCCATTGCTTTT	1,445	526
wMel#11	AGGCGTTTTATTGCAAATGG	TGCAATGTTTGTGGCGTAGT	1,266	347
wMel#12	GTTGCATCGCTCAAGTGAAG	TGCAATTTTCTCGTTTGCTC	1,431	512
wMel#13	TTGCTGCACTTCTCTGCAAT	CTGAAATGATGCCGGTTCTT	1,482	563
wAna#1	TGCCCTGCATCTGACAATAG	ACCCTTTGCCGTTATTGTTG	1,484	565
wAna#4	TCCATGGGAATAGAAAGTGGA	ACTGCTGCACAGTGACCAAG	1,274	355
wAna#7	GCTCTTGGTTGGTGTGGATT	CTTGTGTAGCGGGAATTGGT	1,260	341
wAna#10	CGAAGTCGCAAAAGCAATG	TGGTTAATCTTGTAGAAGTTTGC	1,333	414
wAna#12	ACAAAGGGAAGCTCATGGTG	CGCAACCAATGATCAACATC	1,397	478
wAna#14	GACAAGCCAAAAAGCGAAAG	ATGGTTGCAGCCTTAAGTGG	1,471	555
wSim#2	GCAAGCAGATGGCAGTAAGG	TGGCATAACCAGAGCCAAAG	1,516	597
wSim#10	TTCTCCACGTTTTCTGAAGG	CTTCCAGGAGCTGGTTCTTTAC	1,512	593
wWil#3	CCAGGTAACCTCAAGCAAATAAG	AAGTGGCCTAAGTCTGATG	1,379	460
wWil#5	TACAACGCCCAATGTAAAG	GCCACTACTCAAGACATACGAAC	1,510	591
wWil#6	CAGGATGCAATTGAGAATGC	GCCCTTCAGTTCTGTTCAGC	1,561	642

Table S1: PCR primers and expected product sizes for 24 ISWpi1 loci present in the *wMel*, *wAna*, *wSim* and *wWil* *Wolbachia* genomes.

Source	Locus	wRi	Dana	Lhet	Atab	Ajap	wAu	Dyak	Dmel	Abil	Daur	Dtri	Zsep	wMel	wAna	wSim	wWil	wPip
wMel	1	0	0	0	0	?	0	0	1	?	?	?	0	1	?	?	?	?
wMel	2	0	0	0	0	?	1	1	1	0	0	0	?	1	0	0	1	?
wMel	3	?	?	?	?	0	0	0	1	0	?	?	0	1	0	0	0	?
wMel	5	0	0	0	0	0	0	0	1	?	?	?	0	1	0	?	0	?
wMel	6	0	0	0	?	?	0	0	0	0	0	0	0	1	?	0	0	0
wMel	7	?	?	?	?	0	1	1	1	?	?	?	0	1	?	?	?	?
wMel	8	?	?	0	0	?	0	0	1	?	0	0	0	1	?	?	?	0
wMel	10	?	?	?	0	?	?	1	1	0	?	?	1	1	?	?	?	?
wMel	11	0	0	0	0	0	0	?	1	0	0	0	0	1	0	0	?	0
wMel	12	0	0	0	0	0	0	0	0	0	0	0	0	1	?	?	0	?
wMel	13	0	0	0	0	0	1	1	1	0	0	0	1	1	0	0	?	0
wAna	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	?	0	?
wAna	4	1	1	0	?	0	0	0	0	0	0	0	0	0	1	?	0	?
wAna	7	1	1	1	0	?	?	?	?	?	1	1	?	?	1	?	?	?
wAna	10	1	1	?	0	0	0	?	0	?	1	1	0	0	1	?	0	?
wAna	12	0	0	0	?	?	?	?	0	?	0	0	0	0	1	0	?	?
wAna	14	0	1	0	0	?	0	?	0	0	?	?	?	0	1	?	?	?
wSim	2	0	0	0	0	0	0	?	0	0	0	?	?	0	?	1	0	?
wSim	10	1	0	0	0	?	0	?	0	0	0	0	?	0	0	1	0	0
wWil	3	0	0	0	?	?	1	?	0	0	0	0	0	0	0	0	1	?
wWil	5	?	?	?	?	0	1	?	0	?	?	?	?	0	?	?	1	?
wWil	6	0	0	0	?	?	0	?	0	0	0	0	0	0	0	?	1	?

Codes

0 IS absent 1 IS present ? not determined

confirmed by sequencing not confirmed by sequencing in silico PCR

Table S2: Presence or absence of ISWpi1 elements at 22 orthologous loci in 17 *Wolbachia* strains.

CHAPITRE III :

**LES PROTEINES A DOMAINES
ANKYRINES, DES EFFECTEURS
IMPLIQUES DANS LES
INTERACTIONS ENTRE
WOLBACHIA ET SES HOTES ?**

III. Les protéines à domaines ankyrines, des effecteurs impliqués dans les interactions entre *Wolbachia* et ses hôtes ?

A. Objectifs.

Wolbachia est une bactérie difficilement cultivable et l'élaboration de transformants s'est révélée à ce jour infructueuse (Tanaka *et al.* 2009). Il est donc impossible d'envisager l'étude de mutants pour identifier les facteurs impliqués dans les interactions entre la bactérie et son hôte. Une approche alternative, basée sur l'analyse comparative des génomes, permet cependant d'identifier des facteurs de virulence potentiels comme cela a été le cas chez *L. pneumophila* ou *Rickettsia rickettsii* (Cazalet *et al.* 2008, Ellison *et al.* 2008, Burstein *et al.* 2009). Une telle approche est envisageable pour *Wolbachia* puisque 4 génomes ont été séquencés entièrement, et que de nombreux génomes partiels sont également disponibles pour plusieurs autres souches, y compris la souche *wVulC* (Wu *et al.* 2004, Foster *et al.* 2005, Salzberg *et al.* 2005, Klasson *et al.* 2008, Klasson *et al.* 2009). Les premières analyses des génomes de *Wolbachia* ont révélé la présence d'un nombre étonnamment important de gènes codant des protéines à motif ankyrine (ANK) (Wu *et al.* 2004, Klasson *et al.* 2008). Etant donnée la nature de ces domaines, connus pour être impliqués dans des interactions protéine-protéine chez les eucaryotes, la présence d'un grand nombre de ces protéines ANK chez *Wolbachia* fait de celles-ci des candidats naturels impliqués dans la communication avec l'hôte (McGraw et O'Neill 2004, Fenn et Blaxter 2006).

Cette hypothèse est de plus confortée par les démonstrations récentes chez *L. pneumophila* et *A. phagocytophilum* de l'implication de protéines ankyrines bactériennes dans la perturbation de la physiologie de leur hôte (Lin *et al.* 2007, Pan *et al.* 2008). Chez *L. pneumophila*, l'analyse *in silico* des génomes de plusieurs souches a permis d'identifier 142 effecteurs dont 11 gènes codant des protéines ANK qui sont effectivement sécrétées par la bactérie (Al-Khodori *et al.* 2008, Altman et Segal 2008, de Felipe *et al.* 2008, Habyarimana *et al.* 2008, Pan *et al.* 2008, Burstein *et al.* 2009). Parmi ces protéines, AnkX empêche, après interaction avec les microtubules, le transport de la vacuole contenant les bactéries et sa fusion avec les endosomes de macrophages humains (Pan *et al.* 2008).

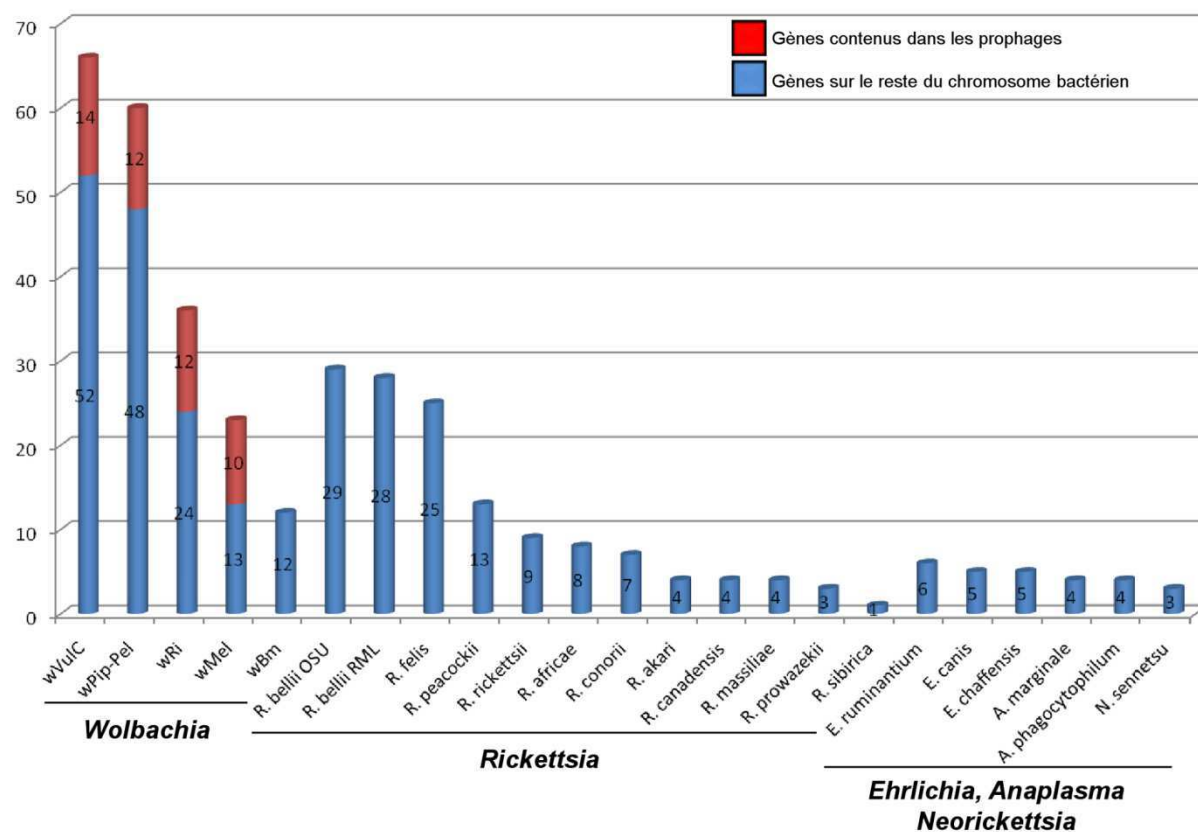


Figure 13: Nombre de gènes ankyrines dans les génomes de Rickettsiales.

Diagramme représentant le nombre de gènes codant des protéines à domaines ankyrines dans les génomes de différentes Rickettsiales (5 souches de *Wolbachia*, puis 2 d'*Anaplasma*, 3 d'*Ehrlichia*, 1 de *Neorickettsia* et 12 de *Rickettsia*). Les chiffres sur chaque barre représentent le nombre de gènes contenus dans les prophages (en rouge) ou dispersés sur le reste du chromosome bactérien (en bleu).

Chez *A. phagocytophilum*, proche parent de *Wolbachia*, AnkA est une protéine sécrétée dans le cytoplasme de la cellule hôte puis dirigée dans le noyau où elle interagit avec de nombreuses protéines dont une ATPase, une tyrosine phosphatase et une protéine de type NADH deshydrogénase (Caturegli *et al.* 2000, Park *et al.* 2004, Lin *et al.* 2007). Il a donc été montré que des protéines bactériennes à domaines ankyrines peuvent interagir avec celles de l'hôte pour assurer l'infection et/ou le maintien du symbiote. Enfin, chez *Wolbachia*, l'équipe de Sinkins a proposé l'implication de deux protéines ANK de *wPip* (*pk1* et *pk2*) dans le phénomène d'IC induit par la bactérie chez *C. pipiens* (Sinkins *et al.* 2005). Ils ont montré que ces gènes étaient polymorphes et permettaient de distinguer des souches génétiquement très proches induisant des effets d'IC différents (Sinkins *et al.* 2005). Ensuite, des variations de l'expression de *pk2*, selon le sexe et les populations d'origine, ont pu être corrélées avec le phénotype d'IC induit par ces mêmes souches (Sinkins *et al.* 2005).

Ainsi, de par leur nature, toutes les protéines ANK peuvent être considérées comme des effecteurs candidats impliqués dans le dialogue entre la bactérie et son hôte, y compris dans les phénomènes d'altération de la reproduction. Nous avons entrepris d'identifier et de caractériser l'ensemble des gènes codant des protéines à motifs ankyrines chez *wVulC* et de les comparer aux autres gènes *ank* connus chez *Wolbachia*. Ce travail a ensuite été en partie associé à une étude plus large comprenant l'étude approfondie des motifs ANK de protéines d'autres souches bactériennes, de virus, ainsi que d'organismes eucaryotes. Enfin, nous avons étudié plus précisément la distribution et l'expression des gènes homologues de *pk1* et de *pk2* chez les souches infectant les isopodes terrestres afin de tester la corrélation entre l'expression de ces gènes et les phénotypes induits par *Wolbachia*.

B. Etude comparative des gènes ankyrines de *wVulC*.

Nous avons mis à profit les avancées du séquençage du génome de *wVulC* afin d'identifier sur les différents contigs assemblés, les gènes codant des protéines à motifs ankyrines. Deux approches complémentaires ont été utilisées, la première par utilisation du serveur BLASTP (Altschul *et al.* 1990) sur les séquences protéiques de *wVulC* en utilisant comme requête l'ensemble des protéines ankyrines de *Wolbachia* disponibles dans GenBank. Chaque séquence de *wVulC* identifiée a secondairement fait l'objet d'une recherche de motifs ANK avec le serveur SMART (e-value < 1) (Letunic *et al.* 2009).

	wVulC	wPip-Pel	wRi	wMel	wBm
Taille du génome (bp)	~1 750 000	1 482 455	1 445 873	1 267 782	1 080 084
Nombre de gènes <i>ank</i>	66	60	36	23	12
Nombre de pseudogènes	0	3	8	0	7
Proportion de pseudogènes	0%	5,0%	22,2%	0%	58,3%
Nombre de gènes associés au phage	14	12	12	10	0
Proportion de gènes de phage*	21,2%	20,0%	33,3%	43,5%	0%
Nombre de gènes uniques	63	54	29	23	12
Nombre de gènes spécifique à chaque souche*	18	24	6	1	2
Taille totale de séquences <i>ank</i> (bp)	73 539	106 994	55 110	30 651	10 918
min	303	450	120	459	246
max	3 510	8 247	7 425	2 901	3 327
Proportion dans le génome	4,2%	7,2%	3,8%	2,4%	1,0%
Contenu en G+C (%) dans le génome	?	34,2%	35,0%	35,2%	34,2%
Contenu en G+C (%) dans les gènes <i>ank</i>	34,6%	34,2%	34,4%	35,2%	33,8%
Contenu en G+C (%) dans les gènes <i>ank</i> de phage	35,1%	35,4%	34,7%	35,2%	0,0%
Nombre de motifs ANK	317	276	183	114	23
min	1	1	1	1	1
max	14	18	34	11	9
moyen (par protéine)	4,8	4,8	6,3	5,0	4,6
Nombre de domaines TM	34	42	15	11	4
Nombre de protéines ANK contenant des TM	19	25	9	6	2
Proportion dans le génome*	28,8%	43,9%	32,1%	26,1%	40,0%
Nombre de domaines autres que TM et ANK	3	3	5	0	0
Nombre de protéines contenant un peptide signal	4	3	2	2	1

* pseudogènes exclus

Figure 14: Comparaison des caractéristiques des gènes *ank* de *wVulC* avec celles des gènes *ank* des 4 souches de *Wolbachia*.

Des analyses par BLASTP ont été réitérées en utilisant les séquences présentant effectivement des domaines ANK, et ainsi de suite jusqu'à épuisement de résultats positifs. Cette approche a permis de révéler 82 fragments d'ADN codant des motifs ANK dont certains étaient des homologues de gènes codant des protéines à motifs ankyrines connues. La seconde approche était basée sur la recherche avec le serveur SMART de motifs ANK directement sur les contigs de *wVulC* traduits dans les 6 cadres de lecture. Ceci nous a permis de confirmer les résultats issus de la première analyse et d'identifier au total 122 séquences. Les 40 séquences codant des motifs ANK identifiées uniquement par cette seconde approche ont ensuite été analysées par BLASTP comme décrit précédemment ce qui a permis de mettre en évidence de nouveaux homologues d'ankyrines de *Wolbachia*.

A la date de l'écriture de ce manuscrit, l'état d'avancement de l'assemblage et de l'annotation du génome de *wVulC* nous a permis d'affiner ces résultats en éliminant les séquences chevauchantes et appartenant à une seule et même ORF. Nous avons ainsi mis en évidence la présence de 66 gènes codant des protéines contenant au moins un motif ANK chez *wVulC*. Ceci représente un nombre minimum dans la mesure où d'autres gènes ank peuvent être localisés dans les régions non encore séquencées ou non annotées du génome. Puisque trois de ces gènes possèdent chacun un paralogue de séquence identique, *wVulC* totalise en fait 63 séquences *ank* uniques. Ce nombre de gènes est déjà supérieur à celui répertorié dans les autres génomes séquencés de *Wolbachia*, le maximum de gènes *ank* ayant été observé dans le génome de *wPip-Pel* (60 gènes dont 56 uniques) (Figure 13). Enfin, cela représente le nombre de gènes *ank* le plus important décrit à ce jour chez un procaryote. Ce résultat renforce l'idée selon laquelle un nombre élevé de gènes *ank* est une caractéristique des bactéries du genre *Wolbachia*. On retrouve cette caractéristique ponctuellement dans d'autres génomes du genre *Rickettsia* (*R. bellii* RML (28), *R. felis* (24)) (Figure 13). En nombre de paires de bases, les séquences ankyrines représentent entre 1 et 7,2% des génomes de *Wolbachia*, et environ 4,2% du génome de *wVulC* (Figure 14).

La comparaison des gènes *ank* de *wVulC* avec ceux des quatre souches de *Wolbachia* dont les génomes ont été publiés (Wu *et al.* 2004, Foster *et al.* 2005, Klasson *et al.* 2008, Klasson *et al.* 2009), a révélé que seuls deux gènes présentent des orthologues dans les cinq souches (Figure 15). Si l'on exclut de l'analyse la souche *wBm* qui ne possède pas de prophage (dont la séquence est riche en gènes *ank*), douze gènes de *wVulC* présentent des orthologues dans les trois souches de *Wolbachia* induisant une IC (*wPip-Pel*, *wRi* et *wMel*) (Figure 15). C'est avec *wPip-Pel*, la souche la plus proche phylogénétiquement de *wVulC* et appartenant elle aussi au supergroupe B, que *wVulC* partage le plus grand nombre de gène *ank* orthologues (27 gènes) (Figure 15). Parmi ces gènes, quinze sont présents uniquement chez ces deux souches du supergroupe B alors qu'ils sont absents chez celles du supergroupe A (Figure 15).

	w Mel	w Ri	w Pip_Pel	w VulC	w Bm
Gènes communs à toutes les souches	WD0438	WRi_003070	WPa_0680	wVulC_ANK54	Wbm0447
	WD0441	WRi_003040	WPa_0677	wVulC_ANK87	Wbm0582
	WD0766	WRi_006090	WPa_0506	wVulC_ANK79	Wbm0296
Gènes communs à toutes les souches sauf w Bm	WD0035	WRi_000290	WPa_1105	wVulC_ANK88	
	WD0147	WRi_001060	WPa_0088		
	WD0191	WRi_001790	WPa_0767	wVulC_ANK71	
	WD0385	WRi_p04040	WPa_1220	wVulC_ANK22-56	
	WD0498	WRi_002480	WPa_0688	wVulC_ANK51	
	WD0754	WRi_007240	WPa_0149	wVulC_ANK86	
	WD0292		WPa_0296		
	WD0566	WRi_006900	WPa_0354	wVulC_ANK59	
			WPa_0452		
			WPa_1337		
	WD0596	WRi_005620	WPa_0256	wVulC_ANK25	
		WRi_010280	WPa_0315	wVulC_ANK32	
		WRi_p07220	WPa_1308	wVulC_ANK46a	
		WRi_p07650		wVulC_ANK46b	
				wVulC_ANK60a	
				wVulC_ANK60b	
	WD0633	WRi_005390	WPa_1293		
		WRi_006690			
		WRi_010050			
	WD0636	WRi_005440	WPa_0299	wVulC_ANK40	
		WRi_010100	WPa_0413	wVulC_ANK48	
			WPa_0340		
	WD0637	WRi_005450	WPa_0300	wVulC_ANK41	
		WRi_010110	WPa_0414	wVulC_ANK47a	
			WPa_0339	wVulC_ANK47b	
Gènes de w Mel ou w Ri n'ayant pas d'homologues chez w VulC ou w Pip-Pel	WD0285				
	WD0286				
	WD0291				
	WD0294				
	WD1213				
	WD0073	WRi_000620			
	WD0434	WRi_003100			
	WD0032-34	WRi_000280			
Gènes communs à w VulC/w Ri	WD0550	WRi_p03640			
		WRi_002880		wVulC_ANK66	
Gènes communs à w VulC/w Pip		WRi_006850		wVulC_ANK85	
			WPa_0383	wVulC_ANK43	
			WPa_0390	wVulC_ANK34	
			WPa_0410	wVulC_ANK16	
			WPa_0648	wVulC_ANK53	
			WPa_0652	wVulC_ANK44-59	
			WPa_0680	wVulC_ANK54	
			WPa_0689	wVulC_ANK36	
			WPa_0730	wVulC_ANK28-73	
			WPa_0731	wVulC_ANK74-75	
			WPa_0733	wVulC_ANK76	
	WD1199		WPa_0763	wVulC_ANK4-72	Wbm0742
			WPa_0768	wVulC_ANK01	
			WPa_0787	wVulC_ANK49-50	
			WPa_1275	wVulC_ANK33	
			WPa_1279	wVulC_ANK80	
Gènes <i>a priori</i> spécifiques à chaque souche	WD0514	WRi_006740	WPa_0112	wVulC_ANK08	Wbm0287
		WRi_006750	WPa_0179	wVulC_ANK17	Wbm0394
		WRi_006810	WPa_0195	wVulC_ANK20	
		WRi_006860	WPa_0217	wVulC_ANK27	
		WRi_006870	WPa_0218	wVulC_ANK29	
		WRi_012460	WPa_0292	wVulC_ANK31	
		WRi_p06670	WPa_0293	wVulC_ANK37	
		WRi_p06840	WPa_0346	wVulC_ANK42	
			WPa_0347	wVulC_ANK62	
			WPa_0388	wVulC_ANK63	
			WPa_0407	wVulC_ANK64	
			WPa_0462	wVulC_ANK67	
			WPa_0606	wVulC_ANK69	
			WPa_0638	wVulC_ANK78	
			WPa_0688	wVulC_ANK81	
			WPa_0729	wVulC_ANK82	
			WPa_0774	wVulC_ANK83	
			WPa_0776	wVulC_ANK89	
			WPa_0862		
			WPa_1014		
			WPa_1147		
			WPa_1351		
			WPa_1378		
			WPa_1379		

 Figure 15: Liste des gènes ankyrines homologues entre les 5 souches de *Wolbachia* étudiées.

Les gènes sont identifiés chez wVulC par rapport à un numéro de séquences donné arbitrairement mais sans rapport avec leur localisation sur les contigs. En rouge sont indiqués les gènes de prophages, en vert les gènes sans domaine ankyrine mais homologues d'ankyrines chez une autre souche.

Enfin, 18 gènes de *wVulC* n'ont aucun homologue chez d'autres souches de *Wolbachia*, ce qui en fait à ce jour des gènes spécifiques (Figures 14 et 15). Des points restent à approfondir notamment concernant l'absence chez *wVulC* de certains gènes conservés dans l'ensemble des autres génomes de *Wolbachia* (Figure 15). Il est possible que ces gènes soient réellement présents dans le génome de *wVulC* mais pas encore identifiés du fait de l'annotation en cours.

En termes de répartition génomique, 14 des 66 gènes identifiés chez *wVulC* sont présents dans des régions du génome de *Wolbachia* contenant des prophages WO (Figures 14 et 15). Deux gènes en particulier, *pk1* et *pk2*, sont présents en plusieurs copies dans les prophages des génomes de *wVulC*, *wPip-Pel* et *wRi* (Figure 15). Cette observation suppose que des événements de duplication ou de recombinaison intra et/ou inter-phage ont eu lieu au cours de l'histoire évolutive de ces gènes, événements nécessairement postérieurs à leur divergence avec *wMel* chez qui ces gènes sont présents en un seul exemplaire (Wu *et al.* 2004). Le nombre de copies de ces gènes semble être lié au nombre de copies du phage WO-B présent au sein d'un génome [*cf.* Article 5 : Pichon *et al.*, p. 90]. Néanmoins, le nombre total de gènes *ank* liés aux phages est relativement stable d'une souche à une autre (entre 10 et 13, en excluant *wBm* qui ne possède pas de phage), ce qui suppose que les différences sur le nombre de gènes *ank* entre souches résident principalement dans le nombre de ces gènes présents dans le reste du chromosome bactérien (Figure 14).

Bien qu'engagées dans des interactions protéine-protéine *via* leurs motifs ANK, la fonction des protéines à domaines ankyrines est encore inconnue. Presqu'un tiers des protéines ankyrines de *Wolbachia* étudiées possèdent trois motifs ANK (Figure 16). Cependant la gamme de motifs ANK dans une protéine est large et s'étend de 1 à 34 domaines (Figure 14 et 16). Nous avons mis à profit l'alignement de 913 motifs ANK contenus dans les protéines ankyrines de *wVulC* et des quatre souches séquencées de *Wolbachia* pour définir une séquence consensus et étudier le degré de conservation de ces domaines. La figure 17 représente plus précisément le consensus obtenu chez *wVulC* avec les 205 motifs ANK longs de 33 acides aminés exactement. Les résultats montrent que certains acides aminés au sein d'un motif ANK sont conservés et font la singularité d'un domaine ankyrine, caractéristique déjà observée par Mosavi *et al.* (2004) (Figure 17). D'après un consensus dessiné à partir de 4000 protéines ANK, Mosavi a pu déterminer le degré de conservation de chaque position. Son analyse a permis de montrer que les positions conservées influençaient la structure 2-D des molécules suivant un enchaînement de feuillets β et d'hélices α . Chaque motif ankyrine a de plus une composition en acides aminés en termes de propriétés biochimiques (hydrophobes, neutres, basiques et acides) assez homogène (Figure 17). Néanmoins, la grande variabilité en d'autres positions du motif consensus reflète la diversité observée des protéines ankyrines chez *Wolbachia*.

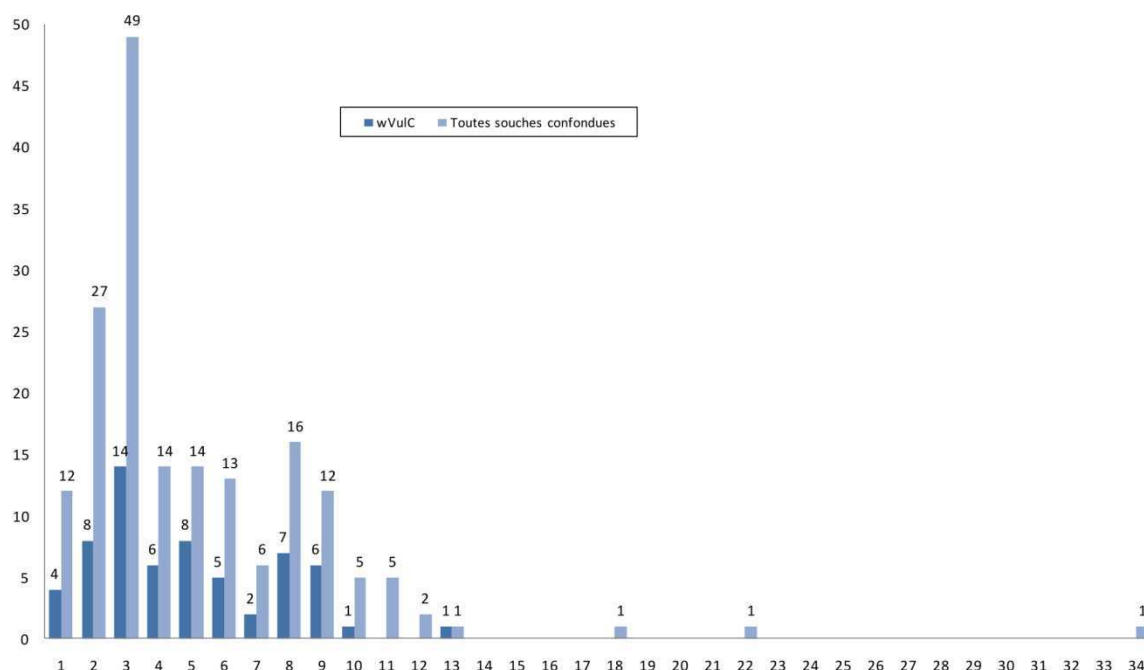


Figure 16: Nombre de protéines ankyrines de *Wolbachia* en fonction du nombre de motifs ANK qu'elles possèdent.

Diagramme représentant l'abondance relative de protéines ankyrines de *Wolbachia* possédant de 1 à 34 motifs ANK. Les analyses ont été menées sur 5 souches: *wVulC* (en bleu foncé), *wMel*, *wBm*, *wPip-Pel*, *wRi* et *wVulC* (valeurs cumulées en bleu clair).

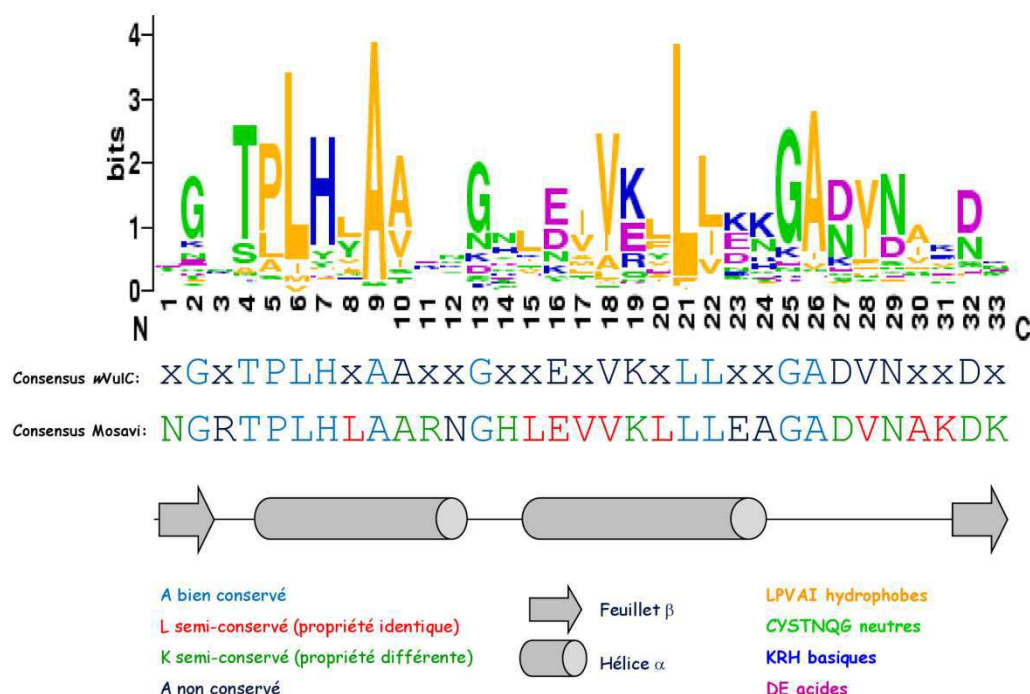


Figure 17: Motif ANK consensus des protéines ankyrines chez *wVulC*.

Séquence consensus réalisée à partir de l'alignement des motifs longs de 33 acides aminés exactement extraits des 66 protéines à domaines ankyrines chez *wVulC*. Le diagramme réalisé avec Weblogo (<http://weblogo.berkeley.edu/logo.cgi>) représente l'abondance relative de chaque acide aminé au niveau des 33 positions du motif. Certains acides aminés aux positions 2,4-6,9,13,21-22,25-26 sont très conservés.

L'identification *via* le serveur SMART de domaines autres qu'ankyrines permet d'appréhender les activités potentielles des protéines ankyrines de *wVulC*. Seules trois des 66 protéines possèdent chacune un domaine nommé DnaJ, Peptidase_M9, Rapsyn_N. Chez les procaryotes, les domaines DnaJ sont retrouvés dans des protéines chaperonnes du même nom et sont supposés intervenir dans les interactions avec d'autres protéines chaperonnes DnaK (Cyr *et al.* 1994). Peptidase_M9 est un domaine propre à la famille des métalloprotéases et plus précisément aux collagénases identifiées notamment chez les bactéries des genres *Vibrio* et *Clostridium* (Matsushita *et al.* 1994). Chez ces dernières, les collagénases sont utiles lors de l'invasion microbienne et interviennent dans la dégradation des barrières de collagène de l'hôte (Matsushita *et al.* 1994). Rapsyn_N est lui un domaine trouvé chez des protéines qui s'accumulent dans les jonctions neuromusculaires (Eckler *et al.* 2005). Rapsyn_N et Peptidase_M9 sont deux domaines uniquement trouvés chez les protéines ankyrines de *wVulC* alors que DnaJ est commun à trois protéines ANK de *wPip-Pel* (Walker *et al.* 2007). D'autres domaines ont été répertoriés mais uniquement chez les protéines ANK d'autres souches de *Wolbachia* [Arch_ATPase, Ovarian Tumor (OTU), ATPase family protein Associated with various cellular Activities (AAA), C-terminal Pox Viral (PRANC) ou encore Tetratricopeptide (TPR)] [cf. Article 4 : Ioannidis *et al.*, p. 54].

Il faut noter la présence dans 19 protéines ANK de *wVulC* d'un à deux domaines transmembranaires (Figure 14). Ces domaines ancreraient ces protéines dans les membranes de la bactérie, soit du côté intracellulaire, soit du côté extracellulaire. Différents serveurs de localisation cellulaire (Psortb, CELLO, TMHMM, Phobius) prédisent tous qu'une protéine ankyrine (*wVulC*_ANK89) et huit autres sont respectivement enchâssées dans la membrane externe et interne de la bactérie. D'après les mêmes analyses, vingt-trois autres protéines ANK seraient cytoplasmiques. D'autres serveurs (SignalP, PrediSi, TargetP, TatP, LipoP et Phobius) prédisent que quatre protéines ANK de *wVulC* possèdent par ailleurs une séquence peptide-signal dans leur région N-terminale dont une (*wVulC*_ANK50) possède un peptide signal plus long (39 aa) que les autres (Figure 14). Bien que la fonction de ces quatre protéines ANK soit encore inconnue, il est envisageable qu'elles soient exportées à travers la membrane interne et probablement par un système Tat pour *wVulC*_ANK50 à cause de son signal caractéristique. Les protéines ANK peuvent être également sécrétées par *Wolbachia*. Or, il a été montré que certaines protéines sécrétées par un T4SS possèdent un signal de translocation dans leur région C-terminale. Bien que difficilement détectable a priori, ces signaux ont été décrits pour la première fois chez les effecteurs d'*A. tumefaciens* (Vergunst *et al.* 2000, Vergunst *et al.* 2005). Ce signal C-terminal a également été mis en évidence dans d'autres protéines sécrétées par les T4aSS de *Mesorhizobium loti* (Hubber *et al.* 2004, Vergunst *et al.* 2005), *Bartonella henselae* (Schulein *et al.* 2005) et par le T4bSS de *L. pneumophila* (Luo et Isberg 2004, Nagai *et al.* 2005).

La recherche d'un signal de type R-x(7)-R-x-R-x-R-x(n) a été réalisée en 2007 par Annette Vergunst en utilisant un modèle caché en chaînes de Markov (Hidden Markov Model HMM) sur 20 protéines à domaines ankyrines de *wVulC*. J'ai plus tard réalisé la recherche d'une signature de type VirF d'*A. tumefaciens* ([RK]-x(7)-[RK]-x-[RK]-x[RK]-x(2)) ou de type Anka d'*A. phagocytophilum* ([RK]-x(2)-[RK]-x(2)-[RK]-x-[RK]-x(2)) à l'aide du serveur de Genouest nommé WAPAM (Weighted Automata PAttern Matching) sur les 66 protéines ankyrines de *wVulC*. Ces deux recherches se sont révélées infructueuses si bien qu'aucune ankyrine de *wVulC* ne possède une signature de translocation identique ou similaire aux effecteurs d'*Anaplasma* et *Agrobacterium*.

L'ensemble des analyses réalisées sur les protéines ANK de *wVulC* et la comparaison aux autres ankyrines de *Wolbachia* ont été intégrés à celles de P. Ioannidis (Université de Ioannina, Grèce) dans le cadre d'une collaboration avec l'équipe de K. Bourtzis [cf. Article 4, Ioannidis *et al.*, p. 54]. Le consensus de motifs ANK obtenus avec les protéines ankyrines de *Wolbachia* a été étendu à 1287 protéines (soit 3899 motifs), comprenant en plus celles présentes chez d'autres bactéries appartenant aux Rickettsiales et à *C. burnetii*, mais aussi chez trois virus et sept espèces d'eucaryotes (Table 1, p. 84). Les résultats montrent que la diversité des ankyrines chez *Wolbachia* est liée en partie à la variabilité de séquence des motifs ANK par le jeu d'évènements d'insertion ou de délétion de résidus au niveau de deux positions spécifiques au sein des motifs (Figures 4 et 5, p. 87-88). Les analyses de BLASTP obtenues avec les protéines à domaines ANK de *Wolbachia* ont également été étendues aux virus et eucaryotes étudiés. Les analyses ont été réalisées sur les protéines entières, les motifs ANK pris séparément ou ensemble, et sur le reste des protéines une fois les motifs ANK soustraits. Les résultats sous forme de relations de parenté montrent une origine eucaryote vraisemblable des protéines ANK de *Wolbachia* (Figure 1, p. 85). Ce résultat a été déterminé essentiellement par la séquence des motifs ANK proprement dits plutôt que par la séquence du reste des protéines (Figure 6, p. 89). Les nombreuses protéines à domaines ankyrines codées par le génome de *wVulC* ont vraisemblablement une origine eucaryote et représentent autant de protéines susceptibles d'interagir avec des protéines de l'hôte. La démonstration de leur implication dans la perturbation de la physiologie hôte se poursuivra en étudiant leur sécrétion effective par les bactéries et leurs effets potentiels quant à la manipulation de la reproduction de l'hôte.

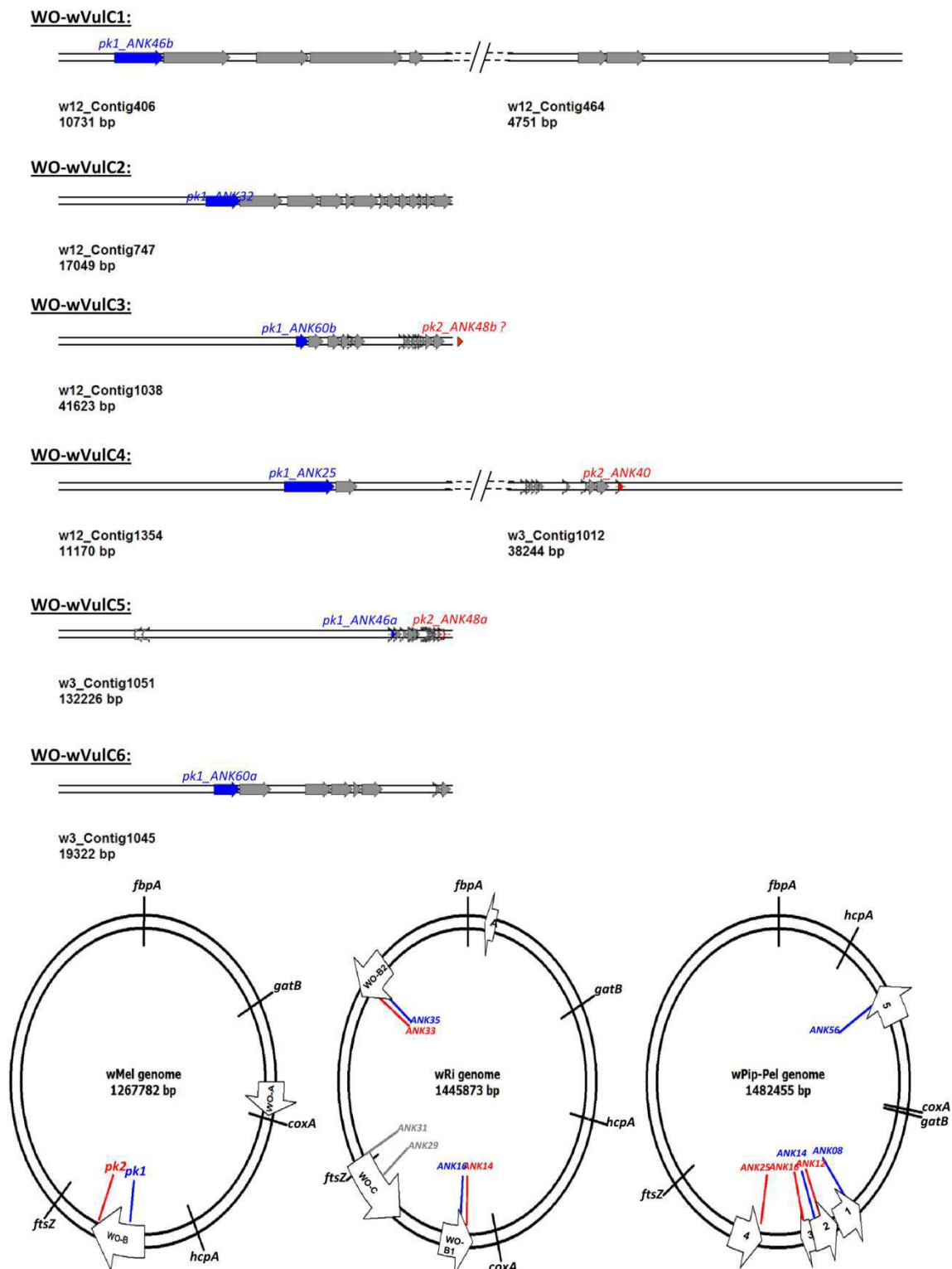


Figure 18: Cartographie génomique des prophages de *Wolbachia*.

A l'aide du logiciel pDRAW 3.2 (www.aacal.com), sont représentées les localisations génomiques des régions de prophage (sous forme de flèches), celle des gènes ankyrines associés aux phages dont *pk1* (en bleu) et *pk2* (en rouge) ainsi que celle de 5 marqueurs phylogénétiques (*fbpA*, *hcpA*, *coxA*, *gatB* et *ftsZ*) chez trois génomes (*wMel*, *wPip-Pel* et *wRi*). Les régions de prophages identifiées chez *wVulC* sont aussi représentées selon leur localisation sur différents contigs (représentation non proportionnelle). Chez *wVulC* et toutes les autres souches de *Wolbachia*, *pk1* et *pk2* encadrent un cluster de 13 gènes de phages (en gris). Les détails de cette figure sont disponibles dans l'article 5.

C. Distribution et expression des gènes ankyrines *pk1* et *pk2* chez les *Wolbachia* d'isopodes.

Steven Sinkins a été le premier dans la communauté scientifique à établir chez *wPip* une corrélation entre le niveau d'expression de certains gènes codant des protéines à domaines ANK et le phénotype d'IC induit par la bactérie chez les hôtes (Sinkins *et al.* 2005). Il a notamment montré que deux gènes (*pk1* et *pk2*) sur les 18 premiers gènes *ank* préalablement isolés du génome de *Wolbachia* de *C. pipiens* étaient polymorphes entre plusieurs souches de *wPip* (Sinkins *et al.* 2005). Ce résultat était de première importance car aucun des marqueurs moléculaires connus de *wPip* ne permettait jusqu'alors de distinguer les nombreux variants de *Wolbachia* infectant *C. pipiens*. De plus, l'analyse de l'expression de *pk2* au sein de deux populations incompatibles de *C. pipiens* (Bei, Beijing, Chine et Pel, Sri Lanka) s'est révélée être sexe-spécifique (ce gène ne s'exprimant que chez les femelles) et corrélée au phénotype d'IC (Sinkins *et al.* 2005).

Chez les isopodes terrestres, nous avons alors cherché à déterminer le degré de conservation des gènes *pk1* et *pk2* et leur patron d'expression chez des souches induisant une IC et chez celles induisant une féminisation, le phénotype le plus répandu dans ce groupe. L'ensemble des résultats sera prochainement soumis à publication [cf. Article 5 : Pichon *et al.*, p. 90]. Parmi les 66 gènes *ank* identifiés chez *wVulC*, nous avons pu mettre en évidence la présence de six paralogues du gène *pk1* et deux du gène *pk2* (Figure 18). Les analyses de Southern Blot et de séquençage nous permettent de considérer que le génome de *wVulC* doit comporter une troisième copie du gène *pk2*, non identifiée sur les contigs, vraisemblablement identique à l'une des deux autres copies de *pk2* (*wVulC*_ANK40), comme cela est le cas chez *wPip*-Pel (Klasson *et al.* 2008). En effet, chez cette souche, le gène *pk2* est contigu au gène *pk2+1*, l'ensemble de ces deux gènes étant présent en trois exemplaires. Or, un gène *pk2+1* orphelin de *pk2* a été localisé à l'extrémité 5' du contig w12_1038 (Figure 18). La troisième copie du gène *pk2* est très certainement localisée en aval de ce gène, dans la région non encore identifiée du génome de *wVulC* (Figure 18). L'avancement de l'assemblage du génome de *wVulC* devrait confirmer la présence de ce gène, que nous avons nommé *wVulC*_ANK40b par anticipation. Enfin, deux allèles du gène *pk1* (*wVulC*_ANK46 et *wVulC*_ANK60) sont retrouvés chacun en deux copies identiques.

Toutes ces copies cartographiées sur les contigs de *wVulC* sont contiguës à treize gènes de phage (sur un locus d'une longueur totale de ~10 kbp) dont la synténie est bien conservée chez les prophages de *Wolbachia* (Figure 18). Nous pensons d'ailleurs utiliser cette cartographie pour affiner l'assemblage du génome de *wVulC*. Nous avons d'ores et déjà utilisé ce cluster de 13 gènes de prophage encadré par les gènes *pk1* en 5' et *pk2* en 3' pour comparer l'histoire évolutive de ces gènes *ank* avec celle des prophages.

Les séquences des gènes *pk1* et *pk2* de *wVulC* sont globalement bien conservées en comparaison avec celles de leurs homologues chez d'autres souches. Les phylogénies réalisées à partir des données dans GenBank ont permis d'établir une nomenclature visant à subdiviser l'ensemble des séquences en familles de gènes. Ainsi, les gènes *pk1* se subdivisent en 3 types (*pk1a,b,c*) et *pk2* en 2 types (*pk2a,b*) (Figure 1, p. 109). Un des gènes *pk1* de *wVulC* se classe alors dans le type *pk1a* (*wVulC*_ANK25) et les cinq autres appartiennent au type *pk1b* (*wVulC*_ANK32, *wVulC*_ANK46a/b et *wVulC*_ANK60a/b) (Figure 1, p. 109). De la même façon, les trois gènes de *pk2* de *wVulC* font partie du type *pk2b* (*wVulC*_ANK40a/b, *wVulC*_ANK48) (Figure 1, p. 109). De plus, le nombre de copies de ces gènes est variable (Table 1 et S2, p. 111/113). Les séquences des produits de PCR et les résultats de Southern blot révèlent, en effet, que ces gènes sont présents en trois copies pour *pk2* et de 3 à 7 copies pour *pk1* selon les souches d'isopodes étudiées (Table 1 et S2, Figure S1, p. 111/113-114). Ces ankyrines de phage sont également retrouvées en plusieurs exemplaires chez *wRi*, *wPip-Pel* et *wCauB* infectant *Cadra cautella*, le nombre de copies étant lié au nombre de prophages (Figure 18; Table 1, p. 111) (Klasson *et al.* 2008, Klasson *et al.* 2009, Tanaka *et al.* 2009). Une analyse par BLASTN des 13 gènes encadrés par *pk1* et *pk2* dans les prophages de *Wolbachia* a permis d'identifier 6 prophages différents dans les contigs de *wVulC*. Par analogie avec ce qui est observé dans le génome de *wPip-Pel*, la présence de 6 copies de *pk1* et de seulement 3 copies de *pk2* chez *wVulC* (Table 1, p. 111 et Figure 18) montre que certains des loci de prophage sont tronqués chez cette souche (Figure 3, p. 110). Bien que dupliqués entièrement, partiellement ou recombinaison, les prophages semblent être à l'origine de l'histoire évolutive des gènes *pk1* et *pk2* (Figures 1 et 2, p. 110). La phylogénie des gènes *pk2* est d'ailleurs congruente avec celle des phages contenant ces gènes (Figure 2, p. 110). L'hypothèse de co-évolution entre prophages et *pk1* semble compromise et laisse à penser que des événements de recombinaison ou transposition ont eu lieu préférentiellement autour de *pk1* (Figure 2, p. 110). Nous avons, en effet, remarqué l'insertion de gènes codant des transposases en aval du gène *gpA* (gène qui suit directement *pk1*). Il reste à déterminer les conséquences évolutives et fonctionnelles de la présence de plusieurs copies des gènes *pk1* et *pk2* chez *wVulC*.

Nous avons étudié l'expression de chacun des paralogues de ces 2 gènes dans les gonades d'*A. vulgare*. A l'exception d'un, tous les allèles des gènes *pk1* et *pk2* s'expriment dans les gonades chez les femelles et les mâles lorsque ceux-ci sont infectés par *Wolbachia* et quelque soit le phénotype de l'interaction (IC ou féminisation) (Table 3, p. 112).

Ce résultat va à l'encontre de ce qui a été observé par Sinkins pour le gène *pk2* (Sinkins *et al.* 2005, Walker *et al.* 2007) et les gènes *wPip_ANK1*, *ANK52* et *ANK57* (Walker *et al.* 2007). De plus, l'une des copies du gène *pk2* (*pk2b1*) s'exprime uniquement dans les souches féminisantes testées (Table 3, p. 112). Aucun transcrit de cet allèle n'a été en effet détecté dans les gonades d'individus appartenant aux trois espèces d'isopodes porteurs de *Wolbachia* induisant une IC (Table 3, p. 112). Ce résultat est quelque peu inattendu étant donné les résultats de Sinkins montrant l'expression du gène *pk2* chez des souches induisant l'IC chez *C. pipiens* (Sinkins *et al.* 2005). Cela suggère que le gène *pk2* n'est pas impliqué dans l'expression de ce phénotype chez les isopodes terrestres ou bien qu'il peut être différentiellement exprimé au cours du temps.

D. Conclusions et perspectives.

Lorsque le génome de *wVulC* sera totalement annoté, nous pourrons finir de caractériser l'ensemble des gènes *ank* et confirmer la présence du gène *pk2* manquant. Nous pourrons également analyser, au sein des souches de *Wolbachia* infectant des isopodes, la distribution et l'expression des 18 gènes spécifiques à la souche *wVulC*, ainsi que celle des orthologues des gènes dont l'expression différentielles entre mâles et femelles a été démontrée chez d'autres souches. Même si il n'y a pas de corrélation sûre entre l'expression des gènes *ank* et les phénotypes induits par la bactérie, le gène *pk2* de *wVulC* semble d'ores et déjà un candidat impliqué dans la manipulation de la reproduction de l'hôte *A. vulgare*. En conséquence, il est primordial de démontrer si les produits du gène *pk2* sont sécrétés vers le cytoplasme de l'hôte et interagissent avec des cibles hôtes. Il est donc nécessaire dès aujourd'hui de mettre au point une technique permettant d'identifier *in vivo* des facteurs de virulence bactérienne chez *Wolbachia*. Comme cela a été fait dans des modèles de bactéries pathogènes intracellulaires, une possibilité est de tester si des protéines sont sécrétées vers le cytoplasme de l'hôte *via* un T4SS (Lin *et al.* 2007).

E. Publications.

E.1 - Ioannidis *et al.* En préparation.

Article 4: Ioannidis P., Klasson L., Chen L., Pichon S., Grève P., Bouchon D., Garrett R.A., Andersson S., Braig H. and K. Bourtzis. *Wolbachia* ankyrins: an abundant and diverse protein group of eukaryotic origin? En préparation.

***Wolbachia* ankyrins: an abundant and diverse protein group of eukaryotic origin?**

Panagiotis Ioannidis¹, Lisa Klasson², Lanming Chen³, Samuel Pichon⁴, Pierre Grève⁴, Didier Bouchon⁴, Roger Garrett³, Siv Andersson², Henk Braig⁵ and Kostas Bourtzis^{1*}

¹Department of Environmental and Natural Resources Management, University of Ioannina, 2 Seferi Street, 30100 Agrinio, Greece.

² Department of Molecular Evolution, Evolutionary Biology Center, Uppsala University, Sweden.

³University of Copenhagen, Institute of Molecular Biology, Biocenter, Ole Maaloes Vej 5, 2200 N Copenhagen, Denmark.

⁴ Université de Poitiers, Laboratoire Ecologie, Evolution, Symbiose, UMR CNRS 6556, 40 avenue du recteur Pineau, F-86022 Poitiers Cedex, France.

⁵ School of Biological Sciences, University of Wales, Deiniol Road, LL57 2UW, Bangor, United Kingdom.

* Corresponding author

E-mail addresses:

Panagiotis Ioannidis - panioan@cc.uoi.gr

Lisa Klasson - lisa.klasson@ebc.uu.se

Lanming Chen - lanmingchen@yahoo.com.cn

Samuel Pichon - samuel_pichon@hotmail.com

Pierre Grève - pierre.greve@univ-poitiers.fr

Didier Bouchon - didier.bouchon@univ-poitiers.fr

Roger Garrett - garrett@mermaid.molbio.ku.dk

Siv Andersson - Siv.Andersson@ebc.uu.se

Henk Braig - h.braig@bangor.ac.uk

Kostas Bourtzis - kbourtz@uoi.gr

Abstract

Background

Bacteria of the genus *Wolbachia* are endosymbionts of a wide variety of arthropods and filarial nematodes. In most symbiotic associations, *Wolbachia* behaves as a reproductive parasite exploiting its host's reproduction in order to spread in a population. Recently, it has been shown that all *Wolbachia* strains, except one, contain an unusually high number of ankyrin repeat containing genes. Ankyrin repeats mediate protein-protein interactions and have been characterized, mainly, in eukaryotes. Their large number and increased diversity in *Wolbachia* suggests that they possibly play an important role in *Wolbachia* biology.

Results

In the present work, 287 ankyrin genes were examined from 9 different *Wolbachia* strains. Additionally, 1297 ankyrin genes were also included in this study originating from eukaryotic organisms, bacteria closely related to *Wolbachia* and viruses. Various aspects were studied such as ankyrin repeat consensus, ankyrin repeat clusters, other domains present in ankyrin genes and a BLAST-based analysis. Comparative analysis of the ankyrin repeat consensus and BLASTP hits clearly suggests a eukaryotic origin of most of the *Wolbachia* ankyrins. The diversity of *Wolbachia* ankyrins is partly due to insertions occurring exclusively in two positions of the repeat without affecting its tertiary structure. A number of additional domains have been detected including AAA ATPase, TPR, PRANC, DnaJ and OUT domains, which provide clues for the potential function of the ankyrin genes harboring them.

Conclusions

The large number and great diversity of *Wolbachia* ankyrin genes is very intriguing and suggestive of a key role in *Wolbachia* biology. The herein suggested eukaryotic origin of these genes, as well as the wealth of additional domains present, supports the hypothesis that *Wolbachia* ankyrins interact with host factors in order to establish symbiotic associations and/or manipulate host reproduction. Study of these genes could unveil key molecular mechanism(s) on the basis of these phenomena.

Background

Wolbachia are Gram-negative, obligatory intracellular alpha-Proteobacteria that infect many invertebrates including insects, terrestrial crustaceans, mites, scorpions, spiders and filarial nematodes (Werren *et al.* 2008; Werren 1997; Taylor *et al.* 2005; Hilgenboecker *et al.* 2008; Duron *et al.* 2008). Filarial nematode *Wolbachia* strains establish mutualistic associations with their hosts (Taylor *et al.* 2005). Most of the arthropod *Wolbachia* strains are considered as reproductive parasites since they promote their proliferation and spread in host populations by inducing reproductive alterations including cytoplasmic incompatibility (CI), parthenogenesis, feminization and male killing (Werren 1997; Stouthamer *et al.* 1999; Werren *et al.* 2008).

The ankyrin (ANK for the purpose of this study) domain is a short, 33-amino acid domain that is common in many proteins (Heringa 1998; Mosavi *et al.* 2004). ANK domains mediate protein-protein interactions (Sedgwick and Smerdon 1999) and can be involved in diverse functions acting as transcription, development and/or cell cycle regulators, or as components of cytoskeleton and cell membranes, or as toxins, ion transporters and signal transducers (Bork 1993; Michaely and Bennett 1992; Mosavi *et al.* 2004; Sedgwick and Smerdon 1999). The first ANK genes were identified in yeast, *Drosophila* and *C. elegans* (Breden and Nasmyth, 1987). To this date, the Pfam database contains 64674 ANK domain sequences. Although ANK genes were initially detected and characterized in eukaryotes, recent studies report that bacteria (Seshadri *et al.* 2003; Wu *et al.* 2004; Madan *et al.* 2007 (R. bellii - unpublished); Ogata *et al.* 2006; Ogata *et al.* 2005; Cho *et al.* 2007; Nakayama *et al.* 2008), archaea (Anderson *et al.* 2008) and viruses (Raoult *et al.* 2004; Afonso *et al.* 2000; Tulman *et al.* 2004; Moreau *et al.* 2009) harbor a large number of ANK genes.

Masui *et al.* (2000) were the first to detect the presence of ANK genes in *Wolbachia*. They reported the presence of 4 ANK genes in the wKue strain, three of which were located in the WO *Wolbachia* prophage. Recent studies and genome projects clearly indicated the presence of an astonishingly high number of ANK genes in *Wolbachia* compared to most bacteria (Wu *et al.* 2004; Foster *et al.* 2005; Salzberg *et al.* 2005a, b; Klasson *et al.* 2008a,b). The 1.27 Mb genome of the CI-inducing wMel strain, which infects the vinegar fly *Drosophila melanogaster*, contains at least 23 ANK genes (Wu *et al.* 2004) while the 1.48 Mb genome of the CI-inducing wPip strain, which infects the mosquito *Culex pipiens*, contains 60 ANK genes (Klasson *et al.* 2008a). On the other hand, the 1.08 Mb genome of the wBm strain, which has established a mutualistic association with the filarial nematode *Brugia malayi*, contains only 5 ANK genes (Foster *et al.* 2005). The large number, particularly in the arthropod *Wolbachia* strains, as well as the considerable diversity of ANK genes suggests that these proteins may play an important role in *Wolbachia* biology.

These initial findings stimulated future research in this field in an attempt to unravel the molecular mechanisms involved in *Wolbachia* symbiosis and *Wolbachia*-induced reproductive alterations. Iturbe-Ormaetxe and colleagues (2005) observed variation in the distribution, sequence variation and expression of *wMel*-like ANK genes in closely related *Wolbachia* strains. Their findings suggest that this variation may be associated with the reproductive phenotypes induced by *Wolbachia* in the *Drosophila* hosts. Sinkins and colleagues (2005) found sequence variation in two *wPip* *Wolbachia* strain ANK genes (*pk1* and *pk2*) between two incompatible *Culex quinquefasciatus* mosquito strains (Pel and Bei). In the same study, it was shown that one of the ANK genes, *pk2* (later renamed *wPip*_ANK12 and *wPip*_ANK25 since it is present in two identical prophage associated copies in the *wPip* genome (Klasson *et al.* 2008a)), is expressed only in females (Sinkins *et al.* 2005). In addition, the same research group reported that three additional *wPip* ANK genes (*wPip*_ANK1, *wPip*_ANK52 and *wPip*_ANK57) were differentially expressed in males and females of the *Culex quinquefasciatus* Pel strain (Walker *et al.* 2007). A similar differential expression was observed for the *wPip*_ANK16 gene (a homolog of *pk2*) in *Culex molestus* (Walker *et al.* 2007) further supporting the notion that these genes may be involved in *Wolbachia*-induced reproductive manipulation. However, Duron and colleagues (2007) investigated the variability, in sequence and in expression, of 58 *wPip* ANK genes in 14 genetically distinct *wPip* variants, which induce CI (uni- and bidirectional) in *Culex pipiens* complex mosquitoes (different than those examined by Sinkins *et al.*, 2005 and Walker *et al.*, 2007) and found no evidence of correlation between CI patterns and variability observed in *wPip* ANK genes (Duron *et al.* 2007).

ANK genes have also been reported in bacterial species of the genera *Ehrlichia*, *Anaplasma* and *Rickettsia*, which are close relatives of *Wolbachia*. In *Anaplasma phagocytophila*, the AnkA protein is secreted by a T4SS (Type IV Secretion System), is phosphorylated by Abl-1 to facilitate infection and has been shown to bind to human granulocyte DNA and nuclear proteins (Caturegli *et al.* 2000; Park *et al.* 2004; Lin *et al.* 2007). ANK proteins of *Legionella pneumophila* and *Coxiella burnetii* have also been implicated in the parasitism of protozoan and human host cells and were recently shown to harbour bacterial Type IV effectors (Seshadri *et al.* 2003; Habyarimana *et al.* 2008; Xiaoxiao *et al.* 2008). Based on the above, and given the fact that *Wolbachia* strains possess a large number of ANK genes and potentially a functional T4SS (Wu *et al.* 2004; Foster *et al.* 2005; Klasson *et al.* 2008a,b; Felix *et al.* 2008; Rances *et al.* 2008), the ANK genes are considered as excellent candidates to be involved in *Wolbachia*-induced mutualistic associations or reproductive phenotypes. In the present study, all available ANK genes present in the complete and / or partial genome of 9 *Wolbachia* strains were thoroughly analyzed *in silico*.

These strains belong to a taxonomically wide spectrum of *Wolbachia* supergroups (A, B and D) and participate in the establishment of mutualistic associations or in the induction of reproductive

phenotypes such as CI, feminization and parthenogenesis. The aim of the *in silico* analysis was to: (a) to determine and to characterize the *Wolbachia* ANK domain consensus, (b) to detect the presence of additional domains and (c) to search for related sequences in eukaryotes, prokaryotes and viruses through a BLAST-based approach. The comparative analysis provides an initial approach to unravel the evolutionary history and the potential functional properties of *Wolbachia* ankyrins.

Methods

Selection of species and strains

Sequence information was used from available genome data of different *Wolbachia* strains, other closely related bacterial species and strains as well as selected viruses and eukaryotic species (Table 1). Nine *Wolbachia* strains from supergroup A (*wMel*, *wRi*, *wUni*, *wAna*, *wSim*, *wWil*), B (*wPip* and *wVulC*) and D (*wBm*) were included in the dataset. Most of the strains are very well characterized and are known to induce cytoplasmic incompatibility (*wMel*, *wRi*, *wPip* and *wAna*), parthenogenesis (*wUni*), feminization (*wVulC*) or to establish mutualistic associations (*wBm*) (Hoffmann *et al.* 1986; Hoffmann, 1988; Stouthamer *et al.* 1990; Rousset *et al.* 1992; Bandi *et al.* 1995; Bourtzis *et al.* 1996). The *wSim* and *wWil* strains have not been biologically characterized yet (although according to Miller and Riegler (2006), *wWil* is most probably a mod-/resc- strain). Close relatives of *Wolbachia* were also included in the dataset. Nineteen species and strains of the *Rickettsiaceae* family were included in this study. These species and strains belong to the genera *Anaplasma*, *Ehrlichia*, *Neorickettsia*, *Orientia* and *Rickettsia* and their genome sequence is publicly available. *Coxiella burnetii* was also included in the dataset, although it is not a closely related species, as a bacterium that has a rather large number of ANK genes. The dataset also included 3 viruses and 7 eukaryotes (Table 1). The three viruses were selected because they have a rather increased number of ANK genes in their genomes. The eukaryotes, on the other hand, are either known *Wolbachia* hosts (*Drosophila melanogaster*, *Apis mellifera*, *Nasonia vitripennis*, *Tribolium castaneum*) or closely related to known *Wolbachia* hosts (*Caenorhabditis elegans*, *Anopheles gambiae*). *Arabidopsis thaliana* was included in this study for comparative purposes (Becerra *et al.*, 2004).

ANK dataset generation

If the proteome of a given organism was available at the time of analysis, then it was directly used for the downstream analysis. If not, as was the case for the *wWil* strain, gene prediction took place using the publicly available glimmer3 program (Delcher *et al.*, 2007) with the default parameters. The gene prediction results were not edited and the protein sequence was extracted using locally

developed Perl scripts. It is important to note that redundant genes present in the annotated proteome of the *wAna* and *wSim* strains were not included in the dataset. The complete proteome of each organism was screened for the presence of ANK genes using the hmmer program of the HMMer 2.3 package (<http://hmmer.wustl.edu/>). The database used was Pfam_ls, obtained from <ftp://selab.janelia.org/pub/Pfam/>. A locally developed Perl script was used to parse the hmmer output and define the ANK gene dataset.

An open reading frame was considered as an ANK gene if it contained at least one ANK domain satisfying all of the following three criteria: (a) to have a positive score according to hmmer, (b) to have a “positional score” (see below) of at least 3 and (c) does not overlap with any other significant domain. ANK genes identified using these criteria are shown in Table 1. The criterion of “positional score” was used in order to filter out those domains not having conserved certain positions of the ANK domain. According to PFAM and Mosavi (2004), there are 9 conserved positions in the ANK domain, 3 of which are strongly conserved. More specifically, positions G-2, T-4, P-5, H-7, G-25 and A-26 are conserved and add 0.5 points while amino acids L-6, A-9 and L-21 are strongly conserved and add 1 point to the “positional score” criterion. A candidate ANK domain should have at least 3 points in order for the gene containing it to be considered as a potential ANK gene. The number of *Wolbachia* phage-related ANK genes (shown within parentheses in Table 1) was also determined by using a BLAST approach and as reference the genome of *wMel*, *wRi* and *wPip* *Wolbachia* strains.

ANK repeat consensus

The ANK repeat consensus was inferred for each organism studied using only the canonical (*i.e.* those being exactly 33 amino acids long) ANK repeats (Table 1). A locally developed Perl script was used to extract the sequence of canonical ANK repeats from the hmmer output. A multiple sequence alignment of all canonical ANK repeats in a given organism was created and that alignment was parsed to find the amino acid frequencies in each of the 33 positions of the ANK repeat using a locally developed Perl script. In a given position of the ANK repeat, a “dominant” amino acid is the one present in more than 40% of the cases while “abundant” amino acids are present in at least 20% of the canonical ANK repeats. The ankyrin repeat consensus determined for *Wolbachia*, *R. bellii* OSU, *R. bellii* RML, *R. felis*, *O. tsutsugamushi* (both strains), *Ehrlichia* genus, *Anaplasma* genus, *C. burnetii*, Mimivirus, Canarypox, Fowlpox, *A. gambiae*, *A. mellifera*, *D. melanogaster*, *T. castaneum*, *N. vitripennis*, *C. elegans* and *A. thaliana* were compared to each other (note that *Ehrlichia* and *Anaplasma* ANK repeat consensus was taken for the whole genus, and not separately for each species, due to the low number of ANK repeats found in each species). Briefly, the frequency of each amino acid in each position of

the ANK repeat was calculated. Then, the absolute difference was calculated between all of the above organisms, for each amino acid at each position. All these absolute differences in a pairwise comparison were added and the resulting number indicated the “divergence” between the ANK repeat consensus of the taxa being compared.

All divergence indicator numbers for each pairwise comparison were subsequently divided by an appropriate number so as to bring the numbers between 0 and 1 and were output in a distance matrix. This matrix was then given as input to the “neighbor” program of the Phylip 3.63 package (Felsenstein, 1989) and, using the UPGMA method, a dendrogram was drawn and viewed using TreeViewX.

ANK repeat clusters

ANK repeat clusters were determined for each organism studied. A cluster of ANK repeats is defined as two or more adjacent ANK repeats in which the distance between any two repeats is at most 20 amino acids (Becerra *et al.*, 2004). ANK repeat clusters were detected based on the hmmpfam output file using locally developed Perl scripts. Strict and loose criteria were used for the identification of ANK repeat clusters. In the dataset generated using strict criteria, the ANK domains participating in clusters should (a) be 33 amino acids long (*i.e.* only canonical ANK repeats were used), (b) have a positive hmmpfam score and (c) have a “positional score” greater than or equal to 3. In contrast, only rules (b) and (c) are taken into account when building clusters using loose criteria. The dataset generated by strict criteria was used in phylogenetic analysis based on clusters while that generated by loose criteria was used in the rest of the cluster-related analyses (see Results and Discussion).

The ANK repeat clusters were also characterized based on: (a) the distribution of ANK repeats of different “quality” (see below) within each cluster and (b) the gaps present within the ANK repeats. This analysis was performed taking into consideration the following two properties of the ANK repeat: (a) its hmmpfam e-value and (b) its length. The lower the e-value and the closer the length to 33 (the typical length of ANK repeats), the better the quality of the ANK repeat. Based on these two criteria, the ANK repeats were divided into four “quality” categories, starting with that of the best quality and ending with that of the worst: Category A includes ANK repeats of good e-value (<0.1) as well as of a normal length (between 30 and 36 amino acids); Category B includes ANK repeats that do not have a good e-value but have a normal length; Category C includes ANK repeats that have good e-value but are relatively long while Category D includes ANK repeats that do not have a good e-value and are relatively long. In order to distinguish between different ANK repeat categories (*i.e.* qualities), a locally developed Perl script was initially used to output repeat quality information for

each gene in a plain text format. This output was subsequently used as input for the DrawClusters program (see below) in order to visualize the results for each gene and calculate the occurrence of each category of domains. Insertion and deletion events were found based on the hmmpfam alignment of each ANK domain for each gene.

Using DrawClusters, the exact position of the insertion/deletion event was recorded and that information was subsequently used in order to find whether these events presented preference for some specific position within the ANK domain. DrawClusters was developed in the framework of this study using Visual Basic 2005 Express Edition.

Phylogenetic study of ANK repeat clusters

Phylogenetic analysis was only performed for the ANK repeat clusters dataset obtained using strict criteria. Briefly, the sequences of ANK repeat clusters having the same number of repeats (hence, exactly the same length) were aligned using ClustalX 1.83 (Thompson *et al.*, 1997). Subsequently, PAUP* 4.0b10 (Swofford, 2000) was used for phylogenetic tree building making use of the Maximum Parsimony algorithm. Tree files generated were viewed with TreeViewX 0.4 (Page, 1996).

Additional domains in ANK genes

A locally developed Perl script was used in order to parse the hmmpfam output file and to determine the presence of additional domains within ANK genes. An additional domain should have a positive hmmpfam score and an e-value of less than 5. Moreover, the alignment length (*i.e.* the area of similarity) should be between 45% and 180% of the corresponding length of the domain model (*i.e.* if a domain is 100 amino acids long then the alignment it produces should be between 45 and 180 amino acids). This latter criterion was used in order to exclude hits containing an excessive number of gaps (either in the hmm model or the query sequence). Subsequently, all candidate domains were also examined by eye. Transmembrane domains (transmembrane helices) were predicted using the TMHMM2.0 web server (Krogh *et al.*, 2001) with the default parameters. When such a helix was located in the first 60 amino acids of a protein, it was tagged as “possible secretion signal” by TMHMM. In case of domain overlap, the conflicting domains were examined by eye to determine which one has better e-value and alignment so as to exclude the least significant one. Finally, the file containing the information about all domains (both ANK and others) of all ANK genes was given as input in the program “DrawAnks” in order to visualize the results and manipulate the data more easily. DrawAnks was written in Visual Basic 2005 Express Edition in the framework of this project.

BLAST analysis of the ANK genes

An extensive BLAST analysis was performed using the dataset of the *Wolbachia* ANK genes (Altschul *et al.*, 1997). The sequence of the whole ANK gene, the ANK repeat cluster sequence and the leftover sequence were extracted and BLASTed against the protein nr database (using a cutoff e-value of 1e-02) downloaded from the NCBI ftp site (<ftp://ftp.ncbi.nih.gov/blast/db/>).

At most 5 eukaryotic, 5 prokaryotic and 5 viral best hits were extracted. It was then calculated whether the first hit was eukaryotic, prokaryotic or viral. This calculation was done for each whole ANK gene, ANK repeat cluster and leftover sequence. The list of hits for whole ANK genes, ANK repeat clusters and leftovers were examined by eye in order to locate interesting hits (that is, hits that for example were not simply annotated as “ANK gene” or “hypothetical protein”) and subsequently the count of each one in whole genes, clusters and leftovers was determined. All calculations mentioned in this section were performed using locally developed Perl scripts.

Results

ANK repeat consensus

Using only the canonical (*i.e.* exactly 33 amino acids long) ANK repeats (Table 1), the ANK repeat consensus was identified in each (group of) organism(s) studied as shown in Table 2. Only the “dominant” amino acids are shown in Table 2, while the exact frequencies of both “dominant” as well as “abundant” amino acids are shown in Supplementary Table 1. It is evident that the *Wolbachia* ANK repeat consensus includes all 9 generally conserved positions (based on Pfam, see Methods): the strongly conserved positions 6, 9 and 21 as well as positions 2, 4, 5, 7, 25 and 26 (for more details see Supplementary Tables 1 and 2). No differences were observed in these 9 conserved positions among the *Wolbachia* strains studied. In addition, 8 more amino acid positions are represented with frequencies >40%. These are positions 10 (46.6% A), 13 (57.1% G), 18 (65.1% V), 19 (40.5% K), 22 (45.5% L), 28 (45.4% V), 29 (61.3% N) and 32 (41.8% D). It is worth noting that Mosavi and colleagues (2002) also classified positions 10, 13, 18, 22, 28, 29 and 32 as conserved.

The ANK repeat consensus was also determined for the closely related genera of *Wolbachia*: *Rickettsia*, *Orientia*, *Anaplasma*, *Ehrlichia* and *Neorickettsia* (Table 2 and Supplementary Tables 1 and 2). The *Rickettsia* ANK repeat consensus includes all 9 generally conserved positions but two: 5 and 7. Surprisingly, these two positions belong to the TPLH “signature”, the most prevalent motif within the ANK repeat (Mosavi *et al.*, 2004). In principle, both positions should be conserved. Other conserved amino acid positions are 14, 22, 27 and 29 (Table 2 and Supplementary Tables 1 and 2). The ANK repeat consensus determined for the genera *Anaplasma*, *Ehrlichia* and *Neorickettsia* should be considered cautiously since the number of the available canonical ANK repeats for these three genera

was rather limited. It is also interesting to note that the *Orientia* and *Coxiella* ANK repeat consensus exhibit variability in the positions 5 and / or 7 of the TPLH “signature”. The ANK repeat consensus determined for the 7 eukaryotic species studied are very similar to each other (Table 2 and Supplementary Tables 1 and 2). The ANK repeat consensus of all 7 species includes all 9 generally conserved amino acid positions (2, 4, 5, 6, 7, 9, 21, 25 and 26). In addition, 3 more amino acid positions are conserved in all eukaryotes examined; positions 13, 18 and 22. All eukaryotes also present amino acid position 10 quite conserved (alanine, ranging from 38.7% to 55.8%).

The ANK repeat consensus was also determined for the 3 viruses included in this study. The most striking observation is that Mimivirus' ANK repeat consensus is very different when compared not only to the other two viruses but, virtually, to any other organism studied (Table 2 and Supplementary Tables 1 and 2). The three major differences are: (a) the amino acid positions 2, 4 and 7 are not conserved in Mimivirus (positions 4 and 7 are 2 of the 4 residues found in the conserved TPLH “signature”, the most prevalent motif within the ANK repeat sequence), (b) amino acid position 6 is represented in Mimivirus by leucine with an average percentage of 49.5% while in all other organisms the observed percentages are over 78% and (c) amino acid positions 13, 14, 15, 16, 18, 19, 20, 22, 27 and 28 are only conserved in Mimivirus. Last, the ANK repeat consensus of some representatives was compared to each other and the result was depicted as a dendrogram (Figure 1). Intriguingly, it shows that *Wolbachia*'s consensus is far more related to that of eukaryotes than to any other (proteo)bacterium. Actually, *Wolbachia*'s consensus is even more related to that of the Fowlpox- and Canarypox-viruses than *Coxiella* or *Rickettsia*. A similar conclusion is drawn from the BLAST results (see below) and strengthens even more the notion that *Wolbachia*'s ANK genes have a eukaryotic rather than prokaryotic origin.

Organization and characterization of ANK repeats in ANK clusters

Adjacent ANK repeats are considered to be essential for the biological function of a single ANK repeat (Becerra *et al.*, 2004). Using the loose criteria (described in Materials and Methods), ANK clusters were identified in each (group of) organism(s) studied as shown in Figure 2. Isolated ANK repeats, although are not considered as ANK clusters, they are herein presented as 1-repeat ANK clusters for comparative purposes. The total number of ANK clusters detected in the 9 *Wolbachia* strains studied is 390. The 1-repeat ANK “clusters” are, by far, the most frequent in *Wolbachia* (194 or 38% of the total ANK clusters) followed by the 3-repeat (60 or 19%), 2-repeat (51 or 16%) and 5-repeat (29 or 9%) ANK clusters (Figure 2 and Suppl Table 3). All *Wolbachia* strains studied contain a large number of isolated ANK repeats (1-repeat clusters). It is worth noting that one of them, the *wBm* strain, has 10 out of 14 ANK clusters present as 1-repeat ANK “clusters”. This relative abundance of

isolated repeats is, by far, the highest of all *Wolbachia* strains. Another interesting observation is that all *Wolbachia* strains studied contain at least one 3-repeat ANK cluster. This class of ANK clusters has more members in the *wRi*, *wAna* and *wPip* strains. Three exceptionally lengthy ANK clusters were observed in certain *Wolbachia* strains: (a) a 25-repeat ANK cluster is present only in *wRi* (ANK26); this ANK gene also contains a cluster of tetratricopeptide repeats (TPR), (b) a 17-repeat ANK cluster is also present in *wRi* (ANK28) and (c) a 11-repeat ANK cluster is detected in *wMel* (WD_0385).

According to Becerra *et al.* (2004), single ANK repeat that appears isolated is probably not functional at all. In addition, it has been reported that the outer repeats of an ANK cluster could be the ones to “decay” first (Mosavi *et al.* 2002, 2004) (by “outer” we mean the two outermost repeats). To investigate these scenarios, the distribution of the ANK repeats according to their quality status (four quality categories A to D were defined in Material and Methods) within a cluster was determined (Figure 3). It is clear that the great majority of the high quality ANK repeats (category A) are not present in isolated domains. Fifty three percent of the category A ANK repeats are located in the inner repeats of a cluster, 38% are detected at the outer repeats of a cluster while only 8% of them are present in isolated ANK repeats. On the other hand, 26% of the category D repeats exist as isolated ANK repeats while 32% are located in the outer ANK repeats and 42% in inner ANK repeats. Categories B and C of ANK repeats have percentages that vary in between the abovementioned values.

Insertion and/or deletion events (based on the hmmpfam alignment) were observed in several repeats of the ANK clusters. The distribution of these indels within an ANK cluster was studied (*i.e.* whether they are in an isolated repeat, in an inner or in an outer repeat of a cluster) and the results are shown in Figure 3 (for insertions and deletions). Three hundred eighty seven insertion events were detected: 73 (19%) of them occur in 148 isolated repeats, 157 (41%) in 484 outer and 157 (40%) in 539 inner repeats of an ANK cluster. On the other hand, a significantly lower number of deletion events (only 110) were observed: the majority of them occur in the inner (46%) and outer (40%) repeats of an ANK cluster while only 14% of the deletion events happen in isolated ANK repeats. The number of deleted amino acids per ANK repeat is about the same for the inner and the outer repeats of an ANK cluster as well as for the isolated ANK repeats.

The site preference of insertions/deletions within ANK repeats was also investigated based on hmmpfam alignments. The deletion events appear to have no site preferences. However, as shown in Figure 4, insertions occur only in two positions of the ANK repeat: (a) between amino acid positions 13 and 14 (194 events – 50.1%) and (b) between amino acid positions 23 and 24 (193 events – 49.9%). The preference for these two insertion positions does not seem to be random. In fact, the first site is located between the two alpha-helices of the ANK repeat while the second site is very close to the end

of the second alpha-helix. It is also worth noting, that despite the fact that the number of insertion events is the same in both positions, the lengthier insertions are observed between amino acids 13 and 14, that is, between the two alpha-helices (Supplementary Figure 1). All inserted amino acid fragments were BLASTed against the protein nr database: no significant domain fragments or hits were detected. In addition, calculation of the charge of the insertion showed that there is no specific charge in the inserted peptide.

Another interesting observation is that the majority of the insertion events have occurred in chromosomal rather than phage ANK genes. In total, 2001 amino acids have been inserted in ANK genes of the nine *Wolbachia* strains studied. Of these, 1410 amino acids have been inserted into chromosomal ANK genes while 583 amino acids have been inserted in phage ANK genes (the remaining 8 amino acids have been inserted in ANK genes of unknown origin). This is a general trend for each *Wolbachia* strain studied except *wWil* (host: *D. willistoni*) (Supplementary Figure 2). The *wWil* strain contains twice as many inserted amino acids in phage ANK genes when compared to chromosomal ones. In *wWil* in particular, the ANK gene WD_0234 contains a rather big insertion, 72 amino acids long. It is the longest insertion observed in a phage ANK gene in *Wolbachia* and it is only comparable to the 42 amino acid insertion found in *wMel*'s WD_0294 phage ANK gene. It is worthwhile to mention that the WD_0385 gene of *wMel* is the most heavily impacted since it contains 11 insertions, one of which is 118 amino acids long.

Twenty-seven out of the 287 *Wolbachia* ANK genes studied contained more than one ANK repeat clusters (Table 3), which were predicted using the strict criteria (and without taking into account the 1-domain clusters). These genes were present in 8 out of 9 *Wolbachia* strains studied; the *wBm* strain did not contain any ANK gene with multiple ANK repeat clusters. The origin of these multiple ANK repeat clusters within a given ANK gene was investigated through a phylogenetic approach. NJ trees were built for each cluster class (*i.e.* one tree for 2-repeat clusters, one for 3-repeat, etc) separately. Subsequently, all genes harboring more than one repeat clusters were identified and the phylogeny of these clusters in the different trees was compared. It was found that in 20 out of the 27 ANK genes harboring more than one ANK repeat clusters the phylogeny of the clusters was congruent, suggesting a common evolutionary history. An example is given in Figure 5 (left part of the picture) for the *wVulC*_ANK26 gene. In the remaining 7 ANK genes, which contained multiple ANK repeat clusters, the NJ trees produced were incongruent suggesting that each ANK cluster has followed its own evolutionary route. An example is also given in Figure 5 (right part of the picture) for the *wVulC*_ANK51 gene. In three cases of ANK genes (*wVulC*_ANK72, *wVulC*_ANK74 and *wRi*_ANK26) containing two ANK repeat clusters of the same length (*i.e.* same number of ANK domains), the two ANK clusters cluster together. This result suggests a possible duplication event of,

probably, the entire ANK repeat cluster. In all other cases, ANK repeat clusters of the same length appear phylogenetically unrelated to each other (data not shown).

Presence of additional domains in *Wolbachia* ANK genes

A major challenge in *Wolbachia* research is to unravel the biological function of the large number of ANK genes present in the genome of various strains. One way to approach this question is by searching for additional domains using the HMMer package (see Methods). In total, 287 *Wolbachia* ANK genes were studied; 181 of them contain only ANK repeats. All other 106 *Wolbachia* ANK genes contain at least one out of 12 additional domains identified. Examples of the structural and potential biological function, variation present in *Wolbachia* ANK genes is shown in Supplementary Figure 3 (see also Supplementary Table 4). Eight out of the 12 additional domains seem to be highly significant exhibiting an e-value of <1 and a high score (>11.6; OTU domain in WP_0462 is excluded since the score is rather low in relation to its length). These domains, in descending order of significance, are the transmembrane (TMHelix), AAA ATPase (AAA), tetratricopeptide repeat (TPR), C-terminal Pox viral (PRANC), DnaJ, ovarian tumor (OTU), osmosensory coiled coil (Osmo_CC) and half-A TPR (HAT) domains. The other 4 domains have a rather high e-value but are included in the list of potential additional domains only because they also have a relatively high score. These less significant domains are the Streptococcal surface antigen (Strep_SA_rep), PEST motifs (Tash_PEST), Pox virus A-type inclusions (Pox_A_type_inc) and M protein repeat (M) domain.

Presence of additional domains in ANK genes was also searched in all Rickettsiales using the same criteria as for *Wolbachia*. In *Anaplasmataceae* (*Anaplasma*, *Ehrlichia* and *Neorickettsia*), the most common additional motif appears to be the transmembrane helix, which is present in 6 genes. In fact, transmembrane helices in ANK genes are only found in *Anaplasmataceae*. PRANC domains are frequently present in both *Orientia tsutsugamushi* strains (8 genes in the Boryong and 17 genes in the Ikeda strain). The most interesting additional domain found in *Rickettsia* species is the RelA/SpoT, which is present in a single gene of *Rickettsia felis* (Cashel, 1974).

***Wolbachia* ANK genes: what can a BLASTP analysis tell us?**

McGraw and O'Neill (2004) suggested that some of the *wMel* ANK genes are more similar to eukaryotic than to prokaryotic genes. We extended their analysis to a much broader spectrum of *Wolbachia* ANK genes and strains using a composite protein BLAST (BLASTP) analysis. Actually, three BLASTP searches were performed for each ANK gene using: (a) the whole ANK gene sequence, (b) the ANK cluster sequence alone and (c) the leftover sequence. Based on the best BLAST hit obtained, each sequence was classified as eukaryotic- or prokaryotic-like (Figure 6).

No viral hits were observed. The two most important findings of this analysis are: (a) based on the whole ANK gene sequence BLASTP results, the majority (about 66%) of the 287 ANK genes studied originating from 9 *Wolbachia* strains seem to be eukaryotic-like and (b) this similarity is essentially due to the ANK domain clusters since leftover sequence BLASTP results show that only 5.7% of the sequences are eukaryotic-like while this percentage increases to 56.2% when ANK domain clusters BLASTP was performed (Figure 6).

A count of the significant BLASTP hits is shown in Supplementary Table 5A while the complete list of hits per gene is shown in Supplementary Table 5B (for each BLASTP run, at most 5 eukaryotic, 5 prokaryotic and 5 viral best hits were extracted). It is evident that the most frequently occurring hits are against unc44 ankyrins, ppGpp pyrophosphohydrolases/synthetases, kinase proteins, ion channels and Pfs/NACHT domain proteins. The majority of these hits are due to sequence similarities at the whole ANK gene and / or ANK domain cluster level (Supplementary Tables 5A and 5B).

Discussion

There are currently four completely sequenced genomes of *Wolbachia* and a few others will soon be available (Wu *et al.* 2004, Foster *et al.* 2005; Klasson *et al.* 2008a,b). All of them have at least one feature in common: an unusually large number of ankyrin repeat-containing genes when compared to any other prokaryotic group of organisms. An exception to this rule is the genome of the *wBm* strain, which contains only 10 ankyrin genes. This peculiarity is probably due to the mutualistic relationship this strain has with its host, the filarial nematode *Brugia malayi* (Foster *et al.* 2005). Actually, the ankyrin gene “density”, which is expressed as the ratio of ankyrin genes divided by the corresponding genome size, is very high in *Wolbachia*, and in viruses, compared to other prokaryotes and/or eukaryotes. Ankyrin genes in *Wolbachia* genomes are also very variable (Iturbe-Ormaetxe *et al.* 2005, Sinkins *et al.* 2005, Walker *et al.* 2007). Variability is evident in the gene nucleotide sequence as well as in the repeat organization. As shown in the present study, ankyrin repeat positioning in genes is highly variable since many ankyrin repeat and ankyrin cluster combinations are found while a wide array of additional domains exists that possibly confer novel functions. Taken together, this abundance and variability of ankyrin genes present in *Wolbachia* raises the following questions: (a) why are there so many, (b) what is their origin and (c) what is their function.

The comparative analysis of the ANK repeat consensus of *Wolbachia* and the other prokaryotic, eukaryotic and viral organisms of the present study resulted to a rather interesting finding: *Wolbachia* ANK repeat consensus is more similar to that of eukaryotes than to that of prokaryotes. It is also

known that the ANK repeat plays an important role in protein-protein interactions and regulates transcription, development, cell cycle, signal transduction, interacts with cytoskeleton, cell membranes, acts as toxins or ion transporters (Michaely and Bennett 1992; Bork, 1993; Heringa 1998; Sedgwick and Smerdon, 1999; Mosavi *et al.* 2004). *Wolbachia* ANK proteins may thus provide all the necessary machinery to allow these bacteria to interact with eukaryotic hosts in order to establish symbiotic relationships of mutualistic or reproductive parasitism. *Wolbachia* ANK repeat consensus has all generally conserved positions (2, 4, 5, 6, 7, 9, 21, 25 and 26th) which include the TPLH motif (positions 4, 5, 6 and 7th). However, the TPLH motif was not conserved among the *Rickettsia* species. More specifically, positions 5 (usually occupied by Pro) and 7 (usually occupied by His) are not conserved in the general ANK repeat consensus calculated from all 10 *Rickettsia* species examined here. Looking at each species separately, only three have an adequate number of canonical ANK repeats for which the ANK repeat consensus can be calculated; *R. felis* and the two *R. bellii* strains.

What is interesting about these ANK repeat consensuses is that *R. felis* appears to have position 7, conserved (41.3% His), in contrast to *R. bellii* strains that have percentages around 20%. Whether this variability has to do with the host specificity (adaptation) has not yet been investigated. It is worth noting, however, that *R. bellii* is mainly found in ticks (not an insect) whereas *R. felis* in cat fleas (an insect). It is noteworthy that the giant Mimivirus, which infects the amoeba *Acanthamoeba polyphaga*, has a unique ANK repeat consensus compared to other viruses, prokaryotes or eukaryotes. The biological role of Mimivirus' ankyrins has not been investigated as yet. It has been reported that isolated ANK repeats (1-repeat ANK clusters) are probably not functional (Becerra *et al.* 2004; Mosavi *et al.* 2002); however, all *Wolbachia* strains studied contain a large number of isolated ANK repeats ranging from 33% in *wUni* to 69% in *wBm* as shown in Supplementary Table 3 (Data about the other Rickettsiales species will be discussed too). Actually, 9 out of a total 13 ANK repeat clusters in *wBm* are present as 1-domain ANK clusters. Interestingly, most of the isolated ANK repeats are not canonical (*i.e.* they are not exactly 33 amino acids long). Almost fifty per cent of the isolated repeats (69/148) have been the target of insertion events compared to 157/484 (32%) in the outer repeats and to 157/539 (29%) in the inner repeats of ANK clusters. The question is whether ANK proteins containing isolated ANK repeats are functional. Actually, ANK proteins containing (exclusively or not) isolated ANK repeats are expressed in diverse *Wolbachia* strains (Iturbe-Ormaetxe *et al.* 2005; Klasson *et al.* 2008a; Walker *et al.* 2007; unpublished data). Another potential role of the isolated repeats could be as a reservoir for future shuffling and/or recombination events to further increase the great diversity of the *Wolbachia* ankyrins. It cannot be excluded that the isolated repeats are in the process of decay; however, the fact that they are not the "primary target" for insertion events and they are expressed (at least at the transcriptional level) makes the scenario that these ANK

proteins are in the process of pseudogenization highly unlikely. Another interesting observation has to do with the length of an ANK repeat cluster and its similarity with known ANK proteins in the databases. ANK proteins with lengthier repeat clusters are more similar to eukaryotic ANK proteins with known functions. The present study provides for first time evidence about how the great diversity of *Wolbachia* ankyrins could result. Many insertion and deletion events were detected in the 287 *Wolbachia* ankyrins studied, which resulted in the “production” of 751 canonical vs. 443 non-canonical ANK repeats. The thorough analysis of these events resulted in the following conclusions: (a) the majority of these events have occurred in the chromosomal and not in the phage regions, (b) most insertion events occurred in isolated repeats followed by the inner repeats of a cluster and (c) in all cases, the insertion events took place in only two positions: about 50% of them occurred between positions 13 and 14, and the other 50% occurred between positions 23 and 24.

Such events, which have been reported in the past for eukaryotic ankyrins (Sedgwick and Smerdon 1999) increase the genetic diversity of *Wolbachia* ankyrins without affecting the overall structure of the repeat. The inserted amino acids are lengthier between positions 13 and 14 (between the two alpha-helices of the motif), however, none of the insertions present any similarity with proteins (or peptides) in the databases. It should also be noted that the currently available data suggest that ANK proteins containing non-canonical repeats are expressed (at least at the transcriptional level). Ten per cent of the total ANK proteins studied (27 out of 287) contain multiple clusters (*i.e.* more than one ANK cluster). This is an unusually high number and the question is how these ANK proteins with multiple ANK clusters have evolved. The phylogenetic analysis of the present study suggested that these ANK proteins could be the result of recombination and/or shuffling events. Several studies have shown that *Wolbachia* is a highly recombining bacterial community (Klasson *et al.* 2008a, b; Baldo *et al.* 2006). It is also known that *Wolbachia* have an active phage (Masui *et al.* 2001; Fujii *et al.* 2004; Bordenstein *et al.* 2006), which could promote exchange of genomic DNA including ankyrin genes (Bordenstein and Wernegreen 2004). In addition, recent studies have suggested that some ankyrins are fast evolving (Iturbe-Ormaetxe *et al.* 2005, Sinkins *et al.* 2005, Klasson *et al.* 2008). Taken together, this evidence suggests that there is a large repertoire of molecular mechanisms which could be responsible for the great genetic diversity observed in *Wolbachia* ANK proteins.

The presence of additional domains in *Wolbachia* ANK genes provides a wealth of information about the potential biological function of these proteins (Supplementary Figure 3; Supplementary Table 4). The most significant domains detected included the transmembrane (TMHelix), ATPase family protein Associated with various cellular Activities (AAA), tetratricopeptide repeat (TPR), C-terminal Pox viral (PRANC), DnaJ, ovarian tumor (OTU), osmosensory coiled coil (Osmo_CC) and

half-A TPR (HAT) domain and secondarily, the Streptococcal surface antigen (Strep_SA_rep), PEST motifs (Tash_PEST), Pox virus A-type inclusions (Pox_A_type_inc) and M protein repeat (M) domain. It is known that transmembrane proteins include numerous receptors, channels, transporters, photosystems and respiratory complexes (Lomize *et al.*, 2007). AAA ATPase is a domain present in proteins performing chaperone-like functions (Confalonieri and Duguet 1995) and exerting their activity through the energy-dependent unfolding of macromolecules (Iyer *et al.*, 2004). They are involved in protein degradation, membrane fusion, microtubule severing, peroxisome biogenesis, signal transduction and the regulation of gene expression (Iyer *et al.*, 2004). The TPR domain is present in a wide range of proteins (Lamb *et al.*, 1995; Goebel and Yanagida, 1991; Das *et al.*, 1998) and organisms (<http://pfam.janelia.org/family?acc=PF00515>).

This domain mediates protein-protein interactions (D'Andrea and Regan, 2003) and has been implicated to play a role in cell cycle regulation, transcriptional control, mitochondrial and peroxisomal protein transport, neurogenesis and protein folding. It has recently been reported that the genome of *Orientia tsutsugamushi* contains a large number of TPR-containing proteins (Cho *et al.* 2007; Nakayama *et al.* 2008). The Pox proteins Repeats of ANkyrin – C terminal (PRANC) domain is usually present at the C-terminus of a variety of Pox virus proteins. Although its function is unclear, it appears to be related to the F-box domain which mediates protein-protein interactions (Mercer *et al.*, 2005; Skowrya *et al.*, 1997). Two *Wolbachia* genes contain the PRANC domain, *wAna21* and *wRi_ANK24*, which are orthologs and are part of a prophage element. The DnaJ or “J” domain is found in the prokaryotic DnaJ heat shock protein and more specifically, at its N-terminus. The J domain is thought to mediate the interaction between DnaJ and DnaK proteins, the latter being a hsp70-like chaperone (Cyr *et al.*, 1994; Kelley, 1998). The presence of J domains in 3 *wPip* ANK genes raises the question whether some *Wolbachia* ANKs may possess chaperone activities. The Ovarian Tumor (OTU) domain is present in a family of proteins found in eukaryotes, viruses and the pathogenic bacterium *Chlamydia pneumoniae*. Although of unknown function, specific amino acids in this domain suggest it could have cysteine protease activity (Makarova *et al.*, 2000). Two genes contain an OTU domain; WP_0462 from *wPip* which is a phage gene and *wRi_ANK21* which is chromosomal gene. The OTU domain is also present in 3 ANK genes of *A. gambiae*, *D. melanogaster* and *N. vitripennis*. The function of the Osmosensory transporter coiled coil (Osmo_CC) is to sense increases in osmotic pressure and respond by changing accordingly the concentration of organic solutes in the cell (Bolen, 2001). There is only one *Wolbachia* gene harboring this domain, WP_0346 (a phage-encoded gene). The Half-A TPR (HAT) repeats are found in several RNA processing proteins as components of larger, macromolecular complexes (Preker and Keller, 1998). They look very much like TPR repeats both in structure and sequence although they lack two highly conserved residues found

in the latter. The Streptococcal surface antigen repeat (Strep_SA_rep) is commonly found in Streptococcal surface antigens although it has been found in other pathogens as well (<http://pfam.janelia.org/family?acc=PF06696>). The Tash protein PEST motif (Tash_PEST) has been detected in *Theileria annulata* proteins (a protozoan parasite that causes fatal diseases in cattle).

These proteins have been found to contain a DNA-binding motif and nuclear localization sequences that allow them to enter into the host nucleus and possibly alter the expression profile of certain cytoskeletal proteins (Shiels *et al.*, 2004; Swan *et al.*, 2001). They are also suggested as being part of the PEST motif (which acts as a signal for rapid proteolytic self-degradation) although this has not been confirmed (Shiels *et al.*, 2004). The only gene having a Tash_PEST motif is WP_0774 of *wPip* strain. The Pox A type inclusion protein repeat (Pox_A_type_inc) has been reported in the corresponding protein of the Poxvirus family of viruses (Funahashi *et al.*, 1988). A-type inclusion protein is one of the most abundant viral proteins and it forms the A-type inclusions which sequester and protect the mature viruses during their transfer to a new host (Funahashi *et al.*, 1988). The M protein repeat is a short repeat (21 amino acids) present in bacterial M proteins. M proteins are immunoglobulin – binding proteins that confer phagocytosis resistance to the bacterial pathogen (Podbielski *et al.*, 1996). Five genes with M protein repeats are found in *Wolbachia*; *wAna23*, *WD_0073*, *wRi_ANK03*, *wSim9* and *wUni_1_1307*. All 5 of them are not related with a prophage element and are encoded in the bacterial chromosome. Contrary to the literature there is only one repeat per gene and not multiple (Podbielski *et al.*, 1996). Their e-values are 4.8-4.9 but this large e-value may be due to the repeat's short length. All of them additionally contain one or two transmembrane domains and their length ranges from 350 to 800 aminoacids. The M protein repeats are located towards the C-terminus of the gene and in all cases before the transmembrane domain(s).

The ankyrin gene sets of the other organisms of this study were also searched for the presence of the additional domains found in *Wolbachia*. Interestingly, it was found that only AAA ATPase, TPR, OTU and DnaJ domains were present together with ankyrin domains and only in eukaryotic organisms. The only exception to this rule is the presence of an AAA domain in a *C. burnetii* ankyrin gene. Another AAA ATPase domain is found in the ankyrin gene set of the plant species *A. thaliana*. TPR repeats are found in all 7 eukaryotes included in this study in a varying number of genes per organism (from 2 to 12) and with varying significance (*i.e.* e-value). As mentioned above, the OTU domain was found in *D. melanogaster*, *N. vitripennis* and *A. gambiae* while the DnaJ domain was only found in a gene of *A. thaliana*. In average, the TPR, AAA and DnaJ domains found in *Wolbachia* have a better e-value than those found in the eukaryotic ankyrin genes while the reverse is true for OTU domains which suggests that these domains may be functional in *Wolbachia*. The search for additional domains in ANK genes of all Rickettsiales species used in the present study indicated the

transmembrane helix as the most common additional motif. Interestingly, transmembrane helices in ANK genes were only found in genera *Anaplasma*, *Ehrlichia* and *Neorickettsia* (not in genera *Rickettsia* and *Orientia*). In other words, ANK proteins with transmembrane helices are present in the family *Anaplasmataceae*, and not in the family *Rickettsiaceae*. The second most common motif was PRANC domain which was shown to be frequently present in both *Orientia tsutsugamushi* strains (8 genes in the Boryong and 17 genes in the Ikeda strain). The most interesting additional domain found in *Rickettsia* species is the RelA/SpoT, which is present in a single gene of *Rickettsia felis*. Different algorithms were used for the prediction of potentially secreted proteins.

The analysis showed that a significant number of *Wolbachia* ANK proteins are probably secreted through the sec-dependent pathway. There is also a significant number of proteins which are secreted by using a non sec-dependent pathway. Interestingly, this study identified *x* *Wolbachia* ANK proteins which probably secreted in host cells by the T4SS machinery. This finding is in accordance with recent observations that ANK proteins are T4SS effector molecules in bacterial pathogens such as *Legionella pneumonophila* and *Coxiella burnetii* (Habyarimana *et al.* 2008; Pan *et al.* 2008). The BLAST analysis provided clear evidence that the majority of the *Wolbachia* ANK genes studied seem to be eukaryotic-like confirming, at a larger scale, previous studies (Masui *et al.* 2000; MacGraw and O'Neill, 2004). The analysis indicated that the “eukaryotic signal” is located in the ankyrin domain clusters rather than in the rest of the sequence (Figure 8). In addition, it was shown that best hits of the BLASTP analysis were against eukaryotic proteins such as unc44 ankyrins, kinase proteins, ion channels, Pfs/NACHT domain proteins and Cdk inhibitors or against the prokaryotic ppGpp pyrophosphohydrolases/synthetases. The most frequently occurring BLAST hits were unc44 ankyrins and ppGpp pyrophosphohydrolase/synthase (RelA/SpoT) (Supplementary Table 2A). The eukaryotic unc44 ankyrins are responsible, in *C. elegans*, for proper axonal outgrowth and guidance (Otsuka *et al.* 1995). Mutations can lead to locomotory defects (Otsuka *et al.*, 2002). In *Wolbachia*, several ankyrin genes present similarity with unc44 ankyrins; 115 and 74 genes have an unc44 ankyrin as a significant hit when BLASTing whole genes and clusters, respectively. None of the leftover fragments matches an unc44 ankyrin. Several *Wolbachia* ANK genes also present significant similarity, although not as best hit, with the prokaryotic enzyme Guanosine polyphosphate (ppGpp) pyrophosphohydrolase/synthetase when BLASTing whole *Wolbachia* ANK genes or repeat clusters.

None of the leftover fragments has similarity to ppGpp. This enzyme is directly implicated in the stringent response, a survival strategy found in a bacterium which is activated during nutrient starvation. ppGpp accumulation is the key event in this strategy and its synthesis is accomplished by this enzyme. Accumulated ppGpp subsequently binds to RNA polymerase in order to regulate the expression of certain genes, DNA replication and virulence induction (Cashel, 1974; Srivatsan and

Wang, 2008). Another possible role for such ppGpp-recognizing proteins was reported recently by Zhao and colleagues (2008) who showed that the insertion specificity for a mobile group II intron hosted in the *E. coli* chromosome is determined by ppGpp concentration. More specifically, while this specific group II intron preferentially inserts at the origin of replication (ori) region, mutations in genes participating in the control of polyphosphate levels, resulted in insertion of this mobile element in chromosomal positions other than ori (Zhao *et al.*, 2008). Whether *Wolbachia* ANK genes exhibiting similarity with the ppGpp homologs present the same properties remains to be investigated.

However, it is worth noting that *Wolbachia* can survive for long periods outside host cells (Rasgon *et al.* 2006). In addition, it is known that the genome of arthropod *Wolbachia* strains contains many mobile elements including group II introns (Bordenstein and Reznikoff, 2005; Yamada *et al.* 2007). The BLASTP analysis also indicated that few *Wolbachia* ANK proteins presented best hits with eukaryotic serine/threonine protein kinases such as integrin-linked kinases (ILK) and cardiac troponin I interacting (TNNI3) kinases which are distantly related to the former. Integrin mediates communication between cells by facilitating signal transduction between them and regulating diverse cell properties such as shape, motility, growth, survival, differentiation and gene transcription (Hannigan *et al.*, 1996). ILKs were found to phosphorylate certain integrin cytoplasmic domains thereby influencing integrin-mediated signal transduction (Hannigan *et al.*, 1996). TNNI3 kinases on the other hand are implicated in proper cardiac troponin phosphorylation which is a prerequisite for the correct functioning of the heart (Luft, 2003; Haworth *et al.*, 2004).

Several *Wolbachia* ANK genes also present significant similarity, although not as best hit, with proteins containing Pfs, NACHT and ankyrin domain. All, but one, of these proteins are of eukaryotic origin. Pfs (nucleoside phosphorylase) domains have been suggested to play a regulatory or signaling role in *Aspergillus* (Fedorova *et al.*, 2005). Indirectly, it is involved in proliferation, differentiation and apoptosis (Fedorova *et al.*, 2005). NACHT is a NTPase domain found in apoptosis proteins and proteins involved in MHC transcription activation. It has recently been shown that *Wolbachia* manipulates the apoptotic pathways of their hosts in order to establish symbiotic associations (Bazzocchi *et al.* 2007; Pannebakker *et al.* 2007). *Wolbachia* ANK proteins also exhibit significant similarity with eukaryotic Ion channels. Ion channels are pore-forming proteins that establish and control voltage gradients across biological membranes). Most of the hits are either cation (most often potassium) or non-potassium (TRPN) channels. The latter are found in mechanosensory organs of animals and mutations to these ion channels is, for example, linked to mechanosensing defects in *Drosophila* and auditory defects in zebrafish (Pedersen *et al.*, 2005). The BLASTP analysis also indicated that at least one *Wolbachia* ANK gene, *wVulC_ANK63* presents high similarity with eukaryotic Cyclin dependent kinase (Cdk) inhibitor. Cyclin dependent kinases are master regulators of cell proliferation that

recognize specific environmental signals and progress or halt cell cycle. Cdk inhibitors, as their name implies, inhibit Cdk activity in response to environmental signals (Blain, 2008). It has recently been suggested that inhibition of Cdk may be involved in the expression of *Wolbachia*-induced reproductive phenotypes such as cytoplasmic incompatibility (Tram and Sullivan 2002; Tram *et al.* 2003; Harris and Braig 2003).

It has been suggested that ANK proteins may play a significant role in *Wolbachia*-host interactions including the establishment of the symbiotic relationship, coping with various environmental stresses and / or the expression of reproductive abnormalities by which *Wolbachia* achieves spreading in naïve host populations (McGraw and O'Neill 2004). Several studies have documented the involvement of ANK proteins in host-bacterial interactions. AnkA, an ankyrin-repeat protein of *Anaplasma phagocytophilum* has been shown to bind chromatin (Caturegli *et al.*, 2000). The AnkA protein contains 11 ankyrin repeats, no transmembrane domain, several kinase phosphorylation sites (tyrosine kinase, protein kinase C and casein kinase) and a N-glycosylation site (Caturegli *et al.*, 2000). Park *et al.* (2004) confirmed AnkA's nuclear localization and also showed that it binds to a variety of human proteins such as ATPase, tyrosine phosphatase and NADH dehydrogenase-like protein. The precise DNA-binding mechanism, though, as well as its effect is not clear. The genome of the bacterial pathogen *Pseudomonas aeruginosa* also contains ANK proteins. One of the ANK genes, ankB, is part of a hydrogen peroxide-induced operon. Interestingly, AnkB is an integral membrane protein that has its ankyrin repeats in the periplasm. It has been shown that the presence of this protein confers hydrogen peroxide resistance in the bacterium (Howell *et al.*, 2000). In a recent elegant study in the plant *Lotus japonicus* it was shown that an ankyrin repeat gene, named ign1, is crucial for development of a nitrogen-fixing symbiotic relationship between *L. japonicus* and *Mesorhizobium loti*; however, the ign1 gene is not hosted by the bacterium, but by the plant (Kumagai *et al.*, 2007).

Recent studies have shown that bacterial species of the genera *Ehrlichia*, *Anaplasma* and *Rickettsia*, which are close relatives of *Wolbachia*, as well as *Legionella pneumophila* and *Coxiella burnetii* contain ANK genes which are secreted to host cells through T4SS and play an important role in host-bacterial interactions (Caturegli *et al.* 2000; Park *et al.* 2004; Lin *et al.* 2007; Seshadri *et al.* 2003; Habyarimana *et al.* 2008; Pan *et al.* 2008). Whether *Wolbachia* ANK genes are also secreted through T4SS and are involved in the establishment of symbiotic associations and the expression of reproductive abnormalities awaits for future experimental confirmation.

Conclusions

The *in silico* analysis of complete and partial genomes of diverse *Wolbachia* strains confirmed the presence of a large number and of great diversity ANK genes. The present study indicated potential mechanisms generating the genetic diversity of ANK genes. In addition, the comparative analysis undertaken suggested that the majority of *Wolbachia* ANK genes are of eukaryotic origin. ANK repeats along with a wealth of additional domains present in these genes clearly suggest that ANK proteins may be a key factor in the establishment of *Wolbachia* symbiosis and the induction of host reproductive manipulations.

Acknowledgements

This work was supported by EUWOL. It was also partially supported by EU CSA-REGPROT 203590 - MicrobeGR and by intramural funds of the University of Ioannina to KB. The authors thank Dr. George Markakis for his valuable help with statistics.

References

- Altschul S, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389 - 3402
- Anderson I, Rodriguez J, Susanti D, Porat I, Reich C, Ulrich LE, Elkins JG, Mavromatis K, Lykidis A, Kim E, Thompson LS, Nolan M, Land M, Copeland A, Lapidus A, Lucas S, Detter C, Zhulin IB, Olsen GJ, Whitman W, Mukhopadhyay B, Bristow J, Kyrpides N. (2008) Genome sequence of *Thermophilum pendens* reveals an exceptional loss of biosynthetic pathways without genome reduction. *J Bacteriol* 190: 2957 - 2965
- Baldo L, Bordenstein S, Wernegreen JJ, Werren JH. (2006) Widespread recombination throughout *Wolbachia* genomes. *Mol Biol Evol* 23: 437 - 449
- Becerra C, Jahrman T, Puigdomènech P, Vicent CM. (2004) Ankyrin repeat-containing proteins in *Arabidopsis*: characterization of a novel and abundant group of genes coding ankyrin-transmembrane proteins. *Gene* 340: 111 - 121
- Blain S (2008) Switching cyclin D-Cdk4 kinase activity on and off. *Cell Cycle* 7: 892 - 898
- Bolen D (2001) Protein stabilization by naturally occurring osmolytes. *Methods Mol Biol* 168: 17 - 36

- Bork P (1993) Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins* 17: 363 - 374
- Bourtzis K, Nirgianaki A, Markakis G, Savakis C (1996) *Wolbachia* infection and cytoplasmic incompatibility in *Drosophila* species. *Genetics* 144: 1063-1073
- Breeden L, Nasmyth K (1987) Similarity between cell-cycle genes of budding yeast and fission yeast and the Notch gene of *Drosophila*. *Nature* 329: 651 - 654
- Cashel M (1974) Preparation of guanosine tetraphosphate (ppGpp) and guanosine pentaphosphate (pppGpp) from *Escherichia coli* ribosomes. *Anal Biochem* 57: 100 - 107
- Cashel M (1974) Preparation of guanosine tetraphosphate (ppGpp) and guanosine pentaphosphate (pppGpp) from *Escherichia coli* ribosomes. *Anal Biochem* 57: 100 - 107
- Caturegli P, Asanovich KM, Walls JJ, Bakken JS, Madigan JE, Popov VL, Dumler JS (2000) ankA: an *Ehrlichia phagocytophila* group gene encoding a cytoplasmic protein antigen with ankyrin repeats. *Infect Immun* 68: 5277-5283
- Cho N, Kim HR, Lee JH, Kim SY, Kim J, Cha S, Kim SY, Darby AC, Fuxelius HH, Yin J, Kim JH, Kim J, Lee SJ, Koh YS, Jang WJ, Park KH, Andersson SG, Choi MS, Kim IS. (2007) The *Orientia tsutsugamushi* genome reveals massive proliferation of conjugative type IV secretion system and host-cell interaction genes. *Proc Natl Acad Sci U S A* 104: 7981 - 7986
- Confalonieri F, Duguet M (1995) A 200-amino acid ATPase module in search of a basic function. *BioEssays* 17: 639 - 650
- Cyr D, Langer T, Douglas MG. (1994) DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. *Trends Biochem Sci* 19: 176 - 181
- D'Andrea L, Regan L. (2003) TPR proteins: the versatile helix. *Trends Biochem Sci* 28: 655 - 662
- Das A, Cohen PW, Barford D. (1998) The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J* 17: 1192 - 1199
- Delcher A, Bratke KA, Powers EC, Salzberg SL. (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23: 673 - 679
- Duron O, Boureux A, Echaubard P, Berthomieu A, Berticat C, Fort P, Weill M. (2007) Variability and expression of ankyrin domain genes in *Wolbachia* variants infecting the mosquito *Culex pipiens*. *J Bacteriol* 189: 4442 - 4448
- Fedorova N, Badger JH, Robson GD, Wortman JR, Nierman WC. (2005) Comparative analysis of programmed cell death pathways in filamentous fungi. *BMC Genomics* 6: 177
- Félix C, Pichon S, Braquart-Varnier C, Braig H, Chen L, Garrett RA, Martin G, Grève P. (2008) Characterization and transcriptional analysis of two gene clusters for type IV secretion machinery in *Wolbachia* of *Armadillidium vulgare*. *Res Microbiol* 159: 481 - 485

- Felsenstein J (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 5: 164 - 166
- Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, Bhattacharyya A, Kapatral V, Kumar S, Posfai J, Vincze T, Ingram J, Moran L, Lapidus A, Omelchenko M, Kyrpides N, Ghedin E, Wang S, Goltsman E, Joukov V, Ostrovskaya O, Tsukerman K, Mazur M, Comb D, Koonin E, Slatko B. (2005) The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol* 3: e121
- Funahashi S, Sato T, Shida H. (1988) Cloning and characterization of the gene encoding the major protein of the A-type inclusion body of cowpox virus. *J Gen Virol* 69: 35 - 47
- Goebel M, Yanagida M (1991) The TPR snap helix: a novel protein repeat motif from mitosis to transcription. *Trends Biochem Sci* 16: 173 - 177
- Habyarimana F, Al-Khodori S, Kalia A, Graham JE, Price CT, Garcia MT, Kwaik YA. (2008) Role for the Ankyrin eukaryotic-like genes of *Legionella pneumophila* in parasitism of protozoan hosts and human macrophages. *Environ Microbiol* 10: 1460 - 1474
- Hannigan G, Leung-Hagesteijn C, Fitz-Gibbon L, Coppolino MG, Radeva G, Filmus J, Bell JC, Dedhar S. (1996) Regulation of cell adhesion and anchorage-dependent growth by a new beta 1-integrin-linked protein kinase. *Nature* 379: 91 - 96
- Harris H, Braig HR. (2003) Sperm chromatin remodelling and *Wolbachia*-induced cytoplasmic incompatibility in *Drosophila*. *Biochem Cell Biol* 81: 229 - 240
- Haworth R, Cuello F, Herron TJ, Franzen G, Kentish JC, Gautel M, Avkiran M. (2004) Protein kinase D is a novel mediator of cardiac troponin I phosphorylation and regulates myofilament function. *Circ Res* 95: 1091 - 1099
- Heringa J (1998) Detection of internal repeats: how common are they? *Curr Opin Struct Biol* 8: 338 - 345
- Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren J (2008) How many species are infected with *Wolbachia*?--A statistical analysis of current data. *FEMS Microbiol Lett* 281: 215 - 220
- Hoffmann AA, M. Turelli and G.M. Simmons (1986) Unidirectional incompatibility between populations of *Drosophila simulans*. *Evolution* 40: 692-701
- Hoffmann AA (1988) Partial cytoplasmic incompatibility between two Australian populations of *Drosophila melanogaster*. *Entomol Exp Appl* 48: 61-67
- Howell M, Alsabbagh E, Ma JF, Ochsner UA, Klotz MG, Beveridge TJ, Blumenthal KM, Niederhoffer EC, Morris RE, Needham D, Dean GE, Wani MA, Hassett DJ. (2000) AnkB, a periplasmic

ankyrin-like protein in *Pseudomonas aeruginosa*, is required for optimal catalase B (KatB) activity and resistance to hydrogen peroxide. *J Bacteriol* 182: 4545 - 4556

Iturbe-Ormaetxe I, Burke GR, Riegler M, O'Neill SL. (2005) Distribution, expression, and motif variability of ankyrin domain genes in *Wolbachia pipientis*. *J Bacteriol* 187: 5136 - 5145

Iyer L, Leipe DD, Koonin EV, Aravind L. (2004) Evolutionary history and higher order classification of AAA+ ATPases. *J Struct Biol* 146: 11 - 31

Kelley W (1998) The J-domain family and the recruitment of chaperone power. *Trends Biochem Sci* 23: 222 - 227

Klasson L, Walker T, Sebahia M, Sanders MJ, Quail MA, Lord A, Sanders S, Earl J, O'Neill SL, Thomson N, Sinkins SP, Parkhill J. (2008) Genome evolution of *Wolbachia* strain *wPip* from the *Culex pipiens* group. *Mol Biol Evol* 25: 1877 - 1887

Klasson L, Westberg J, Sapountzis P, Näslund K, Lutnaes Y, Darby AC, Veneti Z, Chen L, Braig HR, Garrett R, Bourtzis K, Andersson SGE (2008) The mosaic genome structure of the *Wolbachia wRi* strain infecting *Drosophila simulans*. *Proc Natl Acad Sci U S A*: Submitted

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305: 567 - 580

Kumagai H, Hakoyama T, Umehara Y, Sato S, Kaneko T, Tabata S, Kouchi H. (2007) A novel ankyrin-repeat membrane protein, IGN1, is required for persistence of nitrogen-fixing symbiosis in root nodules of *Lotus japonicus*. *Plant Physiol* 143: 1293 - 1305

Lamb J, Tugendreich S, Hieter P. (1995) Tetratricopeptide repeat interactions: to TPR or not to TPR? *Trends Biochem Sci* 20: 257 - 259

Lin M, den Dulk-Ras A, Hooykaas PJ, Rikihisa Y. (2007) *Anaplasma phagocytophilum* AnkA secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection. *Cell Microbiol* 9: 2644 - 2657

Lomize A, Pogozheva ID, Lomize MA, Mosberg HI. (2007) The role of hydrophobic interactions in positioning of peripheral proteins in membranes. *BMC Struct Biol* 7: 44

Luft F (2003) Hearts of this ILK rely on TNNI3K, a MAPKKK that regulated TNNI3. *J Mol Med* 81: 279 - 280

Madan A, Lee H, Madan A, Yoon J-G, Ryu G-Y, Dasch G, Ereemeva M Complete genome sequencing of *Rickettsia bellii*. In Prep

Makarova K, Aravind L, Koonin EV. (2000) A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and *Chlamydia pneumoniae*. *Trends Biochem Sci* 25: 50 - 52

Masui S, Kamoda S, Sasaki T, Ishikawa H (2000) Distribution and evolution of bacteriophage WO in *Wolbachia*, the endosymbiont causing sexual alterations in arthropods. *J Mol Evol* 51: 491 - 497

- McGraw E, O'Neill SL. (2004) *Wolbachia pipientis*: intracellular infection and pathogenesis in *Drosophila*. *Curr Opin Microbiol* 7: 67 - 70
- Mercer A, Fleming SB, Ueda N. (2005) F-box-like domains are present in most poxvirus ankyrin repeat proteins. *Virus Genes* 31: 127 - 133
- Michaely P, Bennett V (1992) The ANK repeat: a ubiquitous motif involved in macromolecular recognition. *Trends Cell Biol* 2: 127 - 129
- Miller W, Riegler M. (2006) Evolutionary dynamics of wAu-like *Wolbachia* variants in neotropical *Drosophila* spp. *Appl Environ Microbiol* 72: 826 - 835
- Mosavi L, Minor DL Jr, Peng ZY. (2002) Consensus-derived structural determinants of the ankyrin repeat motif. *Proc Natl Acad Sci U S A* 99: 16029 - 16034
- Mosavi L, Cammett T, Desrosiers D, Peng Z (2004) The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci* 13: 1435 - 1448
- Nakayama K, Yamashita A, Kurokawa K, Morimoto T, Ogawa M, Fukuhara M, Urakami H, Ohnishi M, Uchiyama I, Ogura Y, Ooka T, Oshima K, Tamura A, Hattori M, Hayashi T. (2008) The Whole-genome sequencing of the obligate intracellular bacterium *Orientia tsutsugamushi* revealed massive gene amplification during reductive genome evolution. *DNA Res* 15: 185 - 199
- Ogata H, Renesto P, Audic S, Robert C, Blanc G, Fournier PE, Parinello H, Claverie JM, Raoult D. (2005) The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biol* 3: e248
- Ogata H, La Scola B, Audic S, Renesto P, Blanc G, Robert C, Fournier PE, Claverie JM, Raoult D. (2006) Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genet* 2: e76
- Otsuka A, Franco R, Yang B, Shim KH, Tang LZ, Zhang YY, Boontrakulpoontawee P, Jeyaprakash A, Hedgecock E, Wheaton VI, Sobery A (1995) An ankyrin-related gene (unc-44) is necessary for proper axonal guidance in *Caenorhabditis elegans*. *J Cell Biol* 129: 1081 - 1092
- Otsuka A, Boontrakulpoontawee P, Rebeiz N, Domanus M, Otsuka D, Velamparampil N, Chan S, Vande Wyngaerde M, Campagna S, Cox A. (2002) Novel UNC-44 AO13 ankyrin is required for axonal guidance in *C. elegans*, contains six highly repetitive STEP blocks separated by seven potential transmembrane domains, and is localized to neuronal processes and the periphery of neural cell bodies. *J Neurobiol* 50: 333 - 349
- Page R (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12: 357 - 358
- Pan X, Lührmann A, Satoh A, Laskowski-Arce MA, Roy CR. (2008) Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* 320: 1651 - 1654

- Pannebakker B, Loppin B, Elemans CP, Humblot L, Vavre F. (2007) Parasitic inhibition of cell death facilitates symbiosis. *Proc Natl Acad Sci U S A* 104: 213 - 215
- Park J, Kim KJ, Choi KS, Grab DJ, Dumler JS. (2004) *Anaplasma phagocytophilum* AnkA binds to granulocyte DNA and nuclear proteins. *Cell Microbiol* 6: 743 - 751
- Pedersen S, Owsianik G, Nilius B. (2005) TRP channels: an overview. *Cell Calcium* 38: 233 - 252
- Podbielski A, Schnitzler N, Beyhs P, Boyle MD. (1996) M-related protein (Mrp) contributes to group A streptococcal resistance to phagocytosis by human granulocytes. *Mol Microbiol* 19: 429 - 441
- Preker P, Keller W. (1998) The HAT helix, a repetitive motif implicated in RNA processing. *Trends Biochem Sci* 23: 15 - 16
- Rancès E, Voronin D, Tran-Van V, Mavingui P. (2008) Genetic and functional characterization of the type IV secretion system in *Wolbachia*. *J Bacteriol* 190: 5020 - 5030
- Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie JM. (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* 306: 1344 - 1350
- Rousset F, D. Bouchon, B. Pintureau, P. Juchault and M. Solignac (1992) *Wolbachia* endosymbionts responsible for various alterations of sexuality in arthropods. *Proceedings of Royal Society of London Biological Series* 250: 91-98
- Salzberg S, Dunning Hotopp JC, Delcher AL, Pop M, Smith DR, Eisen MB, Nelson WC. (2005) Correction: Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. *Genome Biol* 6: 402
- Salzberg S, Hotopp JC, Delcher AL, Pop M, Smith DR, Eisen MB, Nelson WC. (2005) Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. *Genome Biol* 6: R23
- Sedgwick SG, Smerdon SJ (1999) The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem Sci* 24: 311-316
- Seshadri R, Paulsen IT, Eisen JA, Read TD, Nelson KE, Nelson WC, Ward NL, Tettelin H, Davidsen TM, Beanan MJ, Deboy RT, Daugherty SC, Brinkac LM, Madupu R, Dodson RJ, Khouri HM, Lee KH, Carty HA, Scanlan D, Heinzen RA, Thompson HA, Samuel JE, Fraser CM, Heidelberg JF (2003) Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proc Natl Acad Sci U S A* 100: 5455-5460
- Shiels B, McKellar S, Katzer F, Lyons K, Kinnaird J, Ward C, Wastling JM, Swan D. (2004) A *Theileria annulata* DNA binding protein localized to the host cell nucleus alters the phenotype of a bovine macrophage cell line. *Eukaryot Cell* 3: 495 - 505
- Sinkins S, Walker T, Lynd AR, Steven AR, Makepeace BL, Godfray HC, Parkhill J. (2005) *Wolbachia* variability and host effects on crossing type in *Culex* mosquitoes. *Nature* 436

- Sironi M, Bandi C, Sacchi L, Di Sacco B, Damiani G, Genchi C. (1995) Molecular evidence for a close relative of the arthropod endosymbiont *Wolbachia* in a filarial worm. *Mol Biochem Parasitol* 74: 223 - 227
- Skowrya D, Craig KL, Tyers M, Elledge SJ, Harper JW. (1997) F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell* 91: 209 - 219
- Srivatsan A, Wang JD. (2008) Control of bacterial transcription, translation and replication by (p)ppGpp. *Curr Opin Microbiol* 11: 100 - 105
- Stouthamer R, R.F. Luck and W.D. Hamilton (1990) Antibiotics cause parthenogenetic *Trichogramma* (Hymenoptera/Trichogrammatidae) to revert to sex. *Proc Natl Acad Sci U S A* 87: 2424-2427
- Stouthamer R, Breeuwer J, Hurst G (1999) *Wolbachia pipientis*: Microbial manipulator of arthropod reproduction. *Annu Rev Microbiol* 53: 71 - 102
- Swan D, Stern R, McKellar S, Phillips K, Oura CA, Karagenc TI, Stadler L, Shiels BR. (2001) Characterisation of a cluster of genes encoding *Theileria annulata* AT hook DNA-binding proteins and evidence for localisation to the host cell nucleus. *J Cell Sci* 114: 2747 - 2754
- Swofford D (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. In S Sinauer Associates, Massachusetts, ed,
- Taylor M, Bandi C, Hoerauf A (2005) *Wolbachia* bacterial endosymbionts of filarial nematodes. *Adv Parasitol* 60: 245 - 284
- Thompson J, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876 - 4882
- Tram U, Sullivan W. (2002) Role of delayed nuclear envelope breakdown and mitosis in *Wolbachia*-induced cytoplasmic incompatibility. *Science* 296: 1124 - 1126
- Tram U, Ferree PM, Sullivan W. (2003) Identification of *Wolbachia*-host interacting factors through cytological analysis. *Microbes Infect* 5: 999 - 1011
- Tulman E, Afonso CL, Lu Z, Zsak L, Kutish GF, Rock DL. (2004) The genome of canarypox virus. *J Virol* 78: 353 - 566
- Walker T, Klasson L, Sebaihia M, Sanders MJ, Thomson NR, Parkhill J, Sinkins SP. (2007) Ankyrin repeat domain-encoding genes in the *wPip* strain of *Wolbachia* from the *Culex pipiens* group. *BMC Biol* 5: 39
- Werren J, Baldo L, Clark M (2008) *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol* 6: 741 - 751
- Werren JH (1997) Biology of *Wolbachia*. *Annu Rev Entomol* 42: 587-609

Wu M, Sun, LV, Vamathevan, J, Riegler, M, Deboy, R, Brownlie, JC, McGraw, EA, Martin, W, Esser, C, Ahmadinejad, N, Wiegand, Madupu, R, Beanan, MJ, Brinkac, LM, Daugherty, SC, Durkin, AS, Kolonay, JF, Nelson, WC, Mohamoud, Y, Lee, P, Berry, K, Young, MB, Utterback, T, Weidman, J, Nierman, WC, Paulsen, IT, Nelson, KE, Tettelin, H, O'Neill, SL, Eisen, JA (2004) Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol 2: e69

Zhao J, Niu W, Yao J, Mohr S, Marcotte EM, Lambowitz AM. (2008) Group II intron protein localization and insertion sites are affected by polyphosphate. PLoS Biol 6: e150.

Organism	References	Closed/C urated	GenBank	Genome size (Mb)	Number of total ANK genes (phage-encoded)	Ratio (nr of ANK proteins / Mb)	Number of typical ANK domains
<i>Wolbachia</i> w Mel	Wu <i>et al.</i> 2004	Y	NC_002978	1.27	28 (10)	22,0	74
<i>Wolbachia</i> w Bm	Foster <i>et al.</i> 2005	Y	NC_006833	1.08	10 (0)	9,3	12
<i>Wolbachia</i> w Ana	Salzberg <i>et al.</i> 2005	N	NZ_AAGB000000000	1.47	34 (15)	23,1	87
<i>Wolbachia</i> w Sim	Salzberg <i>et al.</i> 2005	N	NZ_AAGC000000000	1.25	13 (3)	10,4	24
<i>Wolbachia</i> w WIl	Salzberg <i>et al.</i> 2005	N	NZ_AAQP000000000	?	14 (8)	N/A	37
<i>Wolbachia</i> w Pip-Pel	Klasson <i>et al.</i> 2008	N	N/A	1.48	58 (19)	32,4	166
<i>Wolbachia</i> w Ri	Klasson <i>et al.</i> 2009	Y	N/A	1.44	38 (20)	26,4	132
<i>Wolbachia</i> w Uni	Klasson <i>et al.</i> 2009	N	N/A	?	16 (0)	N/A	31
<i>Wolbachia</i> w VuIC	-	N	N/A	1.75	76 (24)	44,7	188
<i>Anaplasma marginale</i> St Maries	Brayton <i>et al.</i> 2005	Y	NC_004842	1.2	4	3,3	4
<i>Anaplasma phagocytophilum</i> HZ	Hotopp <i>et al.</i> 2006	Y	NC_007797	1.47	4	2,7	11
<i>Ehrlichia canis</i> Jake	Mavromatis <i>et al.</i> 2006	Y	NC_007354	1.32	5	3,8	9
<i>Ehrlichia chaffeensis</i> Arkansas	Hotopp <i>et al.</i> 2006	Y	NC_007799	1.18	5	4,2	8
<i>Ehrlichia ruminantium</i> Gardel	Frutos <i>et al.</i> 2004	Y	NC_006831	1.5	6	4,0	8
<i>Ehrlichia ruminantium</i> Welgevonden	Collins <i>et al.</i> 2005	Y	NC_005295	1.52	6	3,9	7
<i>Neorickettsia senetsu</i> Miyayama	Hotopp <i>et al.</i> 2006	Y	NC_007798	0.86	3	3,5	6
<i>Rickettsia akari</i> Hartford	-	Y	NC_009881	1.23	4	3,3	3
<i>Rickettsia bellii</i> OSU 85-389	-	Y	NC_009883	1.52	29	19,1	43
<i>Rickettsia bellii</i> RML369-C	Ogata <i>et al.</i> 2006	Y	NC_007940	1.53	28	18,3	41
<i>Rickettsia canadensis</i> McKiel	-	Y	NC_009879	1.16	4	3,4	2
<i>Rickettsia conorii</i> Malish7	Ogata <i>et al.</i> 2000	Y	AE006914	1.27	7	5,5	4
<i>Rickettsia felis</i> URRWXCα2	Ogata <i>et al.</i> 2005	Y	NC_007109	1.49	25	16,8	46
<i>Rickettsia massiliae</i> MTU5	Blanc <i>et al.</i> 2007	Y	NC_009900	1.36	4	2,9	5
<i>Rickettsia prowazekii</i> MadridE	Andersson <i>et al.</i> 1998	Y	AJ235269	1.11	3	2,7	5
<i>Rickettsia rickettsii</i> Sheila Smith	-	Y	NC_009882	1.26	9	7,1	8
<i>Rickettsia typhi</i> Wilmington	McLeod <i>et al.</i> 2004	Y	AE017197	1.11	1	0,9	1
<i>Orientia tsutsugamushi</i> Boryong	Cho <i>et al.</i> 2007	Y	NC_009488	2.13	39	18,3	74
<i>Orientia tsutsugamushi</i> Ikeda	Nakayama <i>et al.</i> 2008	Y	NC_010793	2.01	48	23,9	97
<i>Coxiella burnetii</i> RSA493	Seshadri <i>et al.</i> 2003	Y	NC_002971	2,00	22	11,0	42
<i>Acanthamoeba polyphaga</i> Mimivirus	Raoult <i>et al.</i> 2004	Y	NC_006450	1.18	92	78,0	192
Canarypox virus	Tulman <i>et al.</i> 2004	Y	NC_005309	0.36	51	141,7	212
Fowlpox Virus	Afonso <i>et al.</i> 2000	Y	NC_002188	0.29	31	106,9	122
<i>Drosophila melanogaster</i>	Adams <i>et al.</i> 2000	Y	NT_033778	180	148	0,8	577
			NT_033779				
			NT_033777				
			NT_037436				
			NC_004354				
			NC_004353				
<i>Anopheles gambiae</i>	Holt <i>et al.</i> 2002	Y	NT_078265 – 7268	220	120	0,5	384
			NC_004818				
<i>Apis mellifera</i>	Honeybee Genome Sequencing Consortium, 2006	Y	NC_007070 – 7085	210	100	0,5	370
<i>Caenorhabditis elegans</i>	<i>C. elegans</i> Sequencing Consortium, 1998	Y	NC_003279 – 3284	97	130	1,3	446
<i>Tribolium castaneum</i>	Baylor College of Medicine, unpublished	N	NC_007416 – 7425	158	103	0,7	382
<i>Nasonia vitripennis</i>	Baylor College of Medicine, unpublished	N	N/A	?	117	N/A	499
<i>Arabidopsis thaliana</i>	<i>Arabidopsis</i> Genome Initiative, 2000	Y	NC_003070	120	149	1,2	291
			NC_003071				
			NC_003074				
			NC_003075				
			NC_003076				

Table 1: Species/strains used in this study.

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
Genus	N																																
<i>Wolbachia</i>	751	-	G	-	T	P	L	H	-	A	A	-	-	G	-	-	-	V	K	-	L	L	-	-	G	A	-	V	N	-	-	D	-
<i>Anaplasma</i>	15	-	G	-	T	L	L	H	-	A	A	-	-	G	-	-	-	V	-	-	L	L	-	-	G	A	D	-	S	-	-	-	-
<i>Ehrlichia</i>	32	-	G	-	T	-	L	H	-	A	V	-	-	G	-	-	-	V	K	-	L	L	-	-	G	-	D	-	N	-	-	-	-
<i>Neorickettsia</i>	6	-	G	D	T	-	L	H	-	A	-	-	A	G	N	V	-	-	-	-	L	L	L	-	-	G	A	-	-	-	-	-	-
<i>Rickettsia</i>	158	-	G	-	T	-	L	-	-	A	-	-	-	-	N	-	-	-	-	-	L	L	L	-	-	G	A	D	-	N	-	-	-
<i>Orientia</i>	171	-	G	-	T	A	L	H	-	A	-	-	-	-	I	D	I	-	K	L	L	L	-	-	G	A	D	-	N	-	-	D	-
<i>Coxiella</i>	42	-	G	-	T	P	L	-	-	A	A	-	-	G	-	-	-	-	-	-	L	-	-	-	G	A	D	-	N	-	-	-	-
<i>Mimivirus</i>	192	-	-	-	A	A	L	-	-	A	-	-	-	G	H	L	E	-	V	K	Y	L	V	-	-	G	A	D	I	-	-	-	-
<i>Canarypox</i>	212	-	-	-	T	P	L	H	-	A	V	-	-	-	N	-	-	-	V	K	L	L	L	-	-	G	A	D	-	N	-	-	-
<i>Fowlpox</i>	122	-	-	-	T	P	L	H	-	A	-	-	-	-	N	-	-	-	V	-	-	L	L	-	-	G	A	D	-	N	-	-	-
<i>Drosophila</i>	577	-	G	-	T	P	L	H	-	A	A	-	-	G	-	-	-	-	V	-	-	L	L	-	-	G	A	-	-	-	-	-	-
<i>Anopheles</i>	384	-	G	-	T	P	L	H	-	A	A	-	-	G	-	-	-	-	V	-	-	L	L	-	-	G	A	-	-	-	-	-	-
<i>Apis</i>	370	-	G	-	T	P	L	H	-	A	A	-	-	G	-	-	-	-	V	-	-	L	L	-	-	G	A	-	-	-	-	-	-
<i>Caenorhabditis</i>	446	-	G	-	T	P	L	H	-	A	A	-	-	G	-	-	-	-	V	-	-	L	L	-	-	G	A	-	-	-	-	-	-
<i>Tribolium</i>	382	-	G	-	T	P	L	H	-	A	A	-	-	G	-	-	-	-	V	-	-	L	L	-	-	G	A	D	-	-	-	-	-
<i>Nasonia</i>	499	-	G	-	T	P	L	H	-	A	-	-	-	G	-	-	-	-	V	-	-	L	L	-	-	G	A	-	V	N	-	-	-
<i>Arabidopsis</i>	291	-	G	-	T	P	L	H	-	A	A	-	-	G	-	-	-	-	V	-	-	L	L	-	-	G	A	-	-	-	-	-	-

Table 2: ANK repeat consensus for all genera/species used in this study.

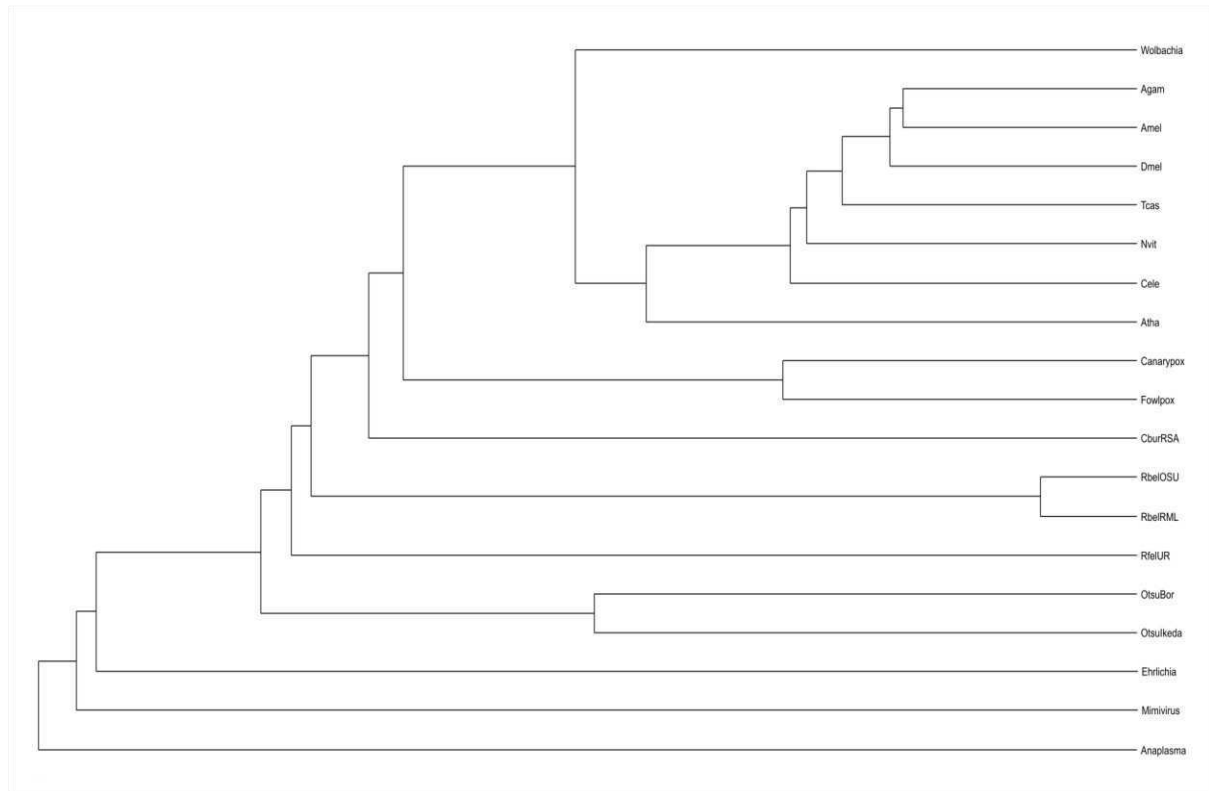


Figure 1: ANK repeat consensus comparison among *Wolbachia* and representatives of viral, prokaryotic and eukaryotic species.

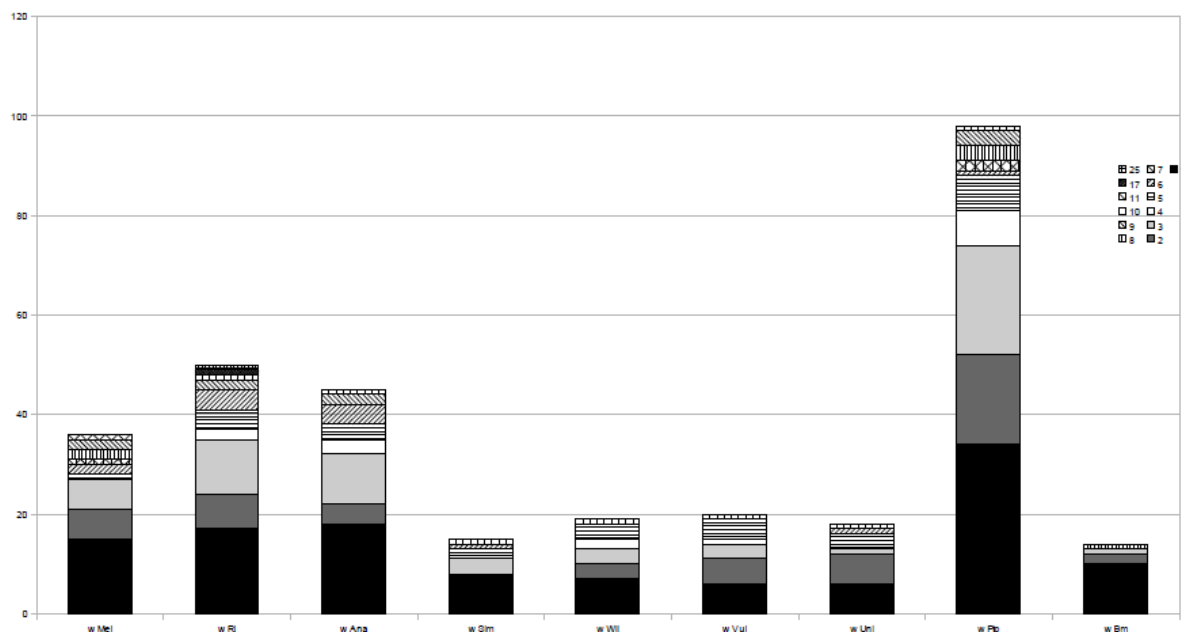


Figure 2: Presence of ANK repeat clusters of different length in *Wolbachia*.

All 9 *Wolbachia* strains used in this study are presented. Clusters were built using the loose criteria (see Methods). In the legend at the upper right portion of the figure it is shown the number of ANK repeats each cluster contains.

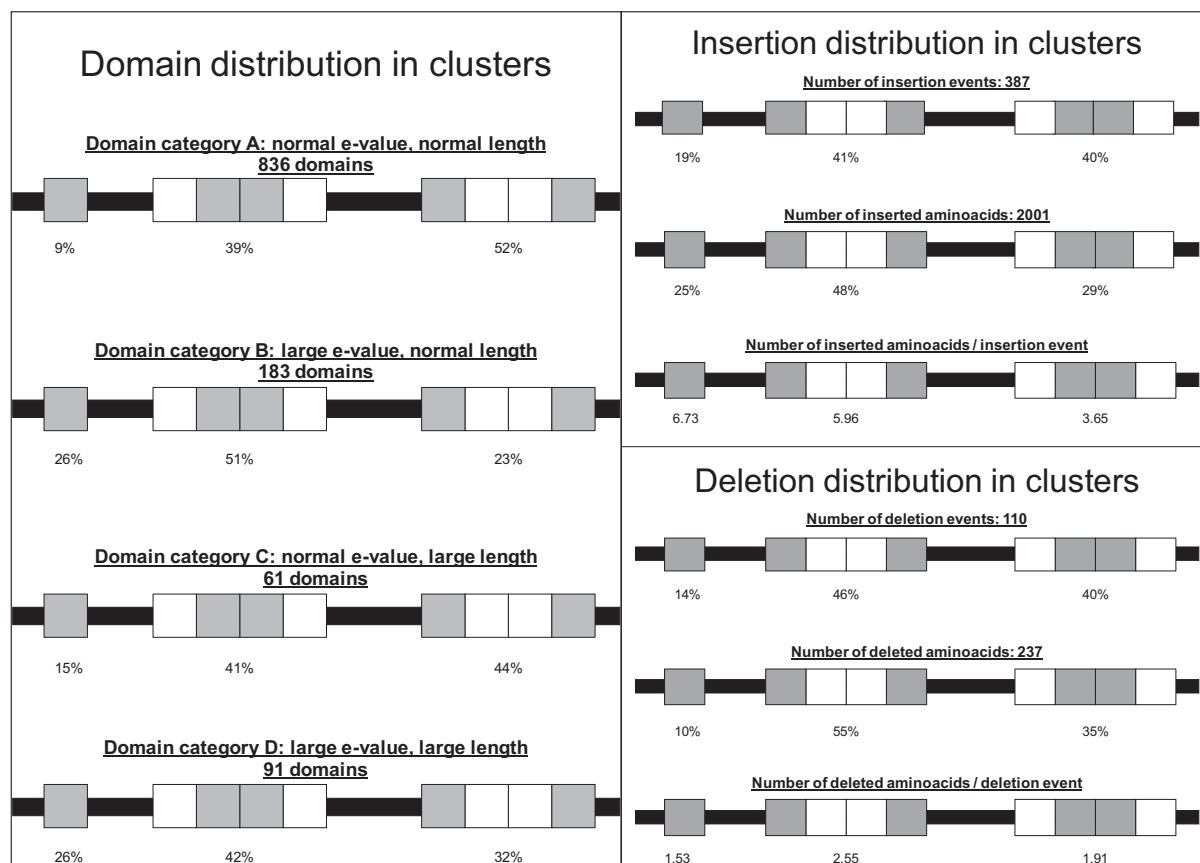


Figure 3: Distribution of ANK repeats within an ANK cluster.

In each domain category (on left), there is a hypothetical gene (shown as a long black rectangle) containing an isolated domain and two other ANK clusters of 4 ANK repeats each. Percentages right below the isolated domains denote the percentage of the corresponding domain category in an isolated domain. Percentages below the ANK clusters are referring to the domains shown in grey (e.g. 39% of category A domains are present in the inner domains of a cluster while another 52% is present in the outer ones). As previously, a hypothetical gene is represented by a black rectangle. An isolated ANK repeat and ANK clusters are drawn on it and ANK repeats in grey are the ones to which the percentages right below, refer to. (on top right) Insertion distribution and (at bottom right) Deletion distribution.

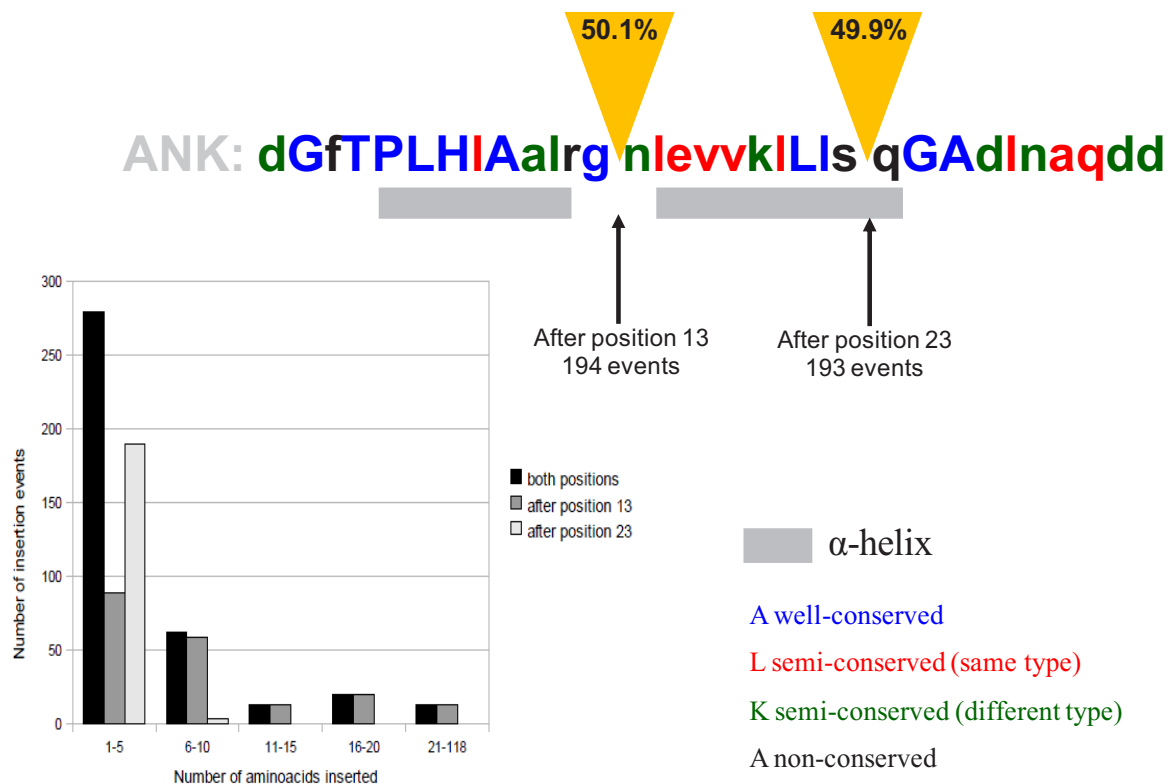


Figure 4: Distribution of insertions within the ANK repeat.

Green bars indicate the position of alpha-helices while in red appear the aminoacid residues between which the insertion events occur. For each bin size the first bar (black) represents the number of insertion events at both positions 13 and 23, the second bar (dark grey) represents the number of insertion events only in position 13 and the third bar (grey) represents the number of insertion events only in position 23.

Gene	Clusters1	Consistency?
wAna3_58534919	3, 3, 22	Yes
wAna5_58534192	5, 2	No
wAna7_58698688	5, 2	Yes
WD_0073	2, 3	Yes
WD_0147	6, 2	No
WD_0291	3, 2	Yes
WD_0498	2, 3	Yes
WP_0383	3, 3, 22	Yes
WP_0410	3, 2, 23	Yes
WP_0688	4, 2	Yes
WP_0763	4, 4	Yes
WP_1220	3, 3, 2	No
wRi_ANK03	2, 3	Yes
wRi_ANK06	2, 4	Yes
wRi_ANK04	5, 2	Yes
wRi_ANK26	2, 4, 2, 24	No
wRi_ANK28	2, 6	No
wRi_ANK20	3, 2, 32	Yes
wSim1_58696982	2, 4	Yes
wUni_1_1601	2, 4	Yes
wUni_1_585	3, 2	No
wVulC_ANK26	2, 4	No
wVulC_ANK51	2, 4	Yes
wVulC_ANK72	4, 4	Yes
wVulC_ANK74	3, 3	Yes
WDW_0144	2, 4	Yes
WDW_0893	3, 2	Yes

Table 3: All cases of consistency/inconsistency in the ankyrin cluster phylogenetic study.

¹ Only clusters having 2, 3, 4, 5 or 6 domains were taken into account. All other clusters did not have enough members, ² The two 3-domain clusters do not present congruent phylogeny, ³ The two 2-domain clusters do not present congruent phylogeny, ⁴ Only two of the three 2-domain clusters present congruent phylogeny

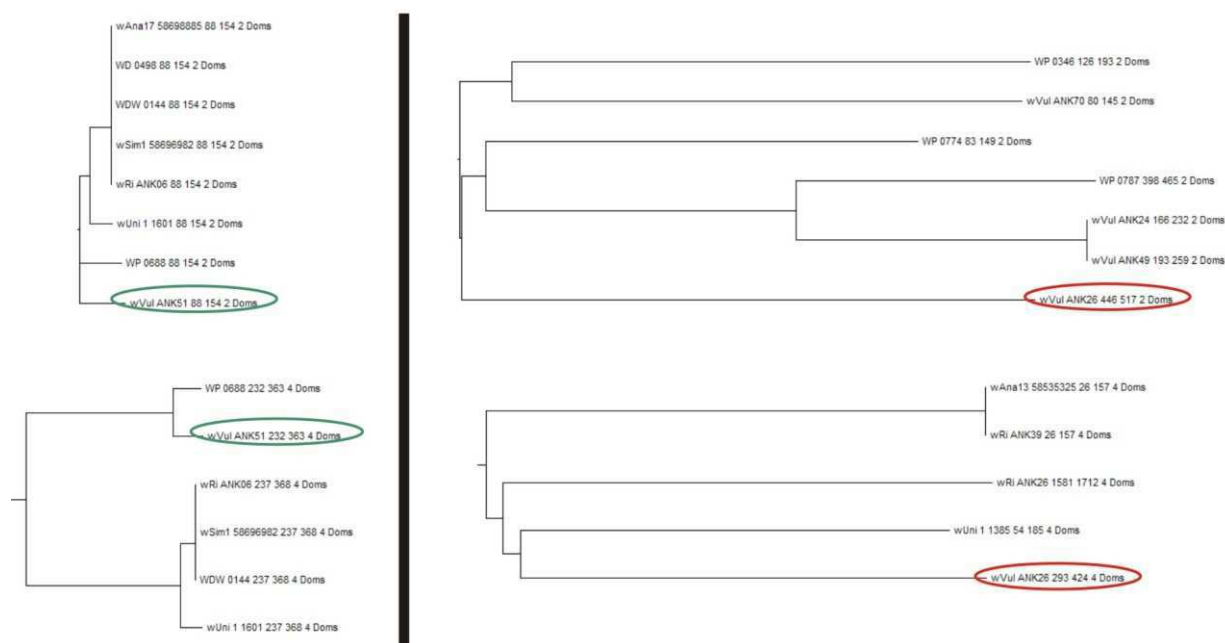


Figure 5: Congruency results from ankyrin cluster phylogenetic study.

Two examples are shown; on the left side there is an example of congruency for the wVulC_ANK51 gene which has a 2-domain and a 4-domain cluster, while in the right side one of incongruency for the gene wVulC_ANK26 which harbors two clusters, one of 2- and another one of 4-domains. Only part of the corresponding phylogenetic trees (for 2- and 4-domain clusters) is shown.

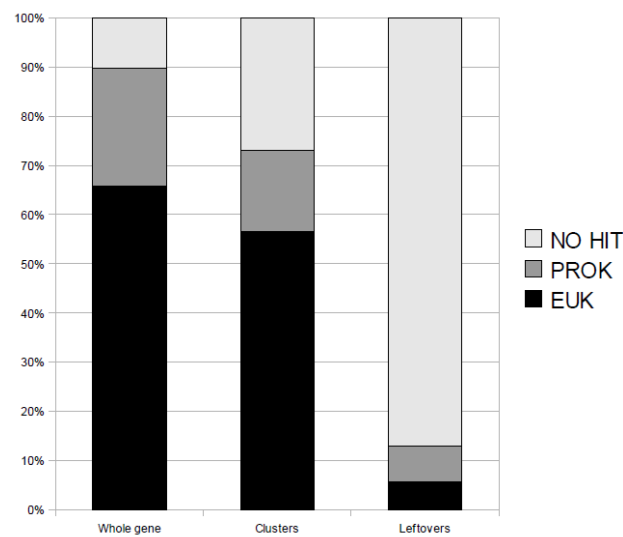


Figure 6: Eukaryotic/Prokaryotic origin of BLAST hits

Percentage of *Wolbachia* whole genes, ANK clusters and leftovers that have a eukaryotic, prokaryotic or no similarity based on BLAST searches against the nr database.

Supplementary files

Supplementary files are available upon request.

E.2 - Pichon *et al.* En préparation.

Article 5: Pichon S., Bouchon D., Liu C., Chen L., Garrett R.A and P. Grève. Specific expression of one ankyrin *pk2* allele in feminizing *Wolbachia* strains. En préparation.

Specific expression of one ankyrin *pk2* allele of the WO prophage in feminizing *Wolbachia* strains

Samuel Pichon¹, Didier Bouchon¹, Chao Liu², Lanming Chen², Roger A. Garrett², and Pierre Grève^{1*}

¹ Ecologie, Evolution, Symbiose, UMR CNRS 6556, Université de Poitiers, 40 avenue du Recteur Pineau, F-86022 Poitiers, France.

² Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 N Copenhagen, Denmark.

* Corresponding author

PG: pierre.greve@univ-poitiers.fr

Phone: +33 (0)5 49 45 39 79, Fax: +33 (0)5 49 45 40 15

Email Addresses:

SP: samuel_pichon@hotmail.com

DB: didier.bouchon@univ-poitiers.fr

CL: chaoliu@bio.ku.dk

LC: lanmingchen@yahoo.com.cn

RG: garrett@bio.ku.dk

Key words: *Wolbachia pipientis*; ankyrin genes; bacterial effectors; prophage; extended phenotype

Abstract

The maternally inherited α -Proteobacteria *Wolbachia pipientis* are obligate endosymbionts of nematodes and a wide range of arthropods, in which they induce a variety of reproductive alterations, including cytoplasmic incompatibility (CI) and feminization. Significant progress at deciphering the molecular basis of these phenotypes resulted from the recent sequencing of a few *Wolbachia* genomes. Both *pk1* and *pk2* phage-related ankyrin genes were the only genes found to be polymorphic between very closely related *Wolbachia* strains expressing different phenotypes in *Culex pipiens*. The *pk2* gene also exhibited differential expression between males and females. Together, these results implicated *pk1* and *pk2* gene products in the *Wolbachia*-induced CI phenotype. The genome of the *Wolbachia* strain *wVulC* that induces feminization of the isopod *Armadillidium vulgare*, contains an unusually high number of ankyrin containing genes (*ank* genes). These encode putative bacterial effector proteins containing ANK repeats that could mediate *Wolbachia*-host protein-protein interactions. Two of the *ank* genes are homologous to *pk1* and *pk2* genes and occur in the *wVulC* genome. To test whether these *Wolbachia ank* gene homologs are potentially involved in altering terrestrial isopod reproduction, we determined the distribution and expression of both *pk1* and *pk2* genes in 8 *Wolbachia* isopod strains that induce either CI or feminization. Although the homologous genes within the *Wolbachia* genomes were highly similar in sequence, our results highlight a large copy number variation amongst strains that may be linked to prophage diversity. Furthermore, transcriptional analyses revealed expression of one *pk2* allele only in the feminizing *Wolbachia* strains of isopods. These results open new perspectives for unravelling functions of ankyrin proteins, in particular *pk2*, with respect to host sex manipulation by *Wolbachia*.

Introduction

The evolutionary success of the maternally inherited α -Proteobacteria *Wolbachia pipientis* is partly due to their ability to manipulate host reproduction to favour their transmission in populations. *Wolbachia* is also able to transfer between hosts *via* horizontal transfers, which contribute to the impressive diversity and range of infected hosts (1). These obligate endosymbionts are found in most filarial nematodes and are estimated to be present in ~60% of arthropod species (2, 3, 4). In arthropods, *Wolbachia* are considered as sex-parasites because they can alter compatibility between eggs and sperm, feminize or kill males, or induce parthenogenesis (2, 5, 6). To date, the mechanisms that mediate these *Wolbachia* extended phenotypes are poorly understood. Since *Wolbachia* remain genetically intractable endosymbionts, comparative genomics and evolutionary approaches are particularly useful for identifying putative bacterial determinants involved in *Wolbachia*-host interactions.

Recent genome analyses of different *Wolbachia* strains revealed a surprisingly high number of Ankyrin domain-containing genes (*ank* genes) whose origin has been proposed to be the result of lateral gene transfer (7, 8, 9, 10, 11). Ankyrin repeats (ANK) constitute 33-residue motifs and are found in numerous eukaryotic proteins and in some bacterial and viral proteins (12, 13). These ANK motifs form tandem arrays that mediate specific protein-protein interactions and have diverse functions in transcription initiation, cell cycle regulation and signalling, cytoskeleton integrity, ion transport, inflammatory responses and development (12, 14). The two closely related intracellular bacteria *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* secrete ankyrin proteins (AnkA and p200, respectively) that bind to the host DNA and/or proteins (15, 16). It has been demonstrated that AnkA plays an important role in facilitating intracellular infection (17) whereas p200 is thought to affect host cell gene transcription and promote the survival of the pathogen (16). Hence it has been proposed that *ank* genes might encode *Wolbachia* effectors that alter host biology (18, 19).

Several studies have challenged the hypothesis that *Wolbachia* ANK proteins are implicated in the molecular basis of Cytoplasmic Incompatibility (CI) (8, 9, 20, 21, 22, 23). Sequence divergence between two closely related *Wolbachia* strains infecting distinct *Culex pipiens* mosquito populations that are incompatible was only recorded in 2 *ank* genes (*pk1* and *pk2*) amongst a subset of 18 *ank* genes and other phylogenetic markers (22). These polymorphic *pk1* and *pk2* *ank* genes, located within a WO prophage region, were thereby suggested to contribute to the CI phenotype. In addition, expression of the *pk2* genes occurred specifically in female mosquitoes consistent with its direct involvement in the CI effect (8, 22, 23). Moreover, in the *Wolbachia* strain of *D. simulans* (*wAu*) which was unable to cause CI a premature stop codon was found in the *pk2* gene (21). The CI phenotype seems to delay recruitment of the histone H3.3/H4 complex to the male pronucleus causing delay of mitotic condensation and pronucleus segregation (24). This interphase cell cycle disruption causes an asynchrony in the development of the male and female pronuclei at fecundation which eventually leads to death of the embryos (24). To date, however, there is no evidence for the direct involvement of bacterial ankyrin genes in the CI phenotype.

In this study, we aimed to determine whether the prophage *pk1* and *pk2* ankyrin genes were conserved and expressed in *Wolbachia* strains that induce CI in terrestrial isopods but also feminization, the main *Wolbachia* phenotype described for this group of hosts (2). First we annotated the *pk1* and *pk2* alleles amongst all *ank* genes identified from the genome assembly of the feminizing *Wolbachia* strain *wVulC* that infects the isopod *Armadillidium vulgare*. Then we investigated the distribution, copy number and expression patterns of both genes in 7 additional *Wolbachia* strains that induce either CI or feminization in isopods. Thus we identified a large copy number variation of the *pk1* and *pk2* genes amongst *Wolbachia* strains that is probably coupled to prophage evolution. Our results also reveal that one *pk2* allele expression is only detected in feminizing strains but never in CI-inducing strains of isopods.

Results

Characterization and distribution of *pk1* and *pk2* genes

Six and three copies of the phage-related *pk1* and *pk2* *ank* genes, respectively, were identified in the contig assembly of the *wVulC* genome (Table 1). Sequence analyses revealed that two *wVulC* *pk1* alleles (ANK46a/b and ANK60a/b) and one *pk2* allele (ANK40a/b) were each found in two identical copies. These results were confirmed by Southern blots (Figure S1) and are consistent with the sequencing of PCR products.

Phylogenetic relationships between these alleles and all the published sequences of *Wolbachia* *pk1* and *pk2* genes were assessed based on the DNA sequences encoding ANK-repeats. No evidence of recombination events was detected in the *pk1* and *pk2* alignments. In the *pk1* sequences, the number of variable sites is 460 out of 1068 of which 405 are informative. Similarly, the number of informative sites is 111 for the 369 bp-long *pk2* sequence alignment. The resulting unrooted phylogenetic trees show that *pk1* sequences group in 3 clusters whereas *pk2* sequences group in 2 clusters (Figure 1). Thus we define three different types of *pk1* genes (named *pk1a*, *pk1b*, *pk1c*) and two different types of *pk2* genes (named *pk2a*, *pk2b*) (Figure 1, Table 2). The *wVulC* genome contains a single copy of the *pk1a* type (ANK25) and 5 copies belonging to the *pk1b* type (ANK32, ANK46a/b & ANK60a/b) (Table 2). The *wVulC* *pk1a* gene (ANK25) is close to the *pk1a* type gene of *wMel* (84.2% of identity) whereas it shares 61.6 to 70.0 % identity with the 5 *wVulC* *pk1b* type sequences. These latter share 84.1 to 91.7 % identities with the three *pk1b* paralogs of *wPip-Pel*. The *wVulC* genome also harbours three genes of the *pk2b* type (ANK40a/b, ANK48) further called *pk2b1* and *pk2b2* alleles respectively (Tables 2 & 3). These genes were closer (87.4 to 90.2 % identity) to one of the three *pk2* genes of *wPip-Pel* (ANK16, *pk2b* type) than to the other two identical copies of *wPip-Pel* (ANK12 & 25, *pk2b* type).

Using the same primer set as for *wVulC* (Table S1), the taxonomic distribution of *pk1* and *pk2* genes was extended to 7 *Wolbachia* strains that induce either CI or feminization in isopods. We also investigated the copy number variation by Southern blot analyses of *EcoRI* or *BamHI* digested DNA

using *pk1a*, *pk1b* and *pk2b1* probes which, according to sequence identities, preferentially hybridized on *pk1a*, *pk1b* and *pk2b* types, respectively (Table 2). In congruence with amplification and sequencing data, the *pk1a* and *pk1b* probes revealed from 2 to 6 copies of *pk1* gene in the studied strains (Table 2). By direct sequencing of the PCR products, we found that the *pk1a* gene of *Wolbachia* strains of *C. convexus*, *P. pruinosus*, *A. vulgare* (*wVulM*) and *A. nasatum* harboured 1, 1, 2 and 3 *EcoRI* sites, respectively, explaining the discrepancy between the number of bands observed by Southern blots, and the number of different sequences obtained (Table 2, Figure S1). Similarly, two *pk1b* alleles of the *Wolbachia* strain of *A. nasatum* contained one *BamHI* restriction site. For *wVulC* *pk1b* alleles, each of the two more intense signals obtained by Southern Blot (Figure S1) revealed the presence of two identical copies as confirmed by the analysis of contigs. Furthermore, in all sampled strains, Southern blots using a *pk2b1* probe in combination with sequencing data revealed 3 copies of the *pk2* gene (Table 2, Figure S1). Once again, sequencing of PCR products identified 2 *pk2* identical alleles each containing one *BamHI* restriction site explaining the 5 signals obtained by Southern blot in the *Wolbachia* strain of *P. pruinosus* (Table 2, Figure S1).

Molecular evolution of *pk1* and *pk2*

The G+C content of *wVulC* *pk1* genes ($34.2 \pm 0.3\%$) is very similar to that of the whole genome (35.0%). The G+C content of *wVulC* *pk2* genes (ANK40a/b: 36.0%, ANK48: 40.0%) is greater to that of the whole genome. Similar results are obtained for G+C content of all sequences (*pk1*: $34.1 \pm 0.1\%$; *pk2*: $37.9 \pm 0.2\%$; genomes: $34.8 \pm 0.3\%$). Protein-coding sequence analysis indicated that Ka/Ks (the ratio of the rate of non-synonymous substitutions (Ka) to the rate of synonymous substitutions (Ks)) of all positions is 0.211 ± 0.009 for Pk1 and 0.245 ± 0.020 for Pk2 and no sites are under positive selection.

All translated *pk1* full-length sequences are predicted to harbour 2 transmembrane domains in their C-terminal region but a variable number of ANK motifs ranging from 8 to 10 (Figure S2, Table S2). In *wVulC*, ANK46a/b and ANK60a/b sequences (*pk1b* type) are shorter in their N-terminal region than all other Pk1 proteins (about 40 and 60 amino acids, respectively). One indel at position 117 of the

DNA sequence of *wVulC* ANK46a/b is responsible for a frameshift and the split into two ORFs homologous to the full length *pk1* of other strains. ANK60 sequences are shortened by a transposase gene insertion in the 5' region of both copies. In contrast, *pk2* translated sequences are relatively more conserved amongst *Wolbachia* strains than *pk1* as identity ranges from 84.5 to 100%. All *Pk2* amino acid sequences harbour 3 ANK motifs except in the *wAu* (host: *D. simulans*) and *wWil* (host: *D. willistoni*) strains in which *Pk2* proteins only harbour 2 repeats due to a premature stop codon that disrupts the third motif (Figure S2, Table S2).

In the *wVulC* assembly, 6 prophage regions were identified (WO-*wVulC*1 to 6) (Table 1). The *pk1* and *pk2* genes are located within a syntenic prophage region containing the *orf7* gene of which phylogeny has been well studied. This gene encodes a minor capsid protein that is located either between the *pk1* and *pk2* genes, or adjoining one *pk1* or one *pk2* gene depending on the prophage region considered (Table 1). Therefore, we compared the *pk1* and *pk2* allele phylogenies to that of *orf7* with respect to their prophage membership in various strains (Table 1). Phylogenies between *pk2* and *orf7* are congruent suggesting a co-evolution of *pk2* and *orf7* (Figure 2). In contrast, 8 of the 15 prophages showed distinct topologies between the trees of *pk1* and *orf7* alleles (Figure 2).

Comparative analysis of *pk1* and *pk2* mRNA expression in CI and feminizing *Wolbachia* strains

RT-PCR using allele-specific primers was performed to examine the expression patterns of *pk1* and *pk2* mRNA in adult gonads of isopods of both CI-inducing and feminizing *Wolbachia* strains (Table 3). Evidence for expression was observed for all copies of *pk1* and *pk2* genes except for one allele of the *pk2b* type (Table 3). Using the *pk2b2* allele specific primers, RT-PCR did not yield any detectable products in all the three isopods CI-inducing *Wolbachia* strains whereas this gene was clearly expressed in feminizing *Wolbachia* strains. When both males and females are infected, no sex-specific differences were observed in *pk1* and *pk2* expression, whatever the reproductive phenotype induced by the bacteria.

Discussion

In this study, we have demonstrated a large copy number variation of *pk1* and *pk2* genes amongst *Wolbachia* strains that is probably coupled to prophage dynamics and evolution. Divergence in the ankyrin *pk1* and *pk2* copy number in *Wolbachia* strains of isopods is consistent with the results of previous Southern blotting analyses using the minor capsid *orf7* phage gene (25). Four different *orf7* paralogs had already been identified in the *wVulC* strain through cloning and sequencing of heterogeneous PCR products (25). Since multiple infections in a single individual have never been demonstrated in natural populations of isopods, we can conclude that the phage WO is likely to be present in several copies in each *Wolbachia* strain. Besides identifying mosaic prophages, our results also reveal the importance of the differentially expressed genes between *Wolbachia* strains inducing CI and feminization. We show that expression of one allele (*pk2b2*) is only detected in feminizing strains but never in CI-inducing strains of isopods. This leads to the hypothesis that this bacterial ankyrin gene is directly involved in the reproductive alteration of isopod hosts. In contrast to the observations for *wPip*, expression of the *pk2* allele was not detected in the three CI-inducing *Wolbachia* strains, suggesting that this gene is not involved in CI in isopods. Furthermore, no differential expression of *pk1* and *pk2* genes was detected between the sexes in hosts where both males and females are infected by either CI-inducing or feminizing *Wolbachia*. This result differs from those of Sinkins and colleagues who showed that in some CI-inducing *wPip* variants, *pk2* genes are highly expressed in females but never in males (22, 23). Expression of *pk2* genes - both identical copies in the *wPip*-Pel genome (*wPip*_ANK12 and *wPip*_ANK25) and one paralog close to that of *Wolbachia* of *C. molestus* (*wPip*_ANK16) - was sex-specific in different *Wolbachia* strains of *C. pipiens* (23). These paralogs are also part of an operon containing three prophage-associated genes with sex-specific expression in *wPip* strains (23). Additionally to *pk2*, three *ank* genes (*wPip*_ANK1, *wPip*_ANK52 and *wPip*_ANK57) showed a sex-specific expression in *wPip*-infected *C. pipiens* group adults (23). Another *ank* gene (WRi_003070) of the CI-inducing *wRi* strain was also shown to be transcribed in ovaries but not in

testes in adult flies (9). In consequence, distribution and transcriptional analyses might be extended to these four supplementary genes, if their homologs are present, to check whether a similar expression pattern is observed which would be consistent with their direct involvement in the CI effect of *Wolbachia* strains harboured by isopods.

Our observations on *Wolbachia* strains of isopods suggest prophage dynamics and/or their multiplication in genomes is similar to that observed in the *wRi* and *wPip-Pel* genomes. The *pk1* and *pk2 ank* gene are each found in a single copy within the WO-B prophage of the *wMel* genome (11). Genome sequencing of the CI-inducing *wPip-Pel* and *wRi* *Wolbachia* strains revealed duplication of these particular genes together with the WO-B-like prophages (8, 9) (Figure 3). The *wPip-Pel* genome contains 5 prophage regions more closely related to the WO-B than to the WO-A *wMel* prophage (8). The *pk1* and *pk2* genes are spread among the 5 different prophages and each occur in 3 copies (identical triplicates for *pk1* or 77-100% identity for *pk2*) (8). The *pk1* gene is found in 4 copies in *wRi* of which two are pseudogenes and spread in the prophage named WO-C (9). Both pseudogenes (*wRi_ANK29* and *ANK31*) may have originally been a single *pk1* gene further disrupted by an insertion sequence ISWpi7. Divergence between strains of the *pk1* and *pk2* gene copy numbers is therefore likely to be the result of insertion sequence events, recombination and/or lateral transfer of prophage (26, 27, 28). These mechanisms have consequences for bacterial genomes which include the modification of genome organization and the modulation of gene copy number and expression (29, 30). Although more complete genome sequences would be helpful, for example those of *wCauB* and *wKue*, comparison of the phylogeny using *orf7* or the *pk1* and *pk2* genes allowed us to decipher the underlying mechanisms of genomic rearrangement occurring in prophages. According to the comparative phylogenies of *pk1* and *orf7* genes, it is likely that sequence insertion and/or recombination events have preferentially occurred in phage region directly around the *pk1* gene. On the other hand, similar topology between *orf7* and *pk2* gene trees plus the high G+C content of *pk2* favour the occurrence of recent lateral transfers of prophage fragments containing the *orf7* and *pk2* genes in the *Wolbachia* genomes. However, we cannot exclude the hypothesis that linkage

disequilibrium occurs between these two genes that are separated by at least 6.7 kilobases, representing less than 0.04% of the genome size. Alternatively, these results highlight the plasticity of the prophage region amongst *Wolbachia* strains as part of the global plasticity observed in the *Wolbachia* genomes. Finally, our results are also consistent with duplications of, for example, the *vir* and *wsp* genes, and with the transfer of insertion sequences within strains as previously highlighted in *wMel* and *wVulC* (7, 8, 9, 10, 11, 31).

Considering bacterial genomes sequenced to date, prophage sequences are particularly abundant in pathogens (32). However, prophages are also missing from some of the most reduced endosymbiont genomes (*i.e.*, *Buchnera*, *Wigglesworthia*, and *Blochmannia*) except for *Ca. Hamiltonella defensa*, *Sodalis glossinidius* and *Wolbachia pipientis* (33, 34, 35). These prophages are noticeably absent in the mutualistic *Wolbachia* strain infecting the nematode *Brugia malayi* (7). Therefore, we suggest that maintenance of such “mobile elements” in *Wolbachia* strains of arthropods may be due to the absence of, or a reduced efficiency of selection on the prophages. However, a recent comparative genome analysis using microarray revealed that phenotype-related effectors or specific strain determinants in *Wolbachia* are likely to be encoded by prophage genes, ankyrin-repeat encoding genes, and predicted genes of unknown function (36). In consequence, our future attention will focus on the *wVulC* specific bacteriophage genes to check whether some of them may be related to virulence proteins as described in WOCauB1 (28) or encode for other ankyrin repeat proteins which could play a role in the host reproductive alteration. To date, the molecular mechanisms that mediate *Wolbachia* feminization are also poorly understood. *Wolbachia*, may target the proteinaceous androgenic hormone or its receptor, or another determinant of sex ultimately leading to the inhibition of androgenic gland differentiation and preventing the androgenic hormone from reaching the target tissues such as gonads and tegumental epithelium (2, 37). This hypothesis suggests a late action of *Wolbachia* on host target(s) during development, as opposed to the very early action of other *Wolbachia* strains that induce parthenogenesis, CI or male killing (5).

In conclusion, our results highlight a large copy number variation amongst strains of both *pk1* and *pk2* genes that may be linked to prophage diversity and the specific expression of one *pk2* allele in the feminizing *Wolbachia* strains of isopods. These results open new perspectives for unravelling functions of ankyrin proteins with respect to host sex manipulation by *Wolbachia*. In particular, identification of putative bacterial determinants such as *Pk2b2* combined with the search for host molecules targeted by *Wolbachia* ankyrins may provide further insights into the feminizing phenotype of *Wolbachia*.

Materials and Methods

Wolbachia-infected isopod species

All isopod species used in this study were collected in France and reared in the laboratory. CI *Wolbachia* were found in three isopod species: *Cylisticus convexus* (*wCon*, Villedaigne) (38), *Porcellio dilatatus petiti* (*wPet*, Saint-Honorat) (39) and *P. d. dilatatus* (*wDil*, Sainte-Marguerite) (Grève, unpublished results). *Wolbachia* strains inducing feminization have been described in *A. vulgare* (*wVulC*, Celles sur Belle and *wVulM*, Mery sur Cher) (40, 41), *A. nasatum* (*wNas*, Poitiers) (42), *Oniscus asellus* (*wAse*, Quinçay) (43), *Porcellionides pruinosus* (*wPruIII*, Nevers) (44).

Total DNA was extracted from males and females of the 8 isopod species as described previously (45). Infection status of each individual was confirmed by a PCR-assay based on the bacterial 16S rDNA gene using *Wolbachia*-specific primers (46).

Distribution of *pk1* and *pk2* genes

Annotation of the *pk1* and *pk2* genes in the genome assembly of the feminizing *Wolbachia* strain *wVulC* was performed by protein and DNA homology searches with BLASTP and BLASTN programs (47) using *wPip-Pel* *pk1* and *pk2* alleles as query. Ankyrin and other functional motif predictions were performed by SMART web server (48) on protein sequences.

wVulC specific primers were designed to amplify full-length or 200-500 bp fragments of *pk1* and *pk2* alleles using a standard PCR protocol (Table S1). The purified PCR products were directly sequenced on both strands on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems) using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) according to the manufacturer's instructions.

pk1 or *pk2* copy number variation among *Wolbachia* strains was assessed by Southern blot. About 15 µg of total DNA were digested at 37°C overnight with *EcoRI* or *BamHI* enzymes that did not cut any of the *wVulC* *pk1* and *pk2* alleles. DNA was electrophoresed on 0.8% agarose gels and southern blotted to nylon membranes. Probes were obtained by PCR amplification of the *wVulC* full-length *pk1* and *pk2* *ank* genes (Table S1), labelled using [α -³²P]-dCTP by the random primer method and hybridized overnight to membranes. The final wash was performed at 52°C in 0.1X SSC. Hybridized blots were imaged and analyzed using a PhosphoImager (Molecular Dynamics, Sunnyval, CA, USA).

Sequence analyses of *pk1* and *pk2* genes

Homologous sequences of both genes were first aligned in the server-based program MAFFT (<http://align.bmr.kyushu-u.ac.jp/mafft/online/server/>) with default parameters. The resulting matrices were manually checked according to the predicted amino acid translation using BioEdit v7.0.9 (49). MEGA software v4.0 (50) was used to calculate nucleotide sequence divergence. For each locus, the G+C content, the number of variable sites and the level of nucleotide diversity per site (Pi) were calculated. Ka/Ks analysis was also performed using the Selecton web server (51).

Recombination analysis was performed with RDP version v3.42 (52) using non-redundant sequence alignment of both *ank* genes. The parameters were set as follows: sequences were considered linear, the highest acceptable P value cut-off was 0.01, a Bonferroni correction was applied, consensus daughter sequences were found, gaps were included, different window sizes of variable sites were

tested and 1,000 permutations were performed. All recombination events detected by the program were visually inspected in the alignment.

Phylogenies based on *pk1* and *pk2* nucleotide region encoding ANK-repeat domains were constructed using MrBayes v3.1.2 software (53, 54). The appropriate model of DNA evolution was estimated with jModelTest v0.1.1 (55), and the best likelihood score was evaluated with the Akaike Information Criterion (56). The selected model was HKY+G for the *pk1* and *pk2* locus encoding ANK domain cluster. Four Metropolis-coupled Markov chain Monte Carlo (MCMC) analyses were run for 1,000,000 generations and sampled every 100 generations. Burn-in included the first 10,000 generations. Statistical support was assessed using Bayesian posterior probabilities (BPPs). Phylogenies of both ankyrin genes were also compared to that of the prophage *orf7* gene following identical methodology (model: TVM+G).

Expression of ankyrin genes

Total RNAs were isolated using the RNeasy Mini Kit (Qiagen) from 20 to 50 gonads dissected from all species. Only ovaries were used in *A. vulgare* and *A. nasatum* where only females are infected. Total RNAs were treated with DNaseI (2U/μL, Ambion) at 37°C for 30 min. Reverse transcription was performed by using Superscript III kit (Invitrogen) as described by the manufacturer, using 1 μg RNA as template and random primers. Reverse transcriptase was omitted in the negative controls. To determine the expression of each gene, 1 μl of the reverse transcriptase reaction was used as template for the RT-PCR experiments.

Acknowledgements

We are grateful to C. Debenest, C. Delaunay, D. Guyonnet, A. Lafitte, J. Lesobre and M. Raimond for technical assistance. We thank I. Mazerie and H. de Verdal for their contributions to the experiments. This research was funded by the Centre National de la Recherche Scientifique (CNRS), the French Ministère de l'Éducation Nationale, de l'Enseignement Supérieur et de la Recherche, the French embassy in Denmark and the Agence Nationale de la Recherche (ANR-06-BLAN-0316). SP was supported by a Ph.D. fellowship from Région Poitou-Charentes.

References

1. Baldo L, Hotopp JC, Jolley KA, Bordenstein SR, Biber SA, *et al.* (2006) Multilocus Sequence Typing System for the Endosymbiont. *Applied and Environmental Microbiology* 72: 7098-7110. doi:10.1128/AEM.00731-06
2. Bouchon D, Cordaux R, Grève P (2008) Feminizing *Wolbachia* and the evolution of sex determination in isopods. In: Bourtzis K, Miller TA. *Insect Symbiosis*. Boca Raton, FL: Taylor & Francis Group, Vol. 3. pp. 273-294.
3. Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH (2008) How many species are infected with *Wolbachia*? A statistical analysis of current data. *FEMS Microbiology Letters* 281: 215-220.
4. Taylor MJ, Bandi C, Hoerauf A (2005) *Wolbachia* bacterial endosymbionts of filarial nematodes. *Advances in Parasitology* 60: 245-284. doi:10.1016/S0065-308X(05)60004-8
5. Bourtzis K, Miller TA (2003) *Insect Symbiosis*. Bourtzis K, Miller TA Boca Raton, FL: Taylor & Francis Group.
6. Werren JH (1997) Biology of *Wolbachia*. *Annual Review of Entomology* 42: 587-609.
7. Foster J, Ganatra M, Kamal I, Ware J, Makarova K, *et al.* (2005) The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biology* 3: e121.
8. Klasson L, Walker T, Sebaihia M, Sanders MJ, Quail MA, *et al.* (2008) Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group. *Molecular Biology and Evolution* 25: 1877-1887.

9. Klasson L, Westberg J, Sapountzis P, Näslund AK, Lutnaes Y, *et al.* (2009) The mosaic genome structure of the *Wolbachia* wRi strain infecting *Drosophila simulans*. *Proceedings of the National Academy of Sciences USA* 106: 5725-5730.
10. Salzberg SL, Hotopp JC, Delcher AL, Pop M, Smith DR, *et al.* (2005) Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. *Genome Biology* 6: R23.
11. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, *et al.* (2004) Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biology* 2: e69.
12. Bork P (1993) Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins* 17: 363-374.
13. Li J, Mahajan A, Tsai M (2006) Ankyrin repeat: a unique motif mediating protein-protein interactions. *Biochemistry* 45: 15168-15178.
14. Mosavi LK, Cammett TJ, Desrosiers DC, Peng Z (2004) The ankyrin repeat as molecular architecture for protein recognition The ankyrin repeat as molecular architecture for protein recognition. *Protein Science* 13: 1435-1448. doi:10.1110/ps.03554604
15. Park J, Kim KJ, Choi K, Grab DJ, Dumler JS (2004) *Anaplasma phagocytophilum* AnkA binds to granulocyte DNA and nuclear proteins. *Cellular Microbiology* 6: 743-751. doi:10.1111/j.1462-5822.2004.00400.x
16. Zhu B, Nethery KA, Kuriakose JA, Wakeel A, Zhang X, *et al.* (2009) Nuclear Translocated *Ehrlichia chaffeensis* Ankyrin Protein Interacts with a Specific Adenine-Rich Motif of Host Promoter and Intronic Alu Elements. *Infection and Immunity* 77: 4243-4255. doi:10.1128/IAI.00376-09
17. Lin M, den Dulk-Ras A, Hooykaas PJ, Rikihisa Y (2007) *Anaplasma phagocytophilum* AnkA secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection. *Cellular Microbiology* 9: 2644-2657. doi:10.1111/j.1462-5822.2007.00985.x
18. Fenn K, Blaxter M (2006) *Wolbachia* genomes: revealing the biology of parasitism and mutualism. *Trends in Parasitology* 22: 60-65.
19. Iturbe-Ormaetxe I, O'Neill SL (2007) *Wolbachia*-host interactions: connecting phenotype to genotype. *Current Opinion in Microbiology* 10: 221-224. doi:10.1016/j.mib.2007.05.002
20. Duron O, Boureux A, Echaubard P, Berthomieu A, Berticat C, *et al.* (2007) Variability and expression of ankyrin domain genes in *Wolbachia* variants infecting the mosquito *Culex pipiens*. *Journal of Bacteriology* 189: 4442-4448.

21. Iturbe-Ormaetxe I, Burke GR, Riegler M, O'Neill SL (2005) Distribution, Expression, and Motif Variability of Ankyrin Domain Genes in *Wolbachia pipientis*. *Journal of Bacteriology* 187: 5136-5145. doi:10.1128/JB.187.15.5136
22. Sinkins SP, Walker T, Lynd AR, Steven AR, Makepeace BL, *et al.* (2005) *Wolbachia* variability and host effects on crossing type in *Culex* mosquitoes. *Nature* 436: 257-260.
23. Walker T, Klasson L, Sebaihia M, Sanders MJ, Thomson NR, *et al.* (2007) Ankyrin repeat domain-encoding genes in the wPip strain of *Wolbachia* from the *Culex pipiens* group. *BMC Biology* 5: 39.
24. Landmann F, Orsi GA, Loppin B, Sullivan W (2009) *Wolbachia*-mediated cytoplasmic incompatibility is associated with impaired histone deposition in the male pronucleus. *PLoS Pathogens* 5: e1000343.
25. Braquart-Varnier C, Grève P, Félix C, Martin G (2005) Bacteriophage WO in *Wolbachia* infecting terrestrial isopods. *Biochemical and Biophysical Research Communications* 337: 580-585. doi:10.1016/j.bbrc.2005.09.091
26. Bordenstein SR, Wernegreen JJ (2004) Bacteriophage flux in endosymbionts (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. *Molecular Biology and Evolution* 21: 1981-1991.
27. Gavotte L, Henri H, Stouthamer R, Charif D, Charlat S, *et al.* (2007) A Survey of the Bacteriophage WO in the Endosymbiotic Bacteria *Wolbachia*. *Molecular Biology and Evolution* 24: 427-435. doi:10.1093/molbev/msl171
28. Tanaka K, Furukawa S, Nikoh N, Sasaki T, Fukatsu T (2009) Complete WO phage sequences reveal their dynamic evolutionary trajectories and putative functional elements required for integration into the *Wolbachia* genome. *Applied and Environmental Microbiology* 75: 5676-5686.
29. Masui S, Kamoda S, Sasaki T, Ishikawa H (2000) Distribution and evolution of bacteriophage WO in *Wolbachia*, the endosymbiont causing sexual alterations in arthropods. *Journal of Molecular Evolution* 51: 491-497. doi:10.1007/s002390010112
©
30. Masui S, Kuroiwa H, Sasaki T, Inui M, Kuroiwa T, *et al.* (2001) Bacteriophage WO and virus-like particles in *Wolbachia*, an endosymbiont of arthropods. *Biochemical and Biophysical Research Communications* 283: 1099-1104.
31. Cordaux R, Pichon S, Ling A, Pérez P, Delaunay C, *et al.* (2008) Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. *Molecular Biology and Evolution* 25: 1889-1896.
32. Brüssow H, Canchaya C, Hardt W (2004) Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiology and Molecular Biology Reviews* 68: 560-602. doi:10.1128/MMBR.68.3.560

33. Bordenstein SR, Reznikoff WS (2005) Mobile DNA in obligate intracellular bacteria. *Nature Reviews Microbiology* 3: 688-699.
34. Moran NA, Degnan PH, Santos SR, Dunbar HE, Ochman H (2005) The players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. *Proceedings of the National Academy of Sciences USA* 102: 16919-16926.
35. Toh H, Weiss BL, Perkin SA, Yamashita A, Oshima K, *et al.* (2006) Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Research* 16: 149-156. doi:10.1101/gr.4106106.1
36. Ishmael N, Dunning Hotopp JC, Ioannidis P, Biber SA, Sakamoto J, *et al.* (2009) Extensive genomic diversity of closely related *Wolbachia* strains. *Microbiology* 155: 2211-2222. doi:10.1099/mic.0.027581-0
37. Juchault P, Legrand JJ (1985) Mechanism of the refractory state of androgen hormone in *Armadillidium vulgare* Latr. (crustacean, isopod, oniscoid) harboring a feminizing bacteria. *General and Comparative Endocrinology* 60: 463-467.
38. Moret Y, Juchault P, Rigaud T (2001) *Wolbachia* endosymbiont responsible for cytoplasmic incompatibility in a terrestrial crustacean: effects in natural and foreign hosts. *Heredity* 86: 325-332.
39. Legrand JJ, Martin G, Artault JC (1978) Correlation between the presence of a bacterial symbiont in oocytes of *Porcellio dilatatus petiti*, and the sterility of the cross: P.d. petiti male x P.d. dilatatus female. *Archives de l'Institut Pasteur de Tunis* 55: 507-514.
40. Cordaux R, Michel-Salzat A, Frelon-Raimond M, Rigaud T, Bouchon D (2004) Evidence for a new feminizing *Wolbachia* strain in the isopod *Armadillidium vulgare*: evolutionary implications. *Heredity* 93: 78-84.
41. Rigaud T, Souty-Grosset C, Raimond R, Mocquard J, Juchault P (1991) Feminizing endocytobiosis in the terrestrial crustacean *Armadillidium vulgare* Latr. (Isopoda): recent acquisitions. *Endocytobiology and Cell Research* 7: 259-273.
42. Rigaud T, Pennings PS, Juchault P (2001) *Wolbachia* Bacteria Effects after Experimental Interspecific Transfers in Terrestrial Isopods. *Journal of Invertebrate Pathology* 77: 251-257. doi:10.1006/jipa.2001.5026
43. Rigaud T, Moreau J, Juchault P (1999) *Wolbachia* infection in the terrestrial isopod *Oniscus asellus*: sex ratio distortion and effect on fecundity. *Heredity* 83: 469-475.
44. Michel-Salzat A, Cordaux R, Bouchon D (2001) *Wolbachia* diversity in the *Porcellionides pruinosus* complex of species (Crustacea: Oniscidea): evidence for host-dependent patterns of infection. *Heredity* 87: 428-434.

45. Bouchon D, Rigaud T, Juchault P (1998) Evidence for widespread *Wolbachia* infection in isopod crustaceans : molecular identification and host feminization. *Proceedings of the Royal Society B: Biological Sciences* 265: 1081-1090.
46. O'Neill SL, Giordano R, Colbert AM, Karr TL, Robertson HM (1992) 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. *Proceedings of the National Academy of Sciences USA* 89: 2699-2702.
47. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215: 403-410.
48. Letunic I, Doerks T, Bork P (2009) SMART 6: recent updates and new developments. *Nucleic Acids Research* 37: D229-232.
49. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95-98.
50. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24: 1596-1599.
51. Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, *et al.* (2007) Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Research* 35: W506-W511.
doi:10.1093/nar/gkm382
52. Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* 21: 260-262.
53. Huelsenbeck JP, Ronquist F (2001) MrBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755.
54. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574.
55. Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* 25: 1253-1256.
56. Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic Biology* 53: 793-808.

Figures

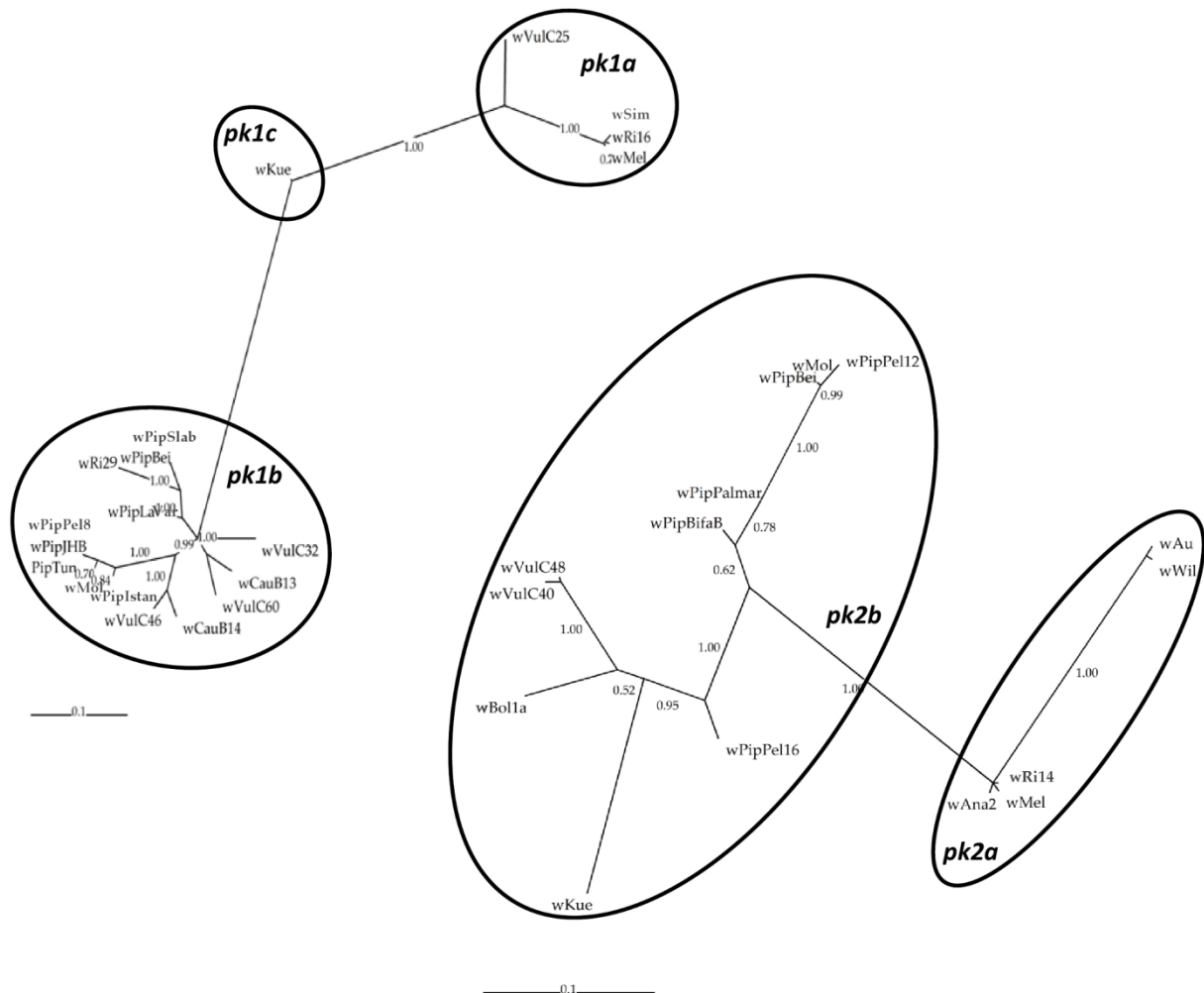


Figure 1: Unrooted ML phylogenetic trees of *pk1* and *pk2* genes.

Unrooted ML phylogenetic trees of *pk1* (on left) and *pk2* (on right) sequences encoding ANK motif clusters. *pk1* alleles are identified by their *Wolbachia* strains and ANK number (details in Table 1). Node supports are shown by bootstraps and by posterior probabilities from Bayesian inferences above each branch. Redundant sequences were not incorporated in the analyses.

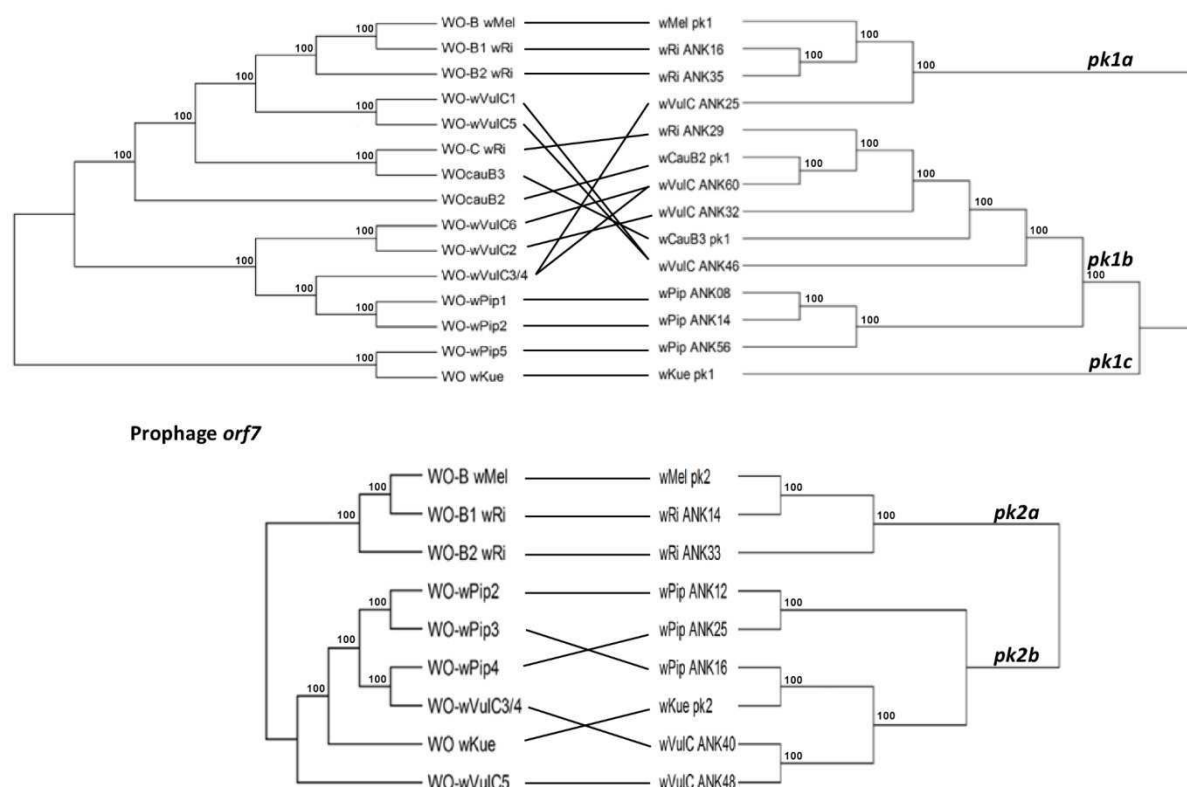


Figure 2: Rooted NJ phylogenetic trees of *pk1*, *pk2* and *orf7* genes.

Rooted NJ phylogenetic trees of *pk2* and *pk1* alleles (on right), respectively, with phage association (on left). *Wolbachia* strains are identified by their prophage regions (details in Table 1). Bootstrap values (based on 1,000 replicates) are shown on branches (%). Bootstrap scores lower than 50% are not given.

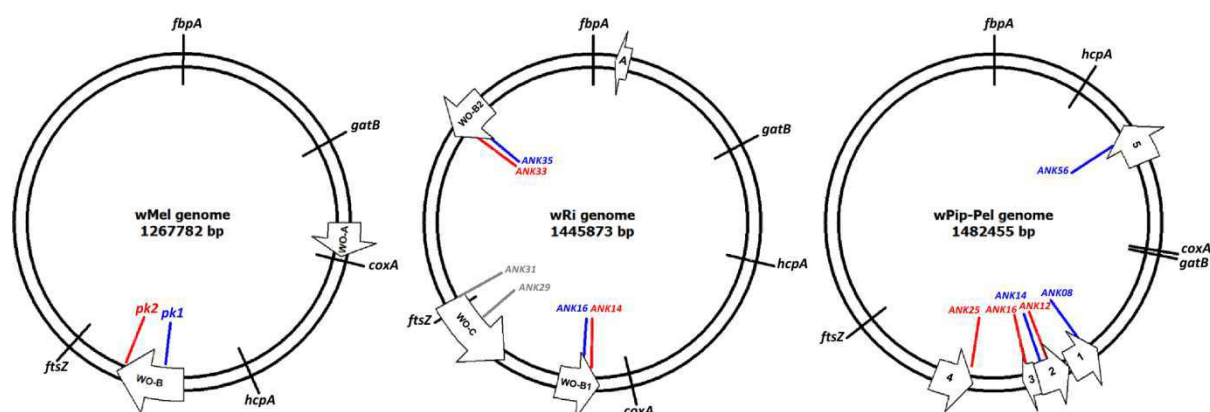


Figure 3: Schematic genomic maps of prophage regions, ankyrin *pk1* and *pk2* alleles in *Wolbachia* genomes.

Schematic genomic maps of prophage regions (orientated open boxes), phage-associated ankyrin *pk1* (blue) and *pk2* (red) alleles and classical MLST markers (*coxA*, *fbpA*, *ftsZ*, *gatB*, *hcpA* shown in black) in the *wMel*, *wRi* and *wPip-Pel* *Wolbachia* genomes represented as circles. For this purpose, genome origin is arbitrarily the *fbpA* gene. *pk1* alleles shown in grey are pseudogenes from the *wRi* genome. Details on genes and prophages are given in Table 1. Genomic maps were generated using the freeware program pDRAW v3.2 (www.acaclone.com). bp, genome size in base pairs.

Tables

Clade	Host species	<i>Wolbachia</i> strain	Genome	Phage	orf7	<i>pk1</i>	<i>pk2</i>
A	<i>Drosophila melanogaster</i>	w Mel	NC_002978	WO-B	WD0602	WD0596	WD0636
A	<i>Drosophila simulans</i>	w Ri	NC_012416	WO-B1	WRi_005560	WRi_005620	ANK16
				WO-B2	WRi_010220	WRi_010280	ANK35
				WO-C	WRi_007170	WRi_p07220	ANK29
				WO-C		WRi_p07650	ANK31
A	<i>Ephesia kuehniella</i>	w Kue	AB036666	WO	gp18	gp14	gp29
B	<i>Cadra cautella</i>	w CauB	AB478515	WOcauB2	B2gp17	B2gp13	
			AB478516	WOcauB3	B3gp18	B3gp14	
B	<i>Culex pipiens quinquefasciatus</i>	w Pip-Pel	NC_010981	WO-wPip1	WPa_0252	WPa_0256	ANK08
				WO-wPip2	WPa_0311	WPa_0315	ANK14
				WO-wPip3	WPa_0326		WPa_0340
				WO-wPip4	WPa_0426		WPa_0413
				WO-wPip5	WPa_1303	WPa_1308	ANK56
B	<i>Armadiillidium vulgare</i>	w VulC		WO-wVulC1	HM452366	HM452371	ANK46b
				WO-wVulC2	HM452369	HM452372	ANK32
				WO-wVulC3	HM452368	HM452373	ANK60b
				WO-wVulC4	HM452368	HM452374	ANK25
				WO-wVulC5	HM452367	HM452371	ANK46a
				WO-wVulC6	HM452370	HM452373	ANK60a

Table 1: List of identified capsid *orf7* as well as ankyrin *pk1* and *pk2* alleles according to their associated phages, in different sequenced and annotated *Wolbachia* prophages.

Accession numbers from this study are in bold.

<i>Wolbachia</i> endosymbiont of	<i>pk2</i> alleles		<i>pk1</i> alleles		
	a	b	a	b	c
<i>Armadiillidium nasatum</i>	0	3	1	5	0
<i>Armadiillidium vulgare</i> w VulC	0	3	1	5	0
<i>Armadiillidium vulgare</i> w VulM	0	3	1	5	0
<i>Oniscus asellus</i>	0	3	1	3	0
<i>Porcellionides pruinosus</i>	0	3	1	5	0
<i>Cylisticus convexus</i>	0	3	1	3	0
<i>Porcellio dilatatus dilatatus</i>	0	3	1	4	0
<i>Porcellio dilatatus petiti</i>	0	3	1	1	0
<i>Drosophila melanogaster</i>	w Mel	1	0	1	0
<i>Drosophila simulans</i>	w Ri	2	0	3(2)	1
<i>Culex pipiens</i>	w Pip-Pel	0	3	0	3
<i>Ephesia kuehniella</i>	w Kue	0	1	0	0
<i>Cadra cautella</i>	w CauB	0	0	0	2

Table 2: Copy number of *pk1* and *pk2* alleles according to their “type” in different *Wolbachia* strains.

In brackets, is indicated number of pseudogenes.

Phenotype	<i>Wolbachia</i> endosymbiont of	sex	<i>pk2</i> alleles			<i>pk1</i> alleles		
			a	b1	b2	a	b	c
fem	<i>Armadillidium nasatum</i>	F	na	+	+	+	+	na
fem	<i>Armadillidium vulgare</i> wVulC	F	na	+	+	+	+	na
fem	<i>Armadillidium vulgare</i> wVulM	F	na	+	+	+	+	na
fem	<i>Oniscus asellus</i>	F/M	na	+	+	+	+	na
fem	<i>Porcellionides pruinosus</i>	F/M	na	+	+	+	+	na
CI	<i>Cylisticus convexus</i>	F/M	na	+	-	+	+	na
CI	<i>Porcellio dilatatus dilatatus</i>	F/M	na	+	-	+	+	na
CI	<i>Porcellio dilatatus petiti</i>	F/M	na	+	-	+	+	na

Table 3: Expression patterns obtained by RT-PCR for *pk1* and *pk2* alleles.

Plus or minus signals indicate expression, or not, of the copy(ies). Distinction is made between the two different *pk2* alleles named *b1* and *b2* within the *pk2b* type. Fem: feminization; CI: cytoplasmic incompatibility. F: female; M: male. na: no *pk1c* or *pk2a* type alleles were amplified in these strains.

Supporting Information

Gene	Use	Forward primer		Reverse primer		PCR product size (bp)
		Name	Sequence (5'-3')	Name	Sequence (5'-3')	
<i>16S rDNA</i>	PCR/RT-PCR	99F	TTGTAGCCTGCTATGGTATAACT	994R	GAATAGGTATGATTTTCATGT	900
<i>pk1a</i>	RT-PCR	Ank-7F2	CACTGATCACTGATGATGATGG	Ank-7R3	GCATTTCTGCCGCTGCCAAC	410
	PCR/SB	Ank-7F	ATGTCTGCTAACTTAAGTTTAG	Ank-7R2	TTAAATTAGGACTACTTGTTTTTC	1500
<i>pk1b</i>	RT-PCR	Ank-1F	ATGGCTGATACTTAAGTTTAG	Ank-1R	GTCAAATTTTAGTAAAGAAGG	430
		Ank-16F2	GGGCATGCTCTGCAGAGAC	Ank-16R	TATGAGTATATGCAGACAATGA	200
	PCR	Ank-16F	ATGCACTATGCAGTTAACAAC	Ank-16R	TATGAGTATATGCAGACAATGA	1000
	PCR/SB	Ank-1F	ATGGCTGATACTTAAGTTTAG	Ank-16R	TATGAGTATATGCAGACAATGA	1000
<i>pk2b1</i>	PCR/RT-PCR	Ank-8F2	GTGGGGTGGATTATGTTAAAG	Ank-8R	GAAAAGTTATTTATGAAAGGCTAA	530
<i>pk2b2</i>	PCR/RT-PCR/SB	Ank-8F	ATGTTAAAGCTTGCCAAATTTA	Ank-8R2	GAAGATTTCCCACTTTGCTGTT	500

Table S1: List of primers used in this study for sequencing (PCR) or for expression analyses (RT-PCR). Expected PCR product size in base pair (bp) was calculated relative to the wVulC reference sequences.

pk1

Clade	Host species	<i>Wolbachia</i> strain	Name	Genome/Locus	pk1	length (bp/aa)	G+C content
A	<i>Drosophila melanogaster</i>	w Mel	wMel	NC_002978	WD0596	1482/493	0.34
A	<i>Drosophila willistoni</i>	w Wil	wRi16	NZ_AAQP01000060		1482/493	0.34
A	<i>Drosophila simulans</i>	w Sim	wSim	NZ_AAQC01000020		948/315	0.35
A	<i>Drosophila simulans</i>	w Au	wRi16	AY971760		1369/456	0.34
A	<i>Drosophila simulans</i>	w Ri	wRi16	NC_012416	WRi_005620	1482/493	0.34
			wRi29		WRi_p07220	1128/375	0.34
			wRi16		WRi_010280	1482/493	0.34
A	<i>Ephesthia kuehniella</i>	w Kue	wKue	AB036666	gp14	1482/493	0.33
B	<i>Cadra cautella</i>	w CauB	wCauB13	AB478515	B2gp13	1494/497	0.33
			wCauB14	AB478516	B3gp14	1485/494	0.34
B	<i>Culex molestus</i>	w Mol	wMol	DQ000470		554/185	0.36
B	<i>Culex pipiens quinquefasciatus</i>	w Pip-Bei	wPipBei	DQ000469		765/255	0.35
B	<i>Culex pipiens</i>	w Pip-Slab	wPipSlab	AM397075	pk1a	1328/443	0.34
		w Pip-Tunis	wPipTunis	AM397076	pk1b	1349/450	0.34
		w Pip-LaVar	wPipLaVar	AM397077	pk1c	1349/450	0.34
		w Pip-Istanbul	wPipIstan	AM397078	pk1d	1343/448	0.34
		w Pip-KaraC	wPipLaVar	AM397079	pk1e	1328/443	0.34
B	<i>Culex pipiens quinquefasciatus</i>	w Pip-JHB	wPipJHB	ABZA01000004		942/313	0.35
B	<i>Culex pipiens quinquefasciatus</i>	w Pip-Pel	wPipPel8	NC_010981	WPa_0256	1473/490	0.34
			wPipPel8		WPa_0315	1473/490	0.34
			wPipPel8		WPa_1308	1473/490	0.34
B	<i>Armadillidium vulgare</i>	w VulC	wVulC25	HM452374	ANK25	1428/475	0.33
			wVulC32	HM452372	ANK32	1485/494	0.34
			wVulC46	HM452371	ANK46a/b	1350/449	0.35
			wVulC60	HM452373	ANK60a/b	1290/429	0.34

pk2

Clade	Host species	<i>Wolbachia</i> strain	Name	Genome/Locus	pk2	length (bp/aa)	G+C content
A	<i>Drosophila melanogaster</i>	w Mel	wMel	NC_002978	WD0636	459/152	0.38
A	<i>Drosophila ananassae</i>	w Ana	wAna2	NZ_AAQB01000292		435/144	0.39
			wRi14	NZ_AAQB01000103		459/152	0.41
A	<i>Drosophila willistoni</i>	w Wil	wWil	NZ_AAQP01000027		339/112	0.38
A	<i>Drosophila simulans</i>	w Au	wAu	AY649752		330/109	0.37
A	<i>Drosophila simulans</i>	w Ri	wRi14	NC_012416	WRi_005440	459/152	0.37
			wRi14		WRi_010100	459/152	0.38
A	<i>Ephesthia kuehniella</i>	w Kue	wKue	AB036666	gp29	486/161	0.38
	<i>Hypolimnas bolina</i>	w Bol1a	wBol1a	FJ842506		465/154	0.38
B	<i>Culex molestus</i>	w Mol	wMol	DQ000472		450/149	0.37
B	<i>Culex pipiens</i>	w Pip-Bei	wPipBei	DQ000471		450/149	0.38
		w Pip-BifaB	wPipBifaB	AM397073	pk2c	450/149	0.38
		w Pip-PalmarA	wPipPalmar	AM397074	pk2d	450/149	0.38
B	<i>Culex pipiens quinquefasciatus</i>	w Pip-JHB	wPipPel12	NZ_ABZA01000004		450/149	0.38
			wPipPel12	NZ_ABZA01000005		450/149	0.37
			wPipPel16	NZ_ABZA01000004(2)		474/157	0.38
B	<i>Culex pipiens quinquefasciatus</i>	w Pip-Pel	wPipPel12	NC_010981	WPa_0299	450/149	0.38
			wPipPel16		WPa_0340	474/157	0.38
			wPipPel12		WPa_0413	450/149	0.37
B	<i>Armadillidium vulgare</i>	w VulC	wVulC40	HM452376	ANK40a/b	498/165	0.36
			wVulC48	HM452375	ANK48	474/157	0.40

Table S2: List of sequences used for Figure 1 and Figure S2. Redundant sequences are under a unique sequence name. Accession numbers from this study are in bold.

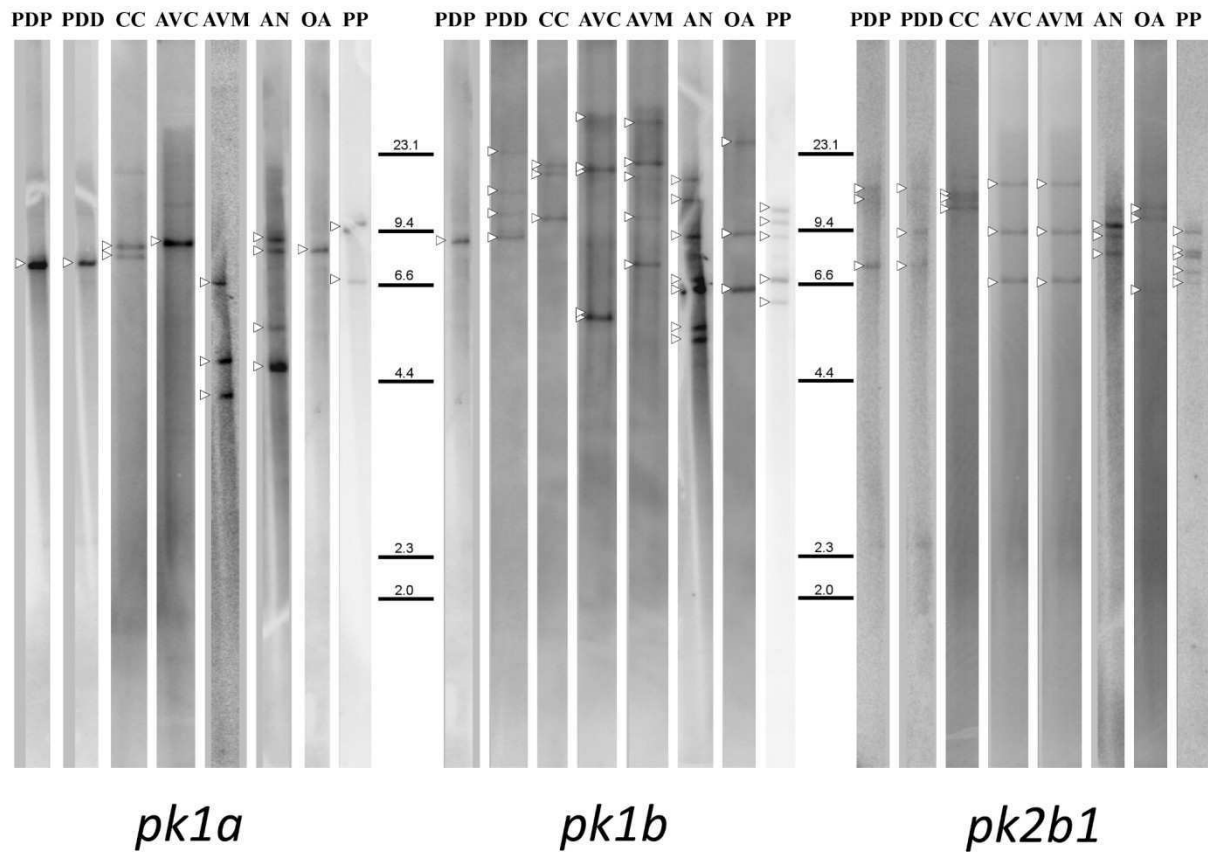


Figure S1: Southern blotting analyses.

Southern blotting of *Eco*RI or *Bam*HI digested DNA from 8 *Wolbachia*-infected terrestrial isopod species based on three probes (see text for details). White triangles highlight the positions of the hybridized fragments. Lines were loaded with DNA from *Wolbachia* strain endosymbionts of PDP as *P. dilatatus petiti*; PDD as *P. dilatatus dilatatus*; CC as *C. convexus*; AVC as *A. vulgare* strain wVulC; AVM as *A. vulgare* strain wVulM; AN as *A. nasatum*; OA as *O. asellus*; PP as *P. pruinosis* strain wPruIII. Number of bands in each line does not necessarily correspond to figures in Table 2 because of restriction site presence in copies and further confirmed by sequencing.

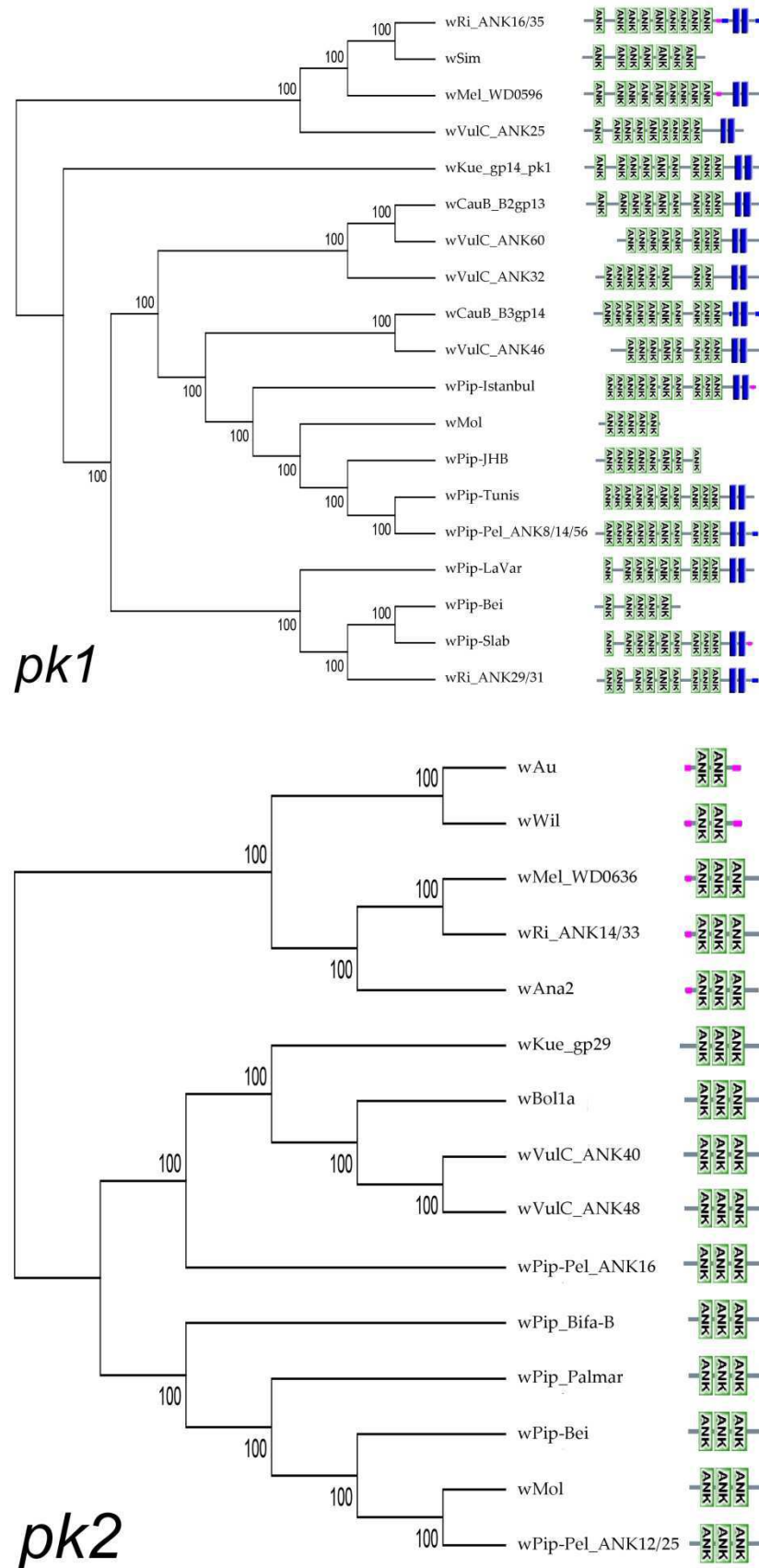


Figure S2: SMART outputs representing the number of ANK motifs found in Pk1 and Pk2 translated sequences.

CHAPITRE IV :

INTERACTION T4SS /

EFFECTEURS DE *WOLBACHIA*.

Technique	Type de T4SS	Effecteur ou Interaction	Références
Genetic complementation			
DnaG Ts mutation	Collb-P9	sog primase	Wilkins & Thomas 2000
Mixed infection	VirB	VirE2, VirF	Ward & Zambryski 2001
Transgenic plants	VirB	VirE2, VirF	Ward & Zambryski 2001
Reporter protein fusions			
Cre recombinase (CRAFT)	VirB	VirE2, VirF, VirE3	Vergunst <i>et al.</i> 2000
YopP	VirB	BvfA	Lavigne <i>et al.</i> 2005
Adenylate cyclase (CyaA)	Dot/Icm	many	<i>in</i> Burstein <i>et al.</i> 2009
Complex formation with T4SS			
<i>In vitro</i> with purified protein	RP4 Tra	TraG-TraI, TraG-pBHR1	Schroder <i>et al.</i> 2002, Szpirer <i>et al.</i> 2000
	F Tra	TraD-TraM	Disque-Kochem & Dreiseikelmann 1997
Bacterial two-hybrid	RP4 Tra	TraG-Mob (pBHR1)	Szpirer <i>et al.</i> 2000
Cytology-based two-hybrid	VirB	VirD4-VirE2	Atmakuri <i>et al.</i> 2003
Substrate-dependent assembly	Dot/Icm	LidA	Conover <i>et al.</i> 2003
Complex formation with chaperone			
Biochemical-IP or pull-down	VirB	VirE1-VirE2	Zhao <i>et al.</i> 2001
	Dot/Icm	IcmR-IcmQ, IcmS-P130	Coers <i>et al.</i> 2000
Yeast two-hybrid	VirB	VirE1-VirE2	Sundberg & Ream 1999, Zhou & Christie 1999
	Dot/Icm	IcmR-IcmQ, IcmS-IcmW	Coers <i>et al.</i> 2000
Localisation			
Milieu	Ptl	PT	<i>in</i> Burns 2003
	Dot/Icm	DotA	Nagai <i>et al.</i> 2001
	VirB	VirE2, VirF	Chen <i>et al.</i> 2000, Pantoja <i>et al.</i> 2002
Translocation-dependent modification			
Tyrosine-phosphorylation	Cag	CagA	Backert <i>et al.</i> 2002

Figure 19: Méthodes utilisées pour identifier des effecteurs bactériens *in vivo*.

IV. Interaction T4SS / effecteurs de *Wolbachia*.

A. Objectifs.

A ce jour, plusieurs travaux ont permis de mettre en évidence chez des endosymbiotes bactériens la sécrétion *in vivo* d'effecteurs vers le cytoplasme de la cellule hôte *via* un T4SS (Figure 19). Les techniques moléculaires développées dans ce but comme par exemple le double-hybride ou le CRAfT (Cre recombinase Reporter Assay for Translocation) ont servi à démontrer la translocation d'un nombre important de déterminants bactériens (Figure 19). Ces deux techniques en particulier s'appuient donc sur l'hypothèse selon laquelle les effecteurs bactériens doivent interagir avec au moins un des composants du T4SS à un instant donné avant d'être sécrétés par cette structure (Cascales et Christie 2003). La technique dite de CRAfT, chez la bactérie intracellulaire *A. phagocytophilum*, proche parent de *Wolbachia*, a ainsi permis de démontrer la sécrétion d'une protéine à domaines ankyrines, AnkA, par un T4SS (Lin *et al.* 2007).

Nous avons adopté une double stratégie pour rechercher les effecteurs bactériens de *wVulC*. Nous avons tout d'abord ciblé des candidats potentiels et testé leurs interactions avec des protéines de T4SS par double-hybride en levures et par CRAfT. Nous avons ensuite réalisé plusieurs expériences préliminaires dans le but d'élaborer une méthode de criblage d'effecteurs bactériens plus exhaustive à partir d'une banque d'ADNc de *wVulC*.

B. Interaction T4SS / effecteurs ciblés de *wVulC*.

Au vu des arguments présentés dans le chapitre II, nous avons voulu tester l'hypothèse selon laquelle les protéines à domaines ankyrines seraient des effecteurs de *Wolbachia* sécrétés par un T4SS. Nous avons donc ciblé dans un premier temps les gènes ankyrines *pk1* et *pk2* pour lesquels nous possédions également des données d'expression. Nous avons également ciblé un autre candidat à la sécrétion par le T4SS de *Wolbachia*, le gène de la phosphatase (*phos*) situé en amont de l'opéron *virB8-D4*. Dans certains cas, et par analogie avec d'autres systèmes comme ceux présents chez *H. pylori* et *B. pertussis*, les gènes codant des protéines effectrices sont, sur le génome, proches des gènes codant le système de sécrétion.

Le gène codant CagA chez *H. pylori* est par exemple localisé au niveau de l'îlot de pathogénicité « *cag* » constitué des gènes du T4SS (Cover *et al.* 1990, Tummuru *et al.* 1995, Censini *et al.* 1996). Chez *B. pertussis*, les gènes codant les 5 sous unités de la toxine pertussique (opéron *ptx*) ont été localisés 55 pb en amont de l'opéron *ptl* regroupant les gènes du T4SS (Nicosia *et al.* 1986).

B.1 - Test par double-hybride chez la levure.

Cette technique permet la détection d'interactions protéine/protéine *in vivo*, chez la levure. Elle a été initialement utilisée pour identifier les effecteurs exportés par un T4SS, dont ceux d'*A. tumefaciens* par exemple (Figure 19). Développé par Fields et Song en 1989, ce système repose sur les propriétés du facteur de transcription Gal4 qui possède deux domaines fonctionnels : un domaine activateur de la transcription et un domaine de fixation à l'ADN. Lorsque ces deux domaines sont séparés, le facteur de transcription n'est plus capable d'induire l'activation de la transcription. Les gènes étudiés sont clonés dans un vecteur de levure permettant leur expression sous forme de protéine de fusion avec l'un et l'autre des domaines du facteur de transcription. L'interaction physique entre les protéines d'intérêt permet alors de reconstituer un activateur entier et fonctionnel capable de stimuler la transcription de gènes rapporteurs (Fields et Song 1989, Chien *et al.* 1991). Cette technique est simple, extrêmement sensible, et permet ultérieurement de délimiter facilement les domaines d'interaction entre les deux protéines (Ding *et al.* 2002).

En plus des protéines cibles ou proies, il faut déterminer les protéines servant d'appâts pour la sécrétion *via* le T4SS. Des approches biochimiques et génétiques (systèmes de double hybride) chez *A. tumefaciens* ont démontré que les protéines, avant d'être sécrétées *via* le T4SS, interagissaient avec une «coupling protein», VirD4, localisée sur la membrane interne de la bactérie (Lin et Kado 1993, Hamilton *et al.* 2000, Atmakuri *et al.* 2003, Atmakuri *et al.* 2004, Cascales *et al.* 2005). Cette protéine sert au recrutement des effecteurs dans le sens où elle permet de les diriger vers le T4SS, au niveau de la protéine VirB10 (Llosa et O'Callaghan 2004). Les protéines VirB4 et VirB11 sont ancrées dans la membrane interne et exposées au cytoplasme de la bactérie. Ces deux protéines ont des fonctions analogues à la protéine VirD4 d'*A. tumefaciens* (Atmakuri *et al.* 2004). En effet, les deux ATPases (VirB4, VirB11) assurent l'apport énergétique nécessaire pour l'assemblage du complexe ainsi que le transport de substrats à travers le pore (Fullner *et al.* 1994, Rashkova *et al.* 1997, Watarai *et al.* 2002, Savvides *et al.* 2003, Atmakuri *et al.* 2004, Middleton *et al.* 2005, Paschos *et al.* 2006). Elles apparaissent comme des moteurs potentiels du système de transport utilisant l'énergie libérée par l'hydrolyse de l'ATP (Stephens *et al.* 1995, Watarai *et al.* 2002).

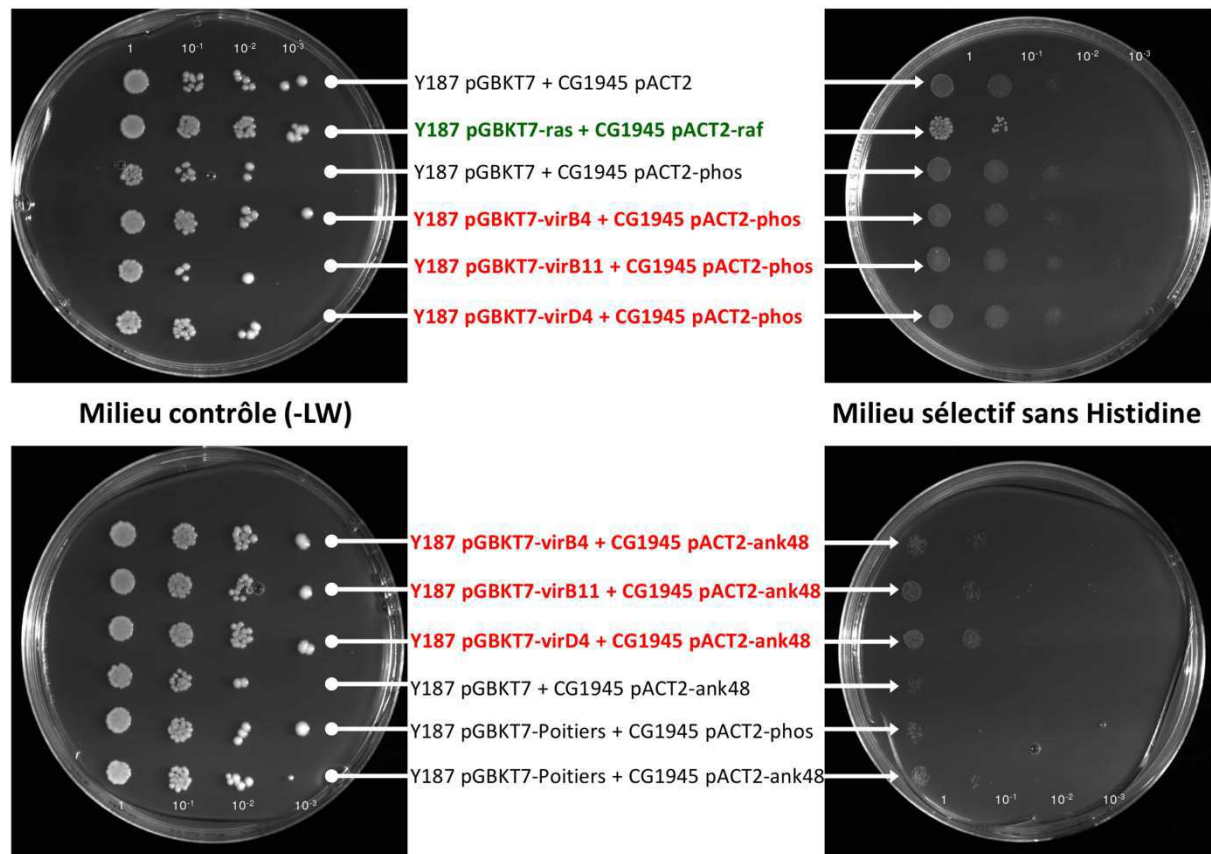


Figure 20: Résultats du double-hybride en levures.

Résultats des tests d'interaction *in vivo* par double-hybride en levures entre les protéines candidates (ici Phos et wVulC_ANK48) et les protéines VirB4, -B11 et VirD4 du T4SS de wVulC.

Etant donné que VirD4, VirB4 et VirB11 sont connus pour interagir avec les toxines bactériennes (Atmakuri *et al.* 2003), nous avons donc choisi d'utiliser ces trois protéines entières en tant qu'appâts en double-hybride pour identifier les effecteurs éventuels du T4SS chez *wVulC*. Nous avons donc cloné les gènes codant les différents allèles de *pk1* (*wVulC*_ANK25, ANK46 et ANK60) et de *pk2* (*wVulC*_ANK40, ANK48), ainsi que le gène codant la phosphatase, dans le vecteur pACT2, en phase avec le domaine d'activation de Gal4 (AD, Activation Domain) [cf. Annexe 1, p. 154 pour le détail des protocoles utilisés dans les expériences décrites de ce chapitre]. Les gènes codant les protéines VirB4, VirB11 et VirD4 ont quant à eux été clonés dans le vecteur pGBKT7 en phase avec le domaine d'interaction à l'ADN de Gal4 (DNA-BD, DNA-binding Domain). Les séquences de chaque gène ont été amplifiées par PCR avec une polymérase haute-fidélité pfu (Promega) puis transformées chez les bactéries. Les clones sélectionnés ont ensuite été utilisés pour la transformation de levures. Le séquençage des produits transformés a permis de vérifier le respect du cadre de lecture avec la séquence fusionnée. Nous avons ensuite procédé aux tests d'interaction entre les composants du T4SS et les protéines d'intérêt de *wVulC* en comparaison avec deux contrôles négatifs constitués par les vecteurs pGBKT7 et pACT2 vides et un contrôle positif d'interaction pGBKT7-ras/pACT2-raf.

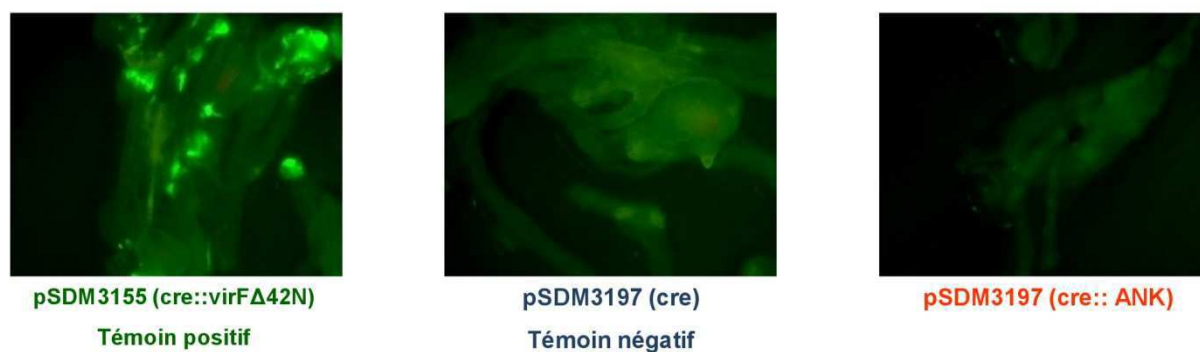
Afin de sélectionner les levures transformées par les clones de pACT2, les souches ont été cultivées en absence de tryptophane (Trp-), celles transformées par pGBKT7 en absence de Leucine (Leu-). Nous avons ensuite réalisé des doubles transformants sélectionnés dans un premier temps sur milieu sans Tryptophane et sans Leucine. Ces doubles transformants ont été repiqués sur milieu sélectif Trp-, Leu-, et sans Histidine (His-) afin de mettre en évidence une interaction entre les deux protéines recombinantes. Aucun clone mis à part le témoin positif Ras-Raf ne s'est développé (Figure 20). Il est possible au vu des résultats qu'aucune des cibles potentielles n'interagisse avec les protéines du T4SS. Il faudrait néanmoins s'assurer que les levures soient capables d'exprimer la protéine codée par le gène cloné. La vérification de cette expression pourra être réalisée par Western blot à l'aide d'anticorps dirigés contre les différents domaines de Gal4. Par ailleurs, des problèmes de toxicité, de taux d'expression, de stabilité ou d'encombrement stérique pour des protéines de hauts poids moléculaires (telles VirB4) sont connues pour empêcher la détection de certaines interactions (Vidal et Legrain 1999). Un problème d'adressage au noyau peut, de la même façon, être rencontré pour des protéines de taille importante (Fields et Sternglanz 1994). En effet, la localisation nucléaire des deux partenaires est requise pour les tests d'activation transcriptionnelle. Ainsi, aucune interaction ne pourra être détectée si l'un des deux partenaires ne peut accéder au noyau. Ceci peut être le cas si une des protéines testées telle VirB4 est membranaire et reste localisée dans le cytoplasme en interaction avec les membranes. Un transport nucléaire inefficace peut donc être à l'origine de résultats négatifs.

Il a aussi été montré que des interactions protéine-protéine qui ne sont pas identifiées en double-hybride avec des protéines entières peuvent potentiellement être mises en évidence en utilisant des peptides plus petits (Ward *et al.* 2002). Il faudrait alors tester des interactions entre protéines Vir elles-mêmes, puisqu'elles interagissent entre elles, pour mettre au point cette technique. Cependant, il est connu que les sites d'interactions intervenant entre protéines Vir ne sont pas les mêmes que ceux utilisés pour les interactions avec les candidats à la sécrétion (Atmakuri *et al.* 2004). De plus, les systèmes de sécrétion sont des systèmes dynamiques, des interactions temporaires ou transitoires peuvent ne pas être identifiées avec cette technique (Vidal et Legrain 1999, Ward *et al.* 2002).

B.2 - Test par CRAfT.

La technique de CRAfT a été mise en place récemment pour démontrer la sécrétion d'effecteurs de bactéries intracellulaires *via* un T4SS. Cette technique permet de tester la sécrétion d'effecteurs potentiels *via* le T4SS d'*A. tumefaciens* (Lin *et al.* 2007). Ce dernier est souvent considéré comme l'archétype du T4aSS des bactéries et est homologue des opérons *vir* de *wVulC*. Fort de la démonstration récente de la sécrétion d'AnkA chez *A. phagocytophilum* (Lin *et al.* 2007) et sur les conseils de D. O'Callaghan, nous avons contacté Annette Vergunst puis ses anciens collègues néerlandais (Sylvius Laboratory de P.J.J Hooykaas, Université de Leiden, Pays-Bas) qui ont mis au point cette technique. Nous avons alors réalisé le clonage des différents gènes cibles (*pk1*, *pk2* et *phos*) dans le vecteur pSDM3197 fourni par A. den Dulk-Ras et P.J.J Hooykaas [*cf.* Annexe 2, p. 159]. Nous leur avons envoyés les plasmides transformés après vérification du respect du cadre de lecture des séquences insérées en fusion avec le gène *cre*. Les plasmides ont ensuite été transformés dans le laboratoire hollandais par électroporation dans la souche LBA1100 d'*A. tumefaciens*. Les premiers tests de sécrétion ont alors été réalisés. Ils requièrent l'infection par *A. tumefaciens* d'une plante transgénique *Arabidopsis thaliana* au niveau de ses racines. Cette plante transgénique C24 lignée pCB1 contient le gène codant la GFP (Green Fluorescent Protein) dont la transcription est bloquée par une séquence cible de l'enzyme située entre le promoteur 35S et le codon « start » du gène codant la GFP (Van Duyne 2001). La méthode de CRAfT repose sur les propriétés de la Cre recombinase, enzyme qui catalyse les réarrangements de l'ADN et dont la séquence est fusionnée aux gènes d'intérêt dans le vecteur pSDM3197. Cette enzyme se fixe sur des sites spécifiques loxP et induit des remaniements génomiques tels l'excision (Van Duyne 2001). Une fois les cellules de la plante infestées par *A. tumefaciens*, la sécrétion de la protéine d'intérêt par la bactérie est déduite si la Cre recombinase vient cliver la séquence cible en amont du gène *gfp* activant ainsi le gène.

Test de translocation des protéines d'intérêt par *A. tumefaciens* dans les racines d'*A. thaliana* pCB1



Analyse de l'expression des protéines d'intérêt par Western Blot

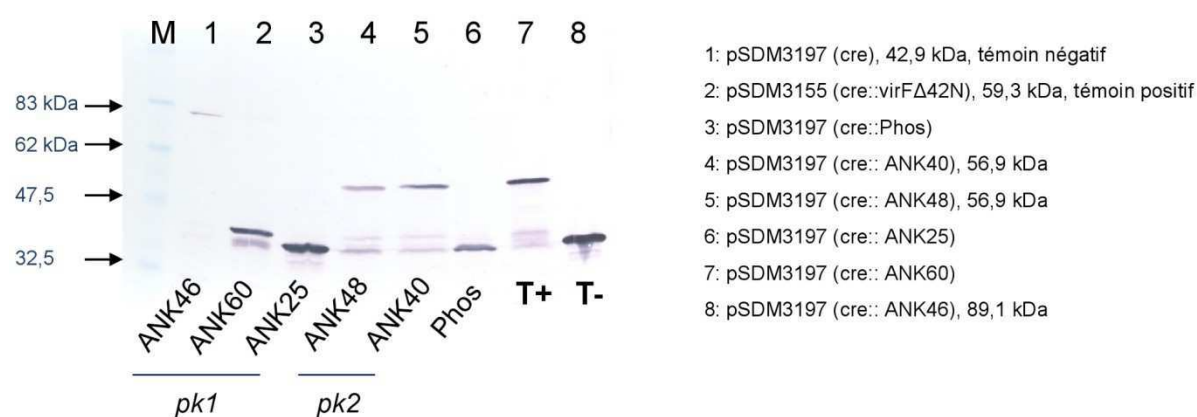


Figure 21: Résultats du CRAfT chez *Agrobacterium tumefaciens*.

Résultats des tests de translocation de protéines candidates de *wVulC* par le T4SS d'*A. tumefaciens* dans les racines d'*A. thaliana*. L'expression des protéines d'intérêt fusionnées à la séquence codant la Cre recombinase a été secondairement testée par Western Blot.

La production de la GFP est ensuite vérifiée par microscopie en épifluorescence. Les tests de translocation par le T4SS d'*A. tumefaciens* ont été réalisés avec les clones contenant les mêmes inserts que ceux utilisés pour les tests de double hybride: *pk1* (*wVulC*_ANK25, ANK46 et ANK 60), *pk2* (*wVulC*_ANK40 et ANK48) et *phos*. Le vecteur pSDM3197 vide (ne possédant que la séquence *cre*) a été utilisé comme contrôle négatif et le vecteur pSDM3155 contenant la séquence codant VirF (*cre::virFΔ42N*), un effecteur connu d'*A. tumefaciens*, comme contrôle positif. Les résultats de microscopie montrent de nombreux points de fluorescence avec l'échantillon pSDM3155 (*cre::virFΔ42N*) et aucun avec le contrôle négatif ou les protéines d'intérêt de *wVulC* (Figure 21). Les protéines ANK et Phos semblent donc ne pas être exportées par le T4SS d'*A. tumefaciens*. Or, la technique du CRAfT suppose la reconnaissance des protéines de *Wolbachia* par le T4SS d'*A. tumefaciens*. Nous ne pouvons donc pas exclure l'hypothèse selon laquelle le T4SS d'*A. tumefaciens* est dans l'incapacité de reconnaître les protéines d'intérêt de *wVulC*. Les résultats obtenus par CRAfT confirment cependant les résultats obtenus en double-hybride. A. den Dulk-Ras a, dans un second temps, vérifié que les protéines de *wVulC* étaient correctement exprimées. Le Western blot réalisé montre l'expression des protéines de *wVulC* chez *A. tumefaciens* détectée avec l'anticorps Cre après induction avec de l'acétoseringone (Figure 21). Les pistes 1 et 2 du Western blot montrent respectivement l'expression des protéines Cre (témoin négatif, poids moléculaire PM = 42,9 kDa) et de la protéine de fusion *cre::virFΔ42N* (témoin positif, PM= 59,3 kDa) (Figure 21). Les protéines Phos et ANK25 (codées par *pk1a*) n'ont pas un PM supérieur à la Cre seule (pistes 3 et 6) (Figure 21). La présence d'un codon stop prématuré, entre la séquence de la Cre et celle de l'insert pourrait expliquer pourquoi la bande observée est de taille similaire du témoin négatif. Cependant, la séquence de chaque clone avait été soigneusement confirmée comme possédant une séquence sans codon stop et en fusion avec le gène *cre*. Nous ne pouvons pas exclure la possibilité de contamination des échantillons par des vecteurs ne contenant que la séquence *cre*. De nouveaux tests de translocation sont prévus après vérification à nouveau de la séquence des clones et des problèmes potentiels de contamination. La protéine ANK32 (piste 7) a elle un PM inférieur à celui attendu (~45 au lieu de 90,6 kDa) (Figure 21). Par contre, les PM de *cre::ANK46* et *cre::pk2* correspondent à ce qui était attendu (89,12 et 56,88 kDa respectivement) (Figure 21). Ainsi, l'absence d'expression des protéines ne permet pas d'expliquer l'ensemble de ces résultats négatifs.

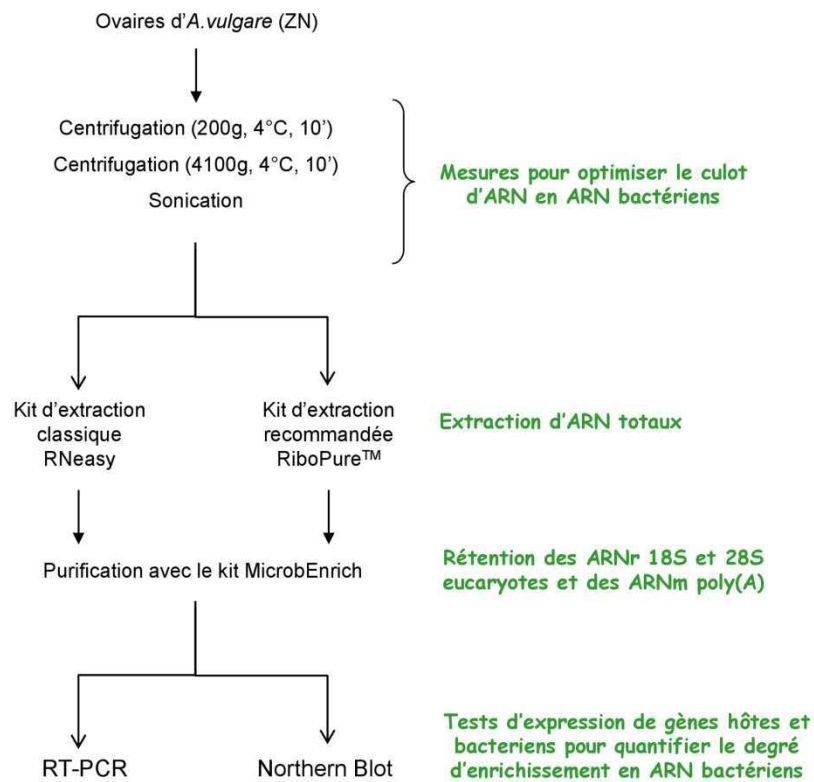


Figure 22: Protocole des étapes préliminaires à la construction d'une banque d'ADNc.

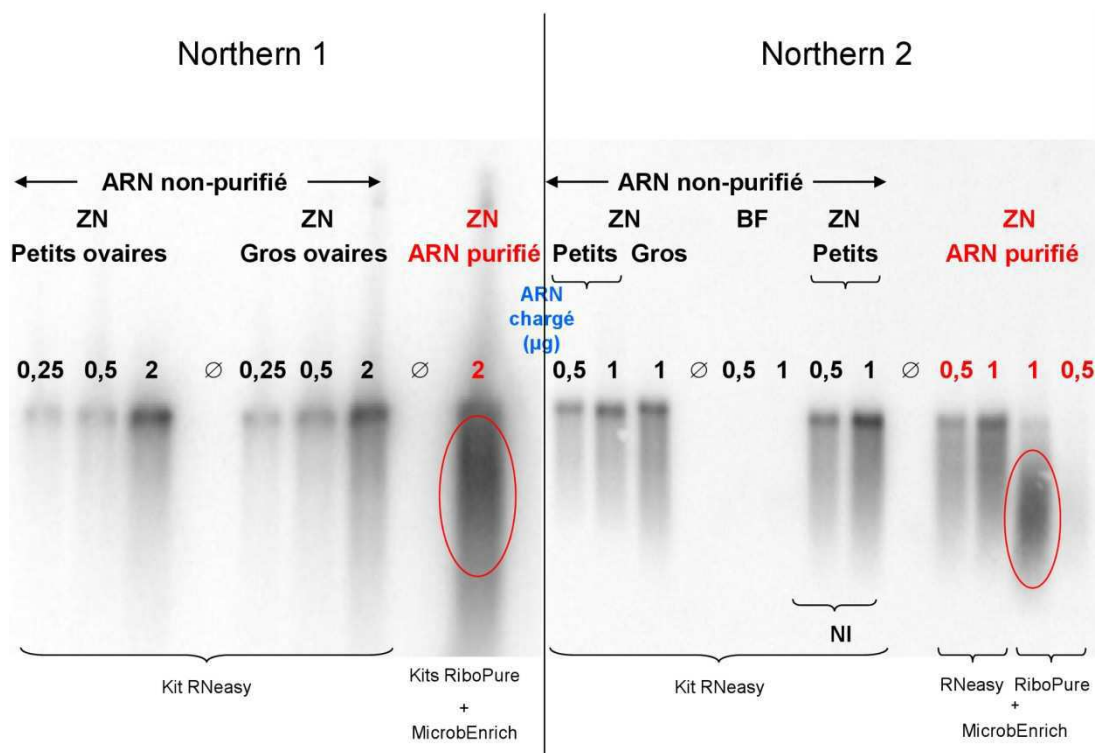


Figure 23: Résultats de l'enrichissement d'un culot d'ARN en ARN bactériens.

Les deux Northern Blots donnent les profils obtenus avec le gène 16S ribosomal de *Wolbachia* en fonction des kits d'extraction et de purification des ARN. *A. vulgare* (ZN) est la population de référence infectée par *wVulC*, *A. vulgare* (BF) est une population non infectée (NI). Ø = piste sans ARN.

C. Elaboration d'une banque d'ADNc de *wVulC*.

Le deuxième volet de ces expérimentations *in vivo* consistait à mettre en place une technique permettant une recherche plus exhaustive des facteurs transitant par le T4SS. Selon le même principe, nous envisagions de cribler par double-hybride une banque enrichie en ADNc de *wVulC* avec les trois mêmes appâts Vir du T4SS de *Wolbachia*. Ainsi, nous avons réalisé les premières étapes préparatoires de la construction d'une banque d'ADNc bactériens en vue, cette fois, de mettre en évidence des interactions entre VirB4, VirD4 ou VirB11 et d'autres protéines bactériennes. La préparation de culots enrichis en ARN bactériens était une étape clé dans la construction de la banque d'ADNc bactériens. Il était indispensable de séparer au mieux les ARN de *Wolbachia* des ARN de l'hôte afin de réaliser une banque la plus enrichie possible en ADNc bactériens.

L'enrichissement d'ARN bactériens a été réalisé avec le kit MicrobEnrich (Ambion) à partir d'ARN totaux extraits suivant deux méthodes, soit avec un kit RiboPure (Ambion) conseillé pour l'utilisation du kit MicrobEnrich, soit à partir d'une extraction classique avec le kit Rneasy (Qiagen) (Figure 22) [cf. Annexe 3, p. 160]. A partir d'un extrait d'ARN totaux, MicrobEnrich est conçu pour hybrider, *via* l'utilisation de billes MagBeads, 90% des ARN ribosomiaux 18S et 28S ainsi que les ARN polyadénylés de l'hôte (Figure 22). Cette technique était alors adaptée puisque les ARN bactériens ne possèdent pas de queues polyA contrairement à ceux des eucaryotes. Aussi la fraction enrichie après utilisation du kit MicrobEnrich doit contenir des ARN bactériens en proportion plus importante. La quantification de la purification par RT-PCR et Northern Blot n'a pas permis de mettre en évidence un enrichissement du culot d'ARN en ARN bactériens (Figures 22 et 23). Les ARN obtenus étaient très souvent dégradés par l'utilisation combinée des kits RiboPure et MicrobEnrich (Figure 23). Nous avons donc réalisé les expériences suivantes à partir d'une extraction classique suivie d'une purification avec MicrobEnrich. Cette méthode nous a permis d'obtenir des ARN non dégradés (Northern 2, Figure 23). Cependant, les hybridations effectuées avec des marqueurs d'ARN eucaryotes n'ont pas permis de mettre en évidence d'enrichissement en ARN bactériens.

Peu de publications font état de l'utilisation de ces kits Ambion. La plupart des études où les résultats se sont avérés encourageants sont issues d'une purification d'ARN bactériens à partir de cellules de Mammifères infectées artificiellement et maintenues en culture cellulaire. Certaines publications admettent la nécessité d'une forte concentration en pathogènes et le manque de reproductibilité concernant la fraction d'ARN purifiés (Orihuela *et al.* 2004). Une autre étude fait état de la dégradation des ARN avec le kit RiboPure (Di Cello *et al.* 2005).

Le problème dans ce cas est que les ARN dégradés et non soustraits lors de la purification MicrobEnrich, rentrent en compétition avec les ARN bactériens pour la synthèse d'ADNc. Cependant, le kit MicrobEnrich a pu être utilisé pour purifier des cultures cellulaires infectées avec les bactéries du genre *Rickettsia* (La *et al.* 2007, Audia *et al.* 2008, Renesto *et al.* 2008). Plus fondamentalement, il est important de déterminer le stade au cours du développement des individus durant lequel l'expression des gènes bactériens est maximale.

D. Conclusions et perspectives.

Les différents problèmes énoncés pour la technique du double-hybride pourront être contournés par l'utilisation de protéines sous formes tronquées. De manière alternative, un autre système de double hybride, développé chez *E. coli*, pourra être utilisé. En effet, ce système a déjà permis de mettre en évidence des réseaux d'interactions entre les protéines effectrices et les protéines VirB de *Rickettsia sibirica* (Malek *et al.* 2004). Ce système pourrait être mieux adapté à l'étude des interactions entre des protéines bactériennes ou membranaires. Ce système est basé sur la reconstitution de l'adénylate cyclase qui conduit à la production d'AMPc. De ce fait, une fois exprimé chez une souche *cya* d'*E. coli*, l'AMPc va induire l'expression d'opérons qui contrôlent des gènes cataboliques (lactose, maltose). Contrairement au système chez la levure, où les interactions entre les protéines d'intérêt se produisent dans le noyau, les interactions peuvent avoir lieu aussi bien dans le cytoplasme qu'au niveau de la membrane interne (Karimova *et al.* 1998).

Par ailleurs, que ce soit par les techniques de CRAfT ou de double-hybride, les démonstrations de sécrétion ne restent qu'indirectes dans la mesure où la sécrétion par le T4SS tout entier n'est évidente qu'après confirmation de la translocation des effecteurs vers le cytoplasme de l'hôte. Classiquement, l'identification d'interactions par double-hybride est suivie par la confirmation par des méthodes biochimiques telles la co-immunoprécipitation, le « transfer DNA immunoprecipitation » (TrIP) ou par la technique du « GST pull down » permettant de mettre en évidence une interaction physique, *in vitro* cette fois, entre deux protéines.

CHAPITRE V :

**DISCUSSION GENERALE ET
PERSPECTIVES.**

V. Discussion générale et perspectives.

Du fait de son très vaste spectre d'hôtes, de la variété des phénotypes induits, de ses impacts écologique et évolutif, et de ses particularités qui échappent au modèle jusque là admis d'évolution des bactéries symbiotiques, *Wolbachia* est un modèle idéal d'étude intégrative des interactions eucaryotes-bactéries. Avant mon travail de thèse, deux génomes complets de *Wolbachia* (*wMel* et *wBm*) avaient été publiés offrant de nouvelles possibilités d'étudier les communications entre la l'endosymbiote et son hôte. Parallèlement, des articles de revue mettaient en avant l'intérêt porté sur les gènes codant un système de sécrétion de type IV (T4SS), ceux codant des protéines à domaines ankyrines (ANK) et ceux contenus dans les séquences de prophage (McGraw et O'Neill 2004, Fenn et Blaxter 2006). Dans ce contexte, mes objectifs de thèse étaient donc de caractériser les gènes du T4SS et d'identifier les effecteurs sécrétés par la souche *Wolbachia wVulC* féminisant les mâles d'*A. vulgare*.

Les éléments rapportés dans ce manuscrit confirment la présence de gènes codant un T4SS chez *wVulC*. Ceux-ci sont co-exprimés dans tous les tissus de l'hôte et s'organisent selon deux opérons distincts chez cette souche. Les analyses comparatives montrent que le T4SS de *Wolbachia* est conservé puisqu'il est composé de deux larges portions génomiques d'environ 20 kb communes à de nombreuses souches. Le T4SS de *Wolbachia* est de plus codé par un ensemble de gènes représentant plus que le minimum nécessaire à la formation fonctionnelle de cette structure. Cette approche évolutive suggère donc que *Wolbachia* puisse utiliser ce système pour la sécrétion vers le cytoplasme de l'hôte d'effecteurs nécessaires à la survie et la transmission de la bactérie. Cela a également permis de mettre en évidence les singularités du T4SS de *Wolbachia* par rapport à l'archétype du T4aSS chez *A. tumefaciens*. Le découpage en deux clusters de gènes ainsi que la présence de paralogues des gènes *virB4*, *-B8* et *-B9* dispersés dans les génomes constituent deux arguments en faveur d'une évolution toute particulière de ce système.

Les efforts menés pour identifier les effecteurs de *wVulC* se sont en outre focalisés sur la caractérisation d'une famille de gènes codant des protéines à domaines ankyrines (*ank*) et sur l'élaboration de tests permettant de savoir si ces dernières étaient effectivement sécrétées par le T4SS de la bactérie. Un nombre important de gènes *ank* (au minimum 66) est présent dans le génome en cours d'assemblage de *wVulC*. La présence d'autant de gènes, principalement connus pour coder des protéines impliquées dans les interactions protéine-protéine chez les eucaryotes, constitue une caractéristique tout à fait originale pour une bactérie.

Cette caractéristique est d'ailleurs partagée, bien qu'en moindre proportion, par toutes les autres souches de *Wolbachia* dont les génomes sont séquencés. Les arguments présentés dans ce mémoire suggèrent l'importance des protéines à motifs ankyrines dans la biologie de *Wolbachia* et en font des candidats à la sécrétion par la bactérie. L'accent a été mis sur la caractérisation génétique et transcriptionnelle des gènes prophagiques *pk1* et *pk2*. Les résultats montrent que ces deux gènes sont retrouvés en plusieurs copies dans le génome des souches infectant les isopodes terrestres et que leur séquence est conservée. Alors que l'ensemble des gènes identifiés est transcrit, une copie du gène *pk2* s'exprime uniquement chez les souches féminisantes hébergées par les isopodes mais aucunement chez les souches induisant l'IC chez les cloportes. Toutefois, les tests de double-hybride et de CRAFT n'ont pas permis de mettre en évidence l'interaction des produits de gènes *pk1* et *pk2* avec les protéines du T4SS de *wVulC*. Ceci suggère que ces protéines en particulier ne sont pas sécrétées par le T4SS de *wVulC* et qu'il est nécessaire de poursuivre les efforts de recherche de déterminants bactériens. Deux stratégies se profilent alors : utiliser les données de génomes de *Wolbachia* et/ou rechercher chez l'hôte les cibles de la bactérie pour identifier des effecteurs potentiels et tester secondairement leur sécrétion par la bactérie.

A. Perspectives concernant les voies de communication de *Wolbachia*.

Les publications récentes de nouveaux génomes de *Wolbachia* (Klasson *et al.* 2008, Klasson *et al.* 2009, Salzberg *et al.* 2009) ajoutée à la liste des génomes en cours de séquençage (Figure 9) vont permettre de rechercher des effecteurs bactériens par l'utilisation de la bioinformatique. A ce jour, aucune publication de génomique comparative concernant le modèle *Wolbachia* n'a émergée. Or, il semble que les trois génomes séquencés (*wBm*, *wRi*, *wPip-Pel*) faisant suite à la publication de celui de la souche *wMel* n'aient pas été exploités en profondeur. Depuis la publication du génome de *Rickettsia prowazekii* (Andersson *et al.* 1998), l'intérêt des chercheurs pour la compréhension de la communication des bactéries Rickettsiales s'est porté principalement vers les gènes codant un T4aSS. Devant le caractère non cultivable de *Wolbachia*, nous sommes aujourd'hui dans une impasse technique pour déterminer la fonctionnalité du T4SS ainsi que son rôle précis dans la survie, le maintien et la transmission de la bactérie. La perspective à court terme est donc d'effectuer des expérimentations d'immunodétection comme cela a pu être fait par nos collègues de Lyon sur la souche de *Wolbachia wAtab3* qui infecte *A. tabida*.

Systems		Composants type	w Bm	w Mel	w Ri	w Pip-Pel	
Export à travers l'IM							
Lol	prolipoprotein diacylglycerol transférase		Lgt	Wbm0079	WD0768	WRi_006110	WP0508
	ABC transporter, ATP-binding protein		LolD	Wbm0023	WD0707	WRi_006470	WP0824
	lipoprotein releasing system transmembrane		LolC/E	Wbm0483	WD1229	WRi_012040	WP1072
	Signal peptidase		SPaseII LspA	Wbm0417	WD0760	WRi_006030	WP0500
TAT	Twin-Arginine Translocation		TatA/E	Wbm0678	WD0152	WRi_001010	WP0074
			Putative TatB	Wbm0183	WD0222	WRi_002110	WP0357
			TatC	Wbm0297	WD0452	WRi_002700	WP0155
			TatD	Wbm0245	WD1120	WRi_012930	WP0856
Sec	Transmembrane component	SecY	Wbm0322	WD0662	WRi_005000	WP1185	
		SecE	?	WD0018	WRi_000170	WP0560	
		SecG	Wbm0170	WD0467	WRi_002840	WP1236	
		SecA	Wbm0266	WD0549	WRi_003630	WP0368	
	ATPase	SecF	Wbm0186	WD0235	WRi_002230	WP0002	
		SecD	Wbm0788	WD0086	WRi_001660	WP0476	
		SecB	Wbm0541	WD0106	WRi_001460	WP1088	
		YajC	Wbm0222	WD0534	WRi_003480	WP1242	
	Signal recognition particle protein	YidC	Wbm0188	WD0237	WRi_002260	WP0004	
		FtsY	Wbm0755	WD0888	WRi_008550	WP1267	
		Ffh	Wbm0049	WD1080	WRi_011140	WP1032	
		SPaseI LepB	Wbm0760	WD0893	WRi_008500	WP1272	
	Signal peptidase	DnaJ	Wbm0785	WD0040	WRi_000340	WP0222	
		HscB	Wbm0758	WD0891	WRi_008520	WP1270	
		DnaK	Wbm0495	WD0928	WRi_008800	WP0140	
		GroES	Wbm0349	WD0308	WRi_004600	WP0026	
		GroEL	Wbm0350	WD0307	WRi_004590	WP0027	
		HtpG	Wbm0138	WD1277	WRi_013040	WP0772	
		Sécrétion à travers l'OM					
		T1SS	ABC transporter, ATP binding protein	HlyB	Wbm0203	WD0400	WRi_003890
HlyD	Wbm0309			WD0649	WRi_004860	WP0084	
T2SS	T2aSS T2bSS	classical secretion-dependent T2SS	ToIC	Wbm0575	WD0068	WRi_000590	WP0609
		outer-membrane protein (OMP)	DsbA	Wbm0382	WD1312	WRi_013390	WP0513
T3SS	T2cSS T3aSS T3bSS	Secretin	PilD,C,B,A,T				
		tight-adherence piliation system (Tad)	PilQ	?	WD0500	WRi_002460	WP0491
		injectisome (non flagellar system)					
		flagellar protein export system					
T4SS	T4aSS	Rickettsiales VirB homologues	VirB3	Wbm0798	WD0859	WRi_008240	WP0604
		VirB4a	Wbm0797	WD0858	WRi_008230	WP0603	
		VirB6a	Wbm0796	WD0857	WRi_008220	WP0602	
		VirB6b	Wbm0795	WD0856	WRi_008210	WP0601	
		VirB6c	Wbm0794	WD0855	WRi_008200	WP0600	
		VirB6d	Wbm0793	WD0854	WRi_008190	WP0599	
		VirB8a	Wbm0279	WD0004	WRi_000050	WP1255	
		VirB9a	Wbm0280	WD0005	WRi_000060	WP1256	
		VirB10	Wbm0281	WD0006	WRi_000070	WP1257	
		VirB11	Wbm0282	WD0007	WRi_000080	WP1258	
		VirD4	Wbm0283	WD0008	WRi_000090	WP1259	
		VirB4b	Wbm0750	WD1291	WRi_011480	WP0636	
		VirB8b	Wbm0641	WD0817	WRi_007830	WP0130	
		VirB9b	Wbm0591	WD1104	WRi_012790	WP0871	
		T5SS	T4bSS T5aSS T5bSS T5cSS	Dot/Icm homologues			
autotransporters of type 1 (AT-1)	IgA1						
two-partner secretion system (TpsA/TpsB)	FhaC						
trimeric autotransporters (AT-2)	YadA-like						
T6SS	T6aSS T6bSS T6cSS T6dSS T6eSS T6fSS T6gSS T6hSS T6iSS T6jSS	Chaperonin	Omp85/YaeT	Wbm0054	WD1085	WRi_011190	WP1037
		ATP-dependent Clp serine protease	ClpB	Wbm0553	WD0319	WRi_004690	WP0355
		chaperone-usher pathway (CU)					
		extracellular nucleation-precipitation pathway (ENP)					

Figure 24: Liste des systèmes de sécrétion de *Wolbachia*.

Tableau bilan issu des données des quatre génomes publiés de *Wolbachia* concernant les systèmes de sécrétion potentiels de *Wolbachia*. IM, membrane interne; OM, membrane externe.

L'anticorps anti-VirB6 produit peut être utilisé dans le modèle *wVulC/A. vulgare* pour déterminer si au moins un des composants du T4SS s'exprime et est dirigé vers l'enveloppe bactérienne.

Parallèlement nous pouvons espérer, par l'exploitation du génome de *wVulC* et par comparaison avec les génomes d'autres modèles bactériens, identifier les différentes voies d'export et de sécrétion de protéines bactériennes. La figure 24 synthétise les résultats préliminaires (bibliographiques et bioinformatiques) concernant la communication de *Wolbachia* que j'ai acquis durant cette thèse. La présence d'homologues de gènes codant des composants de système de sécrétion dans les génomes de *Wolbachia* montre que *Wolbachia* est peut-être capable d'utiliser différentes voies de sécrétion. D'après ces premières observations, *Wolbachia* ne possède pas de systèmes de sécrétion de type II, III, V, VII ni VIII (Figure 24). *Wolbachia* pourrait néanmoins utiliser un système Sec ou Tat pour l'export à travers sa membrane cytoplasmique de protéines dont l'un des devenir est le recrutement par le T1SS ou le T4SS (Desvaux *et al.* 2009). Chez *B. pertussis*, il semble que la toxine pertussique est notamment sécrétée en deux étapes et utilise, une fois dans l'espace périplasmique, le système Ptl pour son transport vers l'extérieur de la bactérie (Shrivastava et Miller 2009). Chacune des cinq sous-unités protéiques de la toxine possède une séquence signal N-terminale, suggérant qu'elles traversent dans un premier temps la membrane interne par un mécanisme général de sécrétion (Sec-dépendant), pour se retrouver dans l'espace périplasmique. À ce niveau, la toxine est assemblée sous sa forme définitive, puis elle est exportée vers l'extérieur de la bactérie à travers la membrane externe par le système Ptl sous la forme d'un complexe macromoléculaire puis est libérée dans le milieu extracellulaire (Shrivastava et Miller 2009). Le système Sec a été récemment caractérisé chez un proche parent de *Wolbachia*, *Rickettsia typhi* (Ammerman *et al.* 2008). Après avoir démontré la fonctionnalité de trois composants de cette voie d'export, les auteurs ont ensuite identifié par prédiction *in silico* les protéines possédant un peptide signal dans leur région N-terminale pouvant les diriger vers cette machinerie d'export (Ammerman *et al.* 2008). Ils ont, dans un second temps, entrepris une analyse transcriptionnelle des candidats dont l'activité du peptide signal avait été positive chez *E. coli* (Ammerman *et al.* 2008). Ils ont ainsi mis en évidence 54 protéines extracytoplasmiques potentiellement impliquées dans la croissance et la virulence de *R. typhi* (Ammerman *et al.* 2008). Par ailleurs, *Wolbachia* semble posséder certains constituants de T1SS et T6SS. Ces observations devront être confirmées s'il l'on veut montrer que *Wolbachia* peut également utiliser ces systèmes de sécrétion d'effecteurs.

B. Recherche *in silico* d'effecteurs.

Comme le nombre et la nature des effecteurs sécrétés par des T4SS ont considérablement été enrichis, nous pouvons également envisager d'utiliser les caractéristiques de ces déterminants bactériens pour identifier *in silico* des effecteurs de *wVulC*. Burstein et ses collègues (2009) ont d'ailleurs mis au point une machine de calcul permettant de filtrer systématiquement tous les candidats à la sécrétion par le T4SS *dot/icm* de *L. pneumophila* (Burstein *et al.* 2009). Chacune des 3005 ORF était retenue comme effecteur potentiel, si elle était située à proximité du système de sécrétion, co-régulée avec celui-ci, régulée à la manière d'autres effecteurs déjà connus, si elle avait une faible composition en GC, si elle codait une protéine à motifs eucaryote ou possédait un signal de sécrétion propre au système *dot/icm* (Burstein *et al.* 2009). Après avoir vérifié leur sécrétion par le T4SS *via* un système analogue au CRAfT (système CyaA), les auteurs ont validé 40 nouveaux effecteurs qui se sont ajoutés au 102 déjà connus (Burstein *et al.* 2009). Nous pouvons donc nous aussi nous baser sur de telles caractéristiques pour identifier des effecteurs potentiels chez *wVulC*.

En ce qui concerne les protéines possédant un signal de translocation par un T4SS dans leur région C-terminale, les recherches effectuées sur les ankyrines de *wVulC* se sont révélées aussi infructueuses que celles menées par L. Klasson sur les protéines ANK de *wPip-Pel* (Klasson *et al.* 2008). J'ai également mené la recherche de signature de type ([RK]-x(7)-[RK]-x-[RK]-x[RK]-x(2)) ou ([RK]-x(2)-[RK]-x(2)-[RK]-x-[RK]-x(2)) sur les 1380 protéines du génome en l'état de *wVulC*. Cette analyse préliminaire n'a révélé aucun motif exactement similaire à ceux d'*A. tumefaciens* ou d'*A. phagocytophilum*. Néanmoins, nous avons pu détecter des motifs s'en rapprochant et dont les résidus arginine (R) variaient en position ou étaient remplacés par des lysines (K) pour quinze protéines de *wVulC*. Parmi celles-ci, toutes étaient des protéines possédant des homologues clairs chez les autres souches de *Wolbachia* sauf une méthyltransférase impliquée dans la régulation de la transcription chez d'autres bactéries. Deux autres protéines supposées s'ancrer dans la membrane externe des bactéries, WspB et OmpA ont aussi été identifiées. Le gène *wspB* localisé en aval de l'opéron *virB8-D4* et cotranscrit avec ces gènes chez *wRi*, est d'ailleurs considéré comme un candidat à la sécrétion dans d'autres modèles comme *wAtab3* (Rancès *et al.* 2008). Une analyse génomique a par ailleurs permis de mettre en évidence que le gène codant l'antigène OmpA était disrupté chez *Rickettsia rickettsii* Iowa (avirulente) en comparaison avec *R. rickettsii* Sheila Smith (virulente) (Ellison *et al.* 2008). Par ailleurs, il est annoncé la découverte d'un nouvel effecteur nommé "Ats" chez *A. phagocytophilum* (Rikihisa *et al.* 2009).

Dès la publication des séquences, nous pourrions vérifier s'il n'existe pas d'homologues de ces gènes chez *Wolbachia* et si ces effecteurs possèdent aussi un signal de translocation identique à AnkA. Ce résultat nous permettra d'affiner notre recherche *in silico* d'effecteurs de *wVulC*. L'achèvement de l'assemblage du génome de *wVulC* est également très attendu puisqu'à court terme il s'agira d'un outil efficace dans une stratégie d'identification des facteurs bactériens susceptibles d'être impliqués dans la féminisation des hôtes. Nous pourrions à ce propos étudier plus spécifiquement les gènes propres à *wVulC* en comparaison avec les séquences de souches induisant une IC. Une approche similaire a permis de mettre en évidence, depuis la publication d'une seconde souche de *Wolbachia* infectant *C. pipiens* *wPip-JHB* (Salzberg *et al.* 2009), des gènes spécifiques de chacune des deux souches. Nous pourrions par ailleurs caractériser les autres protéines bactériennes possédant des domaines impliqués dans les interactions protéine-protéine chez les eucaryotes. Parmi celles-ci, on retrouve les protéines à domaines tétratricopeptides (TPR), F-box, SNAREs, les coiled-coil et les régions riches en leucine (LRR).

C. Recherche des cibles hôtes.

Afin de compléter notre compréhension des processus moléculaires sous-jacents aux différents phénotypes de *Wolbachia*, il nous faut non seulement comprendre comment *Wolbachia* interagit avec la reproduction de l'hôte mais aussi comprendre comment l'hôte peut modifier la transmission et l'expression de *Wolbachia*. Ainsi des différences d'expression de gènes candidats de *Wolbachia* permettraient de déterminer leur implication et les cibles de la bactérie. Dans les cas de féminisation, il a été montré que deux souches féminisantes (*wVulC* et *wVulM*) peuvent cohabiter au sein d'une même population mais que leurs modalités de transmission et de maintien sont différentes, aboutissant à des intensités différentes du phénotype (Cordaux *et al.* 2004). Il ne faut pas oublier que les effecteurs sécrétés par *Wolbachia* peuvent également perturber les mécanismes de défense de l'hôte, facilitant la croissance intracellulaire ou induisant la synthèse de nutriments indispensable à la colonisation. Analyser de manière plus qualitative les mécanismes de l'hôte impliqués dans la régulation des populations bactériennes revient à mieux comprendre les mécanismes de résistance des hôtes (à la transmission ou au maintien de *Wolbachia*).

Au niveau moléculaire, les systèmes génétiques hôtes qui régulent la population symbiotique relèvent entre autres de la réponse immunitaire de l'hôte envers *Wolbachia*. *Wolbachia* est présente dans tous les tissus de l'hôte y compris dans les hémocytes, cellules spécifiques de l'immunité (Herbinière *et al.* 2005, Braquart-Varnier *et al.* 2008, Herbinière *et al.* 2008).

D'autre part des transferts horizontaux entre individus sont possibles par contact hémolympatique (Rigaud et Juchault 1995). Ces observations soulèvent plusieurs questions. Comment *Wolbachia* pénètre-t-elle à l'intérieur des hémocytes ? Les hémocytes servent-ils de vecteurs pour la dissémination de la bactérie dans les tissus hôtes ? De quelle manière *Wolbachia* échappe-t-elle au système immunitaire de l'hôte ? Certains hémocytes peuvent phagocyter la bactérie mais celle-ci semble échapper à la dégradation. De plus, une analyse cytologique menée sur *A. vulgare* montre que dans les hémocytes et les organes hématopoïétiques, *Wolbachia* apparaît comme enfermée dans un endosome qui n'aurait pas fusionné avec les lysosomes (Herbinière *et al.* 2005). Dans les cellules granulaires, la présence de *Wolbachia* ne serait donc pas la conséquence d'une phagocytose mais plutôt d'une endocytose, probablement induite par la bactérie. Ce type d'endocytose provoquée a déjà été montré chez de proches bactéries intracellulaires comme *Neorickettsia risticii* (Messick et Rikihisa 1994). On peut donc supposer que la bactérie induit son internalisation dans les hémocytes *via* l'interaction d'effecteurs bactériens avec l'actine et le cytosquelette des cellules comme chez *Salmonella* spp. (Galan et Zhou 2000). Par ailleurs, les séquences des génomes connus indiquent que les gènes codant les enzymes impliquées dans la synthèse des lipopolysaccharides (LPS) ont été perdus chez *Wolbachia* (Wu *et al.* 2004). L'absence de ces molécules, connues pour être des déterminants antigéniques, pourrait permettre à *Wolbachia* d'être furtive vis-à-vis de la réponse immunitaire. Chez *A. vulgare*, un peptide antimicrobien a été isolé (Herbinière *et al.* 2005) et les protéines majeures impliquées dans le système immunitaire et au niveau du cytosquelette ont été récemment caractérisées (Herbinière *et al.* 2008). L'identification d'autres protéines hôtes impliquées dans la réponse aux invasions bactériennes est en cours. Le degré de l'expression ainsi que de l'activité des protéines hôtes en fonction de l'infection à *Wolbachia* est en effet suivi dans le cadre du projet ANR mené conjointement avec des collègues de Lyon.

Les résultats présentés dans ce manuscrit de thèse constituent en définitive les éléments de base d'une meilleure appréciation des mécanismes des interactions *Wolbachia*-arthropode et en particulier de la souche féminisante *wVulC/A. vulgare*. Les projets à venir permettront de comprendre l'évolution de ces interactions et de proposer de nouvelles stratégies de contrôle basées sur la manipulation des populations symbiotiques.

REFERENCES

REFERENCES

- Al-Khodor S., C.T. Price, F. Habyarimana, A. Kalia et Y. Abu Kwaik. 2008. A Dot/Icm-translocated ankyrin protein of *Legionella pneumophila* is required for intracellular proliferation within human macrophages and protozoa. *Molecular Microbiology*, 70: 908-923.
- Altman E. et G. Segal. 2008. The response regulator CpxR directly regulates expression of several *Legionella pneumophila* icm/dot components as well as new translocated substrates. *Journal of Bacteriology*, 190: 1985-1996.
- Altschul S.F., W. Gish, W. Miller, E.W. Myers et D.J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215: 403-410.
- Aly K.A., L. Krall, F. Lottspeich et C. Baron. 2008. The type IV secretion system component VirB5 binds to the trans-zeatin biosynthetic enzyme Tzs and enables its translocation to the cell surface of *Agrobacterium tumefaciens*. *Journal of Bacteriology*, 190: 1595-1604.
- Ammerman N.C., M.S. Rahman et A.F. Azad. 2008. Characterization of Sec-translocon-dependent extracytoplasmic proteins of *Rickettsia typhi*. *Journal of Bacteriology*, 190: 6234-6242.
- Andersson S.G., A. Zomorodipour, J.O. Andersson, T. Sicheritz-Ponten, U.C. Alsmark, R.M. Podowski, A.K. Naslund, A.S. Eriksson, H.H. Winkler et C.G. Kurland. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature*, 396: 133-140.
- Atmakuri K., Z. Ding et P.J. Christie. 2003. VirE2, a type IV secretion substrate, interacts with the VirD4 transfer protein at cell poles of *Agrobacterium tumefaciens*. *Molecular Microbiology*, 49: 1699-1713.
- Atmakuri K., E. Cascales et P.J. Christie. 2004. Energetic components VirD4, VirB11 and VirB4 mediate early DNA transfer reactions required for bacterial type IV secretion. *Molecular Microbiology*, 54: 1199-1211.
- Audia J.P., M.C. Patton et H.H. Winkler. 2008. DNA microarray analysis of the heat shock transcriptome of the obligate intracytoplasmic pathogen *Rickettsia prowazekii*. *Applied and Environmental Microbiology*, 74: 7809-7812.
- Backert S. et T.F. Meyer. 2006. Type IV secretion systems and their effectors in bacterial pathogenesis. *Current Opinion in Microbiology*, 9: 207-217.
- Baldo L., L. Prendini, A. Corthals et J.H. Werren. 2007. *Wolbachia* are present in southern african scorpions and cluster with supergroup F. *Current Microbiology*, 55: 367-373.
- Bandi C., T.J. Anderson, C. Genchi et M.L. Blaxter. 1998. Phylogeny of *Wolbachia* in filarial nematodes. *Proceedings of the Royal Society B: Biological Sciences*, 265: 2407-2413.

- Bandi C., A.M. Dunn, G.D. Hurst et T. Rigaud. 2001. Inherited microorganisms, sex-specific virulence and reproductive parasitism. *Trends in Parasitology*, 17: 88-94.
- Bao W., Y. Kumagai, H. Niu, M. Yamaguchi, K. Miura et Y. Rikihisa. 2009. Four VirB6 paralogs and VirB9 are expressed and interact in *Ehrlichia chaffeensis*-containing vacuoles. *Journal of Bacteriology*, 191: 278-286.
- Baron C. 2005. From bioremediation to biowarfare: on the impact and mechanism of type IV secretion systems. *FEMS Microbiology Letters*, 253: 163-170.
- Baron C. et B. Coombes. 2007. Targeting bacterial secretion systems: benefits of disarmament in the microcosm. *Infectious Disorders - Drug Targets*, 7: 19-27.
- Beijersbergen A., S.J. Smith et P.J. Hooykaas. 1994. Localization and topology of VirB proteins of *Agrobacterium tumefaciens*. *Plasmid*, 32: 212-218.
- Binet R., S. Letoffe, J.M. Ghigo, P. Delepelaire et C. Wandersman. 1997. Protein secretion by Gram-negative bacterial ABC exporters - a review. *Gene*, 192: 7-11.
- Blöcker A.J., J.E. Deane, A.K. Veenendaal, P. Roversi, J.L. Hodgkinson, S. Johnson et S.M. Lea. 2008. What's the point of the type III secretion system needle? *Proceedings of the National Academy of Sciences U S A*, 105: 6507-6513.
- Bordenstein S.R. et J.J. Wernegreen. 2004. Bacteriophage flux in endosymbionts (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. *Molecular Biology and Evolution*, 21: 1981-1991.
- Bork P. 1993. Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins*, 17: 363-374.
- Bouchon D., T. Rigaud et P. Juchault. 1998. Evidence for widespread *Wolbachia* infection in isopod crustaceans: molecular identification and host feminization. *Proceedings of the Royal Society B: Biological Sciences*, 265: 1081-1090.
- Bouchon D., R. Cordaux et P. Grève. 2008. Feminizing *Wolbachia* and the evolution of sex determination in isopods. In *Insect Symbiosis*. Bourtzis K. et T.A. Miller, Boca Raton, FL. p273-294.
- Bourtzis K. et T.A. Miller. 2003. *Insect Symbiosis*. Bourtzis K. et T.A. Miller, Boca Raton, FL.
- Bourtzis K. 2008. *Wolbachia*-based technologies for insect pest population control. *Advances in Experimental Medicine and Biology*, 627: 104-113.
- Braquart-Varnier C., M. Lachat, J. Herbinière, M. Johnson, Y. Caubet, D. Bouchon et M. Sicard. 2008. *Wolbachia* mediate variation of host immunocompetence. *PLoS One*, 3: e3286.

- Brayton K.A., L.S. Kappmeyer, D.R. Herndon, M.J. Dark, D.L. Tibbals, G.H. Palmer, T.C. McGuire et D.P. Knowles, Jr. 2005. Complete genome sequencing of *Anaplasma marginale* reveals that the surface is skewed to two superfamilies of outer membrane proteins. *Proceedings of the National Academy of Sciences U S A*, 102: 844-849.
- Breeuwer J.A., R. Stouthamer, S.M. Barns, D.A. Pelletier, W.G. Weisburg et J.H. Werren. 1992. Phylogeny of cytoplasmic incompatibility micro-organisms in the parasitoid wasp genus *Nasonia* (Hymenoptera: Pteromalidae) based on 16S ribosomal DNA sequences. *Insect Molecular Biology*, 1: 25-36.
- Breeuwer J.A. et G. Jacobs. 1996. *Wolbachia*: intracellular manipulators of mite reproduction. *Experimental and Applied Acarology*, 20: 421-434.
- Bressac C. et F. Rousset. 1993. The reproductive incompatibility system in *Drosophila simulans*: DAPI-staining analysis of the *Wolbachia* symbionts in sperm cysts. *Journal of Invertebrate Pathology*, 61: 226-230.
- Brownlie J.C., B.N. Cass, M. Riegler, J.J. Witsenburg, I. Iturbe-Ormaetxe, E.A. McGraw et S.L. O'Neill. 2009. Evidence for metabolic provisioning by a common invertebrate endosymbiont, *Wolbachia pipientis*, during periods of nutritional stress. *PLoS Pathogens*, 5: e1000368.
- Brownstein J.S., E. Hett et S.L. O'Neill. 2003. The potential of virulent *Wolbachia* to modulate disease transmission by insects. *Journal of Invertebrate Pathology*, 84: 24-29.
- Bull J.J. et W.R. Rice. 1991. Distinguishing mechanisms for the evolution of co-operation. *Journal of Theoretical Biology*, 149: 63-74.
- Burstein D., T. Zusman, E. Degtyar, R. Viner, G. Segal et T. Pupko. 2009. Genome-scale identification of *Legionella pneumophila* effectors using a machine learning approach. *PLoS Pathogens*, 5: e1000508.
- Callaini G., R. Dallai et M.G. Riparbelli. 1997. *Wolbachia*-induced delay of paternal chromatin condensation does not prevent maternal chromosomes from entering anaphase in incompatible crosses of *Drosophila simulans*. *Journal of Cell Science*, 110: 271-280.
- Cascales E. et P.J. Christie. 2003. The versatile bacterial type IV secretion systems. *Nature Reviews Microbiology*, 1: 137-149.
- Cascales E., K. Atmakuri, Z. Liu, A.N. Binns et P.J. Christie. 2005. *Agrobacterium tumefaciens* oncogenic suppressors inhibit T-DNA and VirE2 protein substrate binding to the VirD4 coupling protein. *Molecular Microbiology*, 58: 565-579.

- Casiraghi M., S.R. Bordenstein, L. Baldo, N. Lo, T. Beninati, J.J. Wernegreen, J.H. Werren et C. Bandi. 2005. Phylogeny of *Wolbachia pipientis* based on *gltA*, *groEL* and *ftsZ* gene sequences: clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the *Wolbachia* tree. *Microbiology*, 151: 4015-4022.
- Caturegli P., K.M. Asanovich, J.J. Walls, J.S. Bakken, J.E. Madigan, V.L. Popov et J.S. Dumler. 2000. *ankA*: an *Ehrlichia phagocytophila* group gene encoding a cytoplasmic protein antigen with ankyrin repeats. *Infection and Immunity*, 68: 5277-5283.
- Caubet Y., M.J. Hatcher, J.P. Mocquard et T. Rigaud. 2000. Genetic conflict and changes in heterogametic mechanisms of sex determination. *Journal of Evolutionary Biology*, 13: 766-777.
- Cazalet C., S. Jarraud, Y. Ghavi-Helm, F. Kunst, P. Glaser, J. Etienne et C. Buchrieser. 2008. Multigenome analysis identifies a worldwide distributed epidemic *Legionella pneumophila* clone that emerged within a highly diverse species. *Genome Research*, 18: 431-441.
- Censini S., C. Lange, Z. Xiang, J.E. Crabtree, P. Ghiara, M. Borodovsky, R. Rappuoli et A. Covacci. 1996. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proceedings of the National Academy of Sciences U S A*, 93: 14648-14653.
- Charlat S., C. Calmet et H. Merçot. 2001. On the mod resc model and the evolution of *Wolbachia* compatibility types. *Genetics*, 159: 1415-1422.
- Cheng Z., Y. Kumagai, M. Lin, C. Zhang et Y. Rikihisa. 2006. Intra-leukocyte expression of two-component systems in *Ehrlichia chaffeensis* and *Anaplasma phagocytophilum* and effects of the histidine kinase inhibitor closantel. *Cellular Microbiology*, 8: 1241-1252.
- Chien C.T., P.L. Bartel, R. Sternglanz et S. Fields. 1991. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proceedings of the National Academy of Sciences U S A*, 88: 9578-9582.
- Christie P.J. 1997. *Agrobacterium tumefaciens* T-complex transport apparatus: a paradigm for a new family of multifunctional transporters in eubacteria. *Journal of Bacteriology*, 179: 3085-3094.
- Christie P.J. et J.P. Vogel. 2000. Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends in Microbiology*, 8: 354-360.
- Christie P.J. 2001. Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. *Molecular Microbiology*, 40: 294-305.
- Coburn B., I. Sekirov et B.B. Finlay. 2007. Type III secretion systems and disease. *Clinical Microbiology Reviews*, 20: 535-549.
- Combes C. 1995. Interactions durables: écologie et évolution du parasitisme. Dunod, Paris. 544 pages.

- Cordaux R., A. Michel-Salzat, M. Frelon-Raimond, T. Rigaud et D. Bouchon. 2004. Evidence for a new feminizing *Wolbachia* strain in the isopod *Armadillidium vulgare*: evolutionary implications. *Heredity*, 93: 78-84.
- Cordaux R., S. Pichon, A. Ling, P. Pérez, C. Delaunay, F. Vavre, D. Bouchon et P. Grève. 2008. Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. *Molecular Biology and Evolution*, 25: 1889-1896.
- Cornelis G.R. 2006. The type III secretion injectisome. *Nature Reviews Microbiology*, 4: 811-825.
- Cover T.L., C.P. Dooley et M.J. Blaser. 1990. Characterization of and human serologic response to proteins in *Helicobacter pylori* broth culture supernatants with vacuolizing cytotoxin activity. *Infection and Immunity*, 58: 603-610.
- Cyr D.M., T. Langer et M.G. Douglas. 1994. DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. *Trends in Biochemical Sciences*, 19: 176-181.
- Dawkins R. 1982. The expended phenotype. Freeman W.H., Oxford.
- de Bary H.A. 1879. De la symbiose. *Revue Internationale des Sciences*, 3: 301-309.
- de Felipe K.S., R.T. Glover, X. Charpentier, O.R. Anderson, M. Reyes, C.D. Pericone et H.A. Shuman. 2008. *Legionella* eukaryotic-like type IV substrates interfere with organelle trafficking. *PLoS Pathogens*, 4: e1000117.
- de Jong M.F., Y.H. Sun, A.B. den Hartigh, J.M. van Dijk et R.M. Tsois. 2008. Identification of VceA and VceC, two members of the VjbR regulon that are translocated into macrophages by the *Brucella* type IV secretion system. *Molecular Microbiology*, 70: 1378-1396.
- de Paz H.D., F.J. Sangari, S. Bolland, J.M. Garcia-Lobo, C. Dehio, F. de la Cruz et M. Llosa. 2005. Functional interactions between type IV secretion systems involved in DNA transfer and virulence. *Microbiology*, 151: 3505-3516.
- Dedene F., F. Vavre, F. Fleury, B. Loppin, M.E. Hochberg et M. Boulétreau. 2001. Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. *Proceedings of the National Academy of Sciences U S A*, 98: 6247-6252.
- Dedene F., M. Boulétreau et F. Vavre. 2005. *Wolbachia* requirement for oogenesis: occurrence within the genus *Asobara* (Hymenoptera, Braconidae) and evidence for intraspecific variation in *A. tabida*. *Heredity*, 95: 394-400.
- Delrue R.M., C. Deschamps, S. Leonard, C. Nijskens, I. Danese, J.M. Schaus, S. Bonnot, J. Ferrooz, A. Tibor, X. De Bolle et J.J. Letesson. 2005. A quorum-sensing regulator controls expression of both the type IV secretion system and the flagellar apparatus of *Brucella melitensis*. *Cellular Microbiology*, 7: 1151-1161.

- den Hartigh A.B., H.G. Rolan, M.F. de Jong et R.M. Tsolis. 2008. VirB3 to VirB6 and VirB8 to VirB11, but not VirB7, are essential for mediating persistence of *Brucella* in the reticuloendothelial system. *Journal of Bacteriology*, 190: 4427-4436.
- Desvaux M., M. Hébraud, R. Talon et I.R. Henderson. 2009. Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. *Trends in Microbiology*, 17: 139-145.
- Di Cello F., Y. Xie, M. Paul-Satyaseela et K.S. Kim. 2005. Approaches to bacterial RNA isolation and purification for microarray analysis of *Escherichia coli* K1 interaction with human brain microvascular endothelial cells. *Journal of Clinical Microbiology*, 43: 4197-4199.
- Ding Z., Z. Zhao, S.J. Jakubowski, A. Krishnamohan, W. Margolin et P.J. Christie. 2002. A novel cytology-based, two-hybrid screen for bacteria applied to protein-protein interaction studies of a type IV secretion system. *Journal of Bacteriology*, 184: 5572-5582.
- Dobson S.L. 2003. Reversing *Wolbachia*-based population replacement. *Trends in Parasitology*, 19: 128-133.
- Dobson S.L., W. Rattanadechakul et E.J. Marsland. 2004. Fitness advantage and cytoplasmic incompatibility in *Wolbachia* single- and superinfected *Aedes albopictus*. *Heredity*, 93: 135-142.
- Duron O., D. Bouchon, S. Boutin, L. Bellamy, L. Zhou, J. Engelstadter et G.D. Hurst. 2008. The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC Biology*, 6: 27.
- Eckler S.A., R. Kuehn et M. Gautam. 2005. Deletion of N-terminal rapsyn domains disrupts clustering and has dominant negative effects on clustering of full-length rapsyn. *Neuroscience*, 131: 661-670.
- Economou A. 1999. Following the leader: bacterial protein export through the Sec pathway. *Trends in Microbiology*, 7: 315-320.
- Economou A., P.J. Christie, R.C. Fernandez, T. Palmer, G.V. Plano et A.P. Pugsley. 2006. Secretion by numbers: protein traffic in prokaryotes. *Molecular Microbiology*, 62: 308-319.
- Ellison D.W., T.R. Clark, D.E. Sturdevant, K. Virtaneva, S.F. Porcella et T. Hackstadt. 2008. Genomic comparison of virulent *Rickettsia rickettsii* Sheila Smith and avirulent *Rickettsia rickettsii* Iowa. *Infection and Immunity*, 76: 542-550.
- Ewald P.W. 1987. Transmission modes and evolution of the parasitism-mutualism continuum. *Annals of the New York Academy of Sciences*, 503: 295-306.
- Félix C. 2004. Etude moléculaire de la bactérie intracellulaire féminisante *Wolbachia* chez *Armadillidium vulgare* (crustacé isopode terrestre). Thèse de doctorat d'Etat, Université de Poitiers, 203 pages.

- Félix C., S. Pichon, C. Braquart-Varnier, H. Braig, L. Chen, R.A. Garrett, G. Martin et P. Grève. 2008. Characterization and transcriptional analysis of two gene clusters for type IV secretion machinery in *Wolbachia* of *Armadillidium vulgare*. *Research in Microbiology*, 159: 481-485.
- Fenn K. et M. Blaxter. 2006. *Wolbachia* genomes: revealing the biology of parasitism and mutualism. *Trends in Parasitology*, 22: 60-65.
- Fields S. et O. Song. 1989. A novel genetic system to detect protein-protein interactions. *Nature*, 340: 245-246.
- Fields S. et R. Sternglanz. 1994. The two-hybrid system: an assay for protein-protein interactions. *Trends in Genetics*, 10: 286-292.
- Filloux A., A. Hachani et S. Bleves. 2008. The bacterial type VI secretion machine: yet another player for protein transport across membranes. *Microbiology*, 154: 1570-1583.
- Foster J., M. Ganatra, I. Kamal, J. Ware, K. Makarova, N. Ivanova, A. Bhattacharyya, V. Kapatral, S. Kumar, J. Posfai, T. Vincze, J. Ingram, L. Moran, A. Lapidus, M. Omelchenko, N. Kyrpides, E. Ghedin, S. Wang, E. Goltsman, V. Joukov, O. Ostrovskaya, K. Tsukerman, M. Mazur, D. Comb, E. Koonin et B. Slatko. 2005. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biology*, 3: e121.
- Fronzes R., P.J. Christie et G. Waksman. 2009. The structural biology of type IV secretion systems. *Nature Reviews Microbiology*, 7: 703-714.
- Fry A.J., M.R. Palmer et D.M. Rand. 2004. Variable fitness effects of *Wolbachia* infection in *Drosophila melanogaster*. *Heredity*, 93: 379-389.
- Fullner K.J., K.M. Stephens et E.W. Nester. 1994. An essential virulence protein of *Agrobacterium tumefaciens*, VirB4, requires an intact mononucleotide binding domain to function in transfer of T-DNA. *Molecular and General Genetics*, 245: 704-715.
- Galan J.E. et D. Zhou. 2000. Striking a balance: modulation of the actin cytoskeleton by *Salmonella*. *Proceedings of the National Academy of Sciences U S A*, 97: 8754-8761.
- Gentschev I., G. Dietrich et W. Goebel. 2002. The *E. coli* alpha-hemolysin secretion system and its use in vaccine development. *Trends in Microbiology*, 10: 39-45.
- Gillespie J.J., N.C. Ammerman, S.M. Dreher-Lesnick, M.S. Rahman, M.J. Worley, J.C. Setubal, B.S. Sobral et A.F. Azad. 2009. An anomalous type IV secretion system in *Rickettsia* is evolutionarily conserved. *PLoS One*, 4: e4833.
- Goodacre S.L., O.Y. Martin, C.F. Thomas et G.M. Hewitt. 2006. *Wolbachia* and other endosymbiont infections in spiders. *Molecular Ecology*, 15: 517-527.

- Habyarimana F., S. Al-Khodori, A. Kalia, J.E. Graham, C.T. Price, M.T. Garcia et Y.A. Kwaik. 2008. Role for the Ankyrin eukaryotic-like genes of *Legionella pneumophila* in parasitism of protozoan hosts and human macrophages. *Environmental Microbiology*, 10: 1460-1474.
- Hamilton C.M., H. Lee, P.L. Li, D.M. Cook, K.R. Piper, S.B. von Bodman, E. Lanka, W. Ream et S.K. Farrand. 2000. TraG from RP4 and TraG and VirD4 from Ti plasmids confer relaxosome specificity to the conjugal transfer system of pTiC58. *Journal of Bacteriology*, 182: 1541-1548.
- Hedges L.M., J.C. Brownlie, S.L. O'Neill et K.N. Johnson. 2008. *Wolbachia* and virus protection in insects. *Science*, 322: 702.
- Henderson I.R., J.P. Nataro, J.B. Kaper, T.F. Meyer, S.K. Farrand, D.L. Burns, B.B. Finlay et J.W. St Geme, 3rd. 2000. Renaming protein secretion in the Gram-negative bacteria. *Trends in Microbiology*, 8: 352.
- Herbinière J., C. Braquart-Varnier, P. Grève, J.M. Strub, J. Frère, A. Van Dorsselaer et G. Martin. 2005. Armadillidin: a novel glycine-rich antibacterial peptide directed against Gram-positive bacteria in the woodlouse *Armadillidium vulgare* (Terrestrial Isopod, Crustacean). *Developmental and Comparative Immunology*, 29: 489-499.
- Herbinière J., P. Grève, J.M. Strub, D. Thierse, M. Raimond, A. van Dorsselaer, G. Martin et C. Braquart-Varnier. 2008. Protein profiling of hemocytes from the terrestrial crustacean *Armadillidium vulgare*. *Developmental and Comparative Immunology*, 32: 875-882.
- Hilgenboecker K., P. Hammerstein, P. Schlattmann, A. Telschow et J.H. Werren. 2008. How many species are infected with *Wolbachia*? A statistical analysis of current data. *FEMS Microbiology Letters*, 281: 215-220.
- Hiroki M., Y. Kato, T. Kamito et K. Miura. 2002. Feminization of genetic males by a symbiotic bacterium in a butterfly, *Eurema hecabe* (Lepidoptera: Pieridae). *Naturwissenschaften*, 89: 167-170.
- Hiroki M., Y. Tagami, K. Miura et Y. Kato. 2004. Multiple infection with *Wolbachia* inducing different reproductive manipulations in the butterfly *Eurema hecabe*. *Proceedings of the Royal Society B: Biological Sciences*, 271: 1751-1755.
- Hoerauf A., K. Nissen-Pahle, C. Schmetz, K. Henkle-Duhrsen, M.L. Blaxter, D.W. Buttner, M.Y. Gallin, K.M. Al-Qaoud, R. Lucius et B. Fleischer. 1999. Tetracycline therapy targets intracellular bacteria in the filarial nematode *Litomosoides sigmodontis* and results in filarial infertility. *The Journal of Clinical Investigation*, 103: 11-18.

- Hotopp J.C., M. Lin, R. Madupu, J. Crabtree, S.V. Angiuoli, J.A. Eisen, R. Seshadri, Q. Ren, M. Wu, T.R. Utterback, S. Smith, M. Lewis, H. Khouri, C. Zhang, H. Niu, Q. Lin, N. Ohashi, N. Zhi, W. Nelson, L.M. Brinkac, R.J. Dodson, M.J. Rosovitz, J. Sundaram, S.C. Daugherty, T. Davidsen, A.S. Durkin, M. Gwinn, D.H. Haft, J.D. Selengut, S.A. Sullivan, N. Zafar, L. Zhou, F. Benahmed, H. Forberger, R. Halpin, S. Mulligan, J. Robinson, O. White, Y. Rikihisa et H. Tettelin. 2006. Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genetics*, 2: e21.
- Huang Y., K. Magori, A.L. Lloyd et F. Gould. 2007. Introducing transgenes into insect populations using combined gene-drive strategies: modeling and analysis. *Insect Biochemistry and Molecular Biology*, 37: 1054-1063.
- Hubber A., A.C. Vergunst, J.T. Sullivan, P.J. Hooykaas et C.W. Ronson. 2004. Symbiotic phenotypes and translocated effector proteins of the *Mesorhizobium loti* strain R7A VirB/D4 type IV secretion system. *Molecular Microbiology*, 54: 561-574.
- Hueck C.J. 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiology and Molecular Biology Reviews*, 62: 379-433.
- Hurst G.D. et F.M. Jiggins. 2000. Male-killing bacteria in insects: mechanisms, incidence, and implications. *Emerging Infectious Diseases*, 6: 329-336.
- Hurst L.D. et J.P. Randerson. 2002. Parasitic sex puppeteers. *Scientific American*, 286: 56-61.
- Ishikawa H. 2003. Insect symbiosis: an introduction. In *Insect Symbiosis*. Bourtzis K. et T.A. Miller, Boca Raton, FL. p1-21.
- Iturbe-Ormaetxe I., G.R. Burke, M. Riegler et S.L. O'Neill. 2005. Distribution, expression, and motif variability of ankyrin domain genes in *Wolbachia pipientis*. *Journal of Bacteriology*, 187: 5136-5145.
- Jiggins F.M., G.D. Hurst et M.E. Majerus. 2000. Sex-ratio-distorting *Wolbachia* causes sex-role reversal in its butterfly host. *Proceedings of the Royal Society B: Biological Sciences*, 267: 69-73.
- Juchault P. et J.J. Legrand. 1972. Croisements de néo-mâles expérimentaux chez *Armadillidium vulgare* Latr. (Crustacé, Isopode, Oniscoïde). Mise en évidence d'une hétérogamétie femelle. *Comptes Rendus de l'Académie des Sciences de Paris*, 274: 1387-1389.
- Juchault P. et J.J. Legrand. 1985. Mechanism of the refractory state of androgen hormone in *Armadillidium vulgare* Latr. (crustacean, isopod, oniscoid) harboring a feminizing bacteria. *General and Comparative Endocrinology*, 60: 463-467.
- Judd P.K., R.B. Kumar et A. Das. 2005. The type IV secretion apparatus protein VirB6 of *Agrobacterium tumefaciens* localizes to a cell pole. *Molecular Microbiology*, 55: 115-124.

- Karimova G., J. Pidoux, A. Ullmann et D. Ladant. 1998. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proceedings of the National Academy of Sciences U S A*, 95: 5752-5756.
- Klasson L., T. Walker, M. Sebahia, M.J. Sanders, M.A. Quail, A. Lord, S. Sanders, J. Earl, S.L. O'Neill, N. Thomson, S.P. Sinkins et J. Parkhill. 2008. Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group. *Molecular Biology and Evolution*, 25: 1877-1887.
- Klasson L., J. Westberg, P. Sapountzis, K. Naslund, Y. Lutnaes, A.C. Darby, Z. Veneti, L. Chen, H.R. Braig, R. Garrett, K. Bourtzis et S.G. Andersson. 2009. The mosaic genome structure of the *Wolbachia* wRi strain infecting *Drosophila simulans*. *Proceedings of the National Academy of Sciences U S A*, 106: 5725-5730.
- Kostakioti M., C.L. Newman, D.G. Thanassi et C. Stathopoulos. 2005. Mechanisms of protein export across the bacterial outer membrane. *Journal of Bacteriology*, 187: 4306-4314.
- Kremer N., D. Charif, H. Henri, M. Bataille, G. Prevost, K. Kraaijeveld et F. Vavre. 2009a. A new case of *Wolbachia* dependence in the genus *Asobara*: evidence for parthenogenesis induction in *Asobara japonica*. *Heredity*, 103: 248-256.
- Kremer N., D. Voronin, D. Charif, P. Mavingui, B. Mollereau et F. Vavre. 2009b. *Wolbachia* interferes with ferritin expression and iron metabolism in insects. *PLoS Pathogens*, 5: e1000630.
- La M.V., P. Francois, C. Rovey, S. Robineau, P. Barbry, J. Schrenzel, D. Raoult et P. Renesto. 2007. Development of a method for recovering Rickettsial RNA from infected cells to analyze gene expression profiling of obligate intracellular bacteria. *Journal of Microbiological Methods*, 71: 292-297.
- Lachat M. 2009. Impact de deux souches de *Wolbachia* sur les traits d'histoire de vie de leurs hôtes *Armadillidium vulgare*. Thèse de doctorat d'Etat, Université de Poitiers, 167 pages.
- Landmann F., G.A. Orsi, B. Loppin et W. Sullivan. 2009. *Wolbachia*-mediated cytoplasmic incompatibility is associated with impaired histone deposition in the male pronucleus. *PLoS Pathogens*, 5: e1000343.
- Legrand J.J., G. Martin et J.C. Artault. 1978. Correlation between the presence of a bacterial symbiont in oocytes of *Porcellio dilatatus petiti*, and the sterility of the cross: P.d. petiti male x P.d. dilatatus female. *Archives de l'Institut Pasteur de Tunis*, 55: 507-514.
- Legrand J.J., E. Legrand-Hamelin et P. Juchault. 1987. Sex determination in crustacea. *Biological Reviews*, 62: 439-470.
- Letunic I., T. Doerks et P. Bork. 2009. SMART 6: recent updates and new developments. *Nucleic Acids Research*, 37: D229-232.

- Lin M., A. den Dulk-Ras, P.J. Hooykaas et Y. Rikihisa. 2007. *Anaplasma phagocytophilum* Anka secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection. *Cellular Microbiology*, 9: 2644-2657.
- Lin T.S. et C.I. Kado. 1993. The virD4 gene is required for virulence while virD3 and orf5 are not required for virulence of *Agrobacterium tumefaciens*. *Molecular Microbiology*, 9: 803-812.
- Lipsitch M., M.A. Nowak, D. Ebert et R.M. May. 1995. The population dynamics of vertically and horizontally transmitted parasites. *Proceedings of the Royal Society B: Biological Sciences*, 260: 321-327.
- Llosa M. et D. O'Callaghan. 2004. Euroconference on the Biology of Type IV Secretion Processes: bacterial gates into the outer world. *Molecular Microbiology*, 53: 1-8.
- Lo N., M. Casiraghi, E. Salati, C. Bazzocchi et C. Bandi. 2002. How many *Wolbachia* supergroups exist? *Molecular Biology and Evolution*, 19: 341-346.
- Lo N., C. Paraskevopoulos, K. Bourtzis, S.L. O'Neill, J.H. Werren, S.R. Bordenstein et C. Bandi. 2007. Taxonomic status of the intracellular bacterium *Wolbachia pipientis*. *International Journal of Systematic and Evolutionary Microbiology*, 57: 654-657.
- Lopez J.E., G.H. Palmer, K.A. Brayton, M.J. Dark, S.E. Leach et W.C. Brown. 2007. Immunogenicity of *Anaplasma marginale* type IV secretion system proteins in a protective outer membrane vaccine. *Infection and Immunity*, 75: 2333-2342.
- Luo Z.Q. et R.R. Isberg. 2004. Multiple substrates of the *Legionella pneumophila* Dot/Icm system identified by interbacterial protein transfer. *Proceedings of the National Academy of Sciences U S A*, 101: 841-846.
- Malek J.A., J.M. Wierzbowski, W. Tao, S.A. Bosak, D.J. Saranga, L. Doucette-Stamm, D.R. Smith, P.J. McEwan et K.J. McKernan. 2004. Protein interaction mapping on a functional shotgun sequence of *Rickettsia sibirica*. *Nucleic Acids Research*, 32: 1059-1064.
- Marcadé I., C. Souty-Grosset, D. Bouchon, T. Rigaud et R. Raimond. 1999. Mitochondrial DNA variability and *Wolbachia* infection in two sibling woodlice species. *Heredity*, 83: 71-78.
- Martin D.P., C. Williamson et D. Posada. 2005. RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics*, 21: 260-262.
- Martin G., P. Juchault et J.J. Legrand. 1973. Mise en évidence d'un micro-organisme intracytoplasmique de l'oniscoïde *Armadillidium vulgare* Latr. dont la présence accompagne l'intersexualité ou la féminisation totale des mâles génétiques de la lignée thélygène. *Comptes Rendus de l'Académie des Sciences de Paris*, 276: 2313-2316.

- Martin G., P. Juchault, O. Sorokine et A. van Dorsselaer. 1990. Purification and characterization of androgenic hormone from the terrestrial isopod *Armadillidium vulgare* Latr. (Crustacea, Oniscidea). *General and Comparative Endocrinology*, 80: 349-354.
- Martin G., O. Sorokine, M. Moniatte, P. Bulet, C. Hetru et A. Van Dorsselaer. 1999. The structure of a glycosylated protein hormone responsible for sex determination in the isopod, *Armadillidium vulgare*. *European Journal of Biochemistry*, 262: 727-736.
- Masui S., T. Sasaki et H. Ishikawa. 2000. Genes for the type IV secretion system in an intracellular symbiont, *Wolbachia*, a causative agent of various sexual alterations in arthropods. *Journal of Bacteriology*, 182: 6529-6531.
- Matsushita O., K. Yoshihara, S. Katayama, J. Minami et A. Okabe. 1994. Purification and characterization of *Clostridium perfringens* 120-kilodalton collagenase and nucleotide sequence of the corresponding gene. *Journal of Bacteriology*, 176: 149-156.
- McGraw E.A., D.J. Merritt, J.N. Droller et S.L. O'Neill. 2002. *Wolbachia* density and virulence attenuation after transfer into a novel host. *Proceedings of the National Academy of Sciences U S A*, 99: 2918-2923.
- McGraw E.A. et S.L. O'Neill. 2004. *Wolbachia pipientis*: intracellular infection and pathogenesis in *Drosophila*. *Current Opinion in Microbiology*, 7: 67-70.
- Medini D., A. Covacci et C. Donati. 2006. Protein homology network families reveal step-wise diversification of Type III and Type IV secretion systems. *PLoS Computational Biology*, 2: e173.
- Méresse S., O. Steele-Mortimer, E. Moreno, M. Desjardins, B. Finlay et J.P. Gorvel. 1999. Controlling the maturation of pathogen-containing vacuoles: a matter of life and death. *Nature Cell Biology*, 1: e183-188.
- Messick J.B. et Y. Rikihisa. 1994. Inhibition of binding, entry, or intracellular proliferation of *Ehrlichia risticii* in P388D1 cells by anti-E. risticii serum, immunoglobulin G, or Fab fragment. *Infection and Immunity*, 62: 3156-3161.
- Michel-Salzat A., R. Cordaux et D. Bouchon. 2001. *Wolbachia* diversity in the *Porcellionides pruinosus* complex of species (Crustacea: Oniscidea): evidence for host-dependent patterns of infection. *Heredity*, 87: 428-434.
- Middleton R., K. Sjolander, N. Krishnamurthy, J. Foley et P. Zambryski. 2005. Predicted hexameric structure of the *Agrobacterium* VirB4 C terminus suggests VirB4 acts as a docking site during type IV secretion. *Proceedings of the National Academy of Science U S A*, 102: 1685-1690.
- Min K.T. et S. Benzer. 1997. *Wolbachia*, normally a symbiont of *Drosophila*, can be virulent, causing degeneration and early death. *Proceedings of the National Academy of Sciences U S A*, 94: 10792-10796.

- Moran N.A. et G.R. Plague. 2004. Genomic changes following host restriction in bacteria. *Current Opinion in Genetics & Development*, 14: 627-633.
- Moreau J., A. Bertin, Y. Caubet et T. Rigaud. 2001. Sexual selection in an isopod with *Wolbachia*-induced sex reversal: males prefer real females. *Journal of Evolutionary Biology*, 14: 388-394.
- Moret Y., P. Juchault et T. Rigaud. 2001. *Wolbachia* endosymbiont responsible for cytoplasmic incompatibility in a terrestrial crustacean: effects in natural and foreign hosts. *Heredity*, 86: 325-332.
- Mosavi L.K., T.J. Cammett, D.C. Desrosiers et Z.Y. Peng. 2004. The ankyrin repeat as molecular architecture for protein recognition. *Protein Science*, 13: 1435-1448.
- Nagai H., E.D. Cambronne, J.C. Kagan, J.C. Amor, R.A. Kahn et C.R. Roy. 2005. A C-terminal translocation signal required for Dot/Icm-dependent delivery of the *Legionella* RalF protein to host cells. *Proceedings of the National Academy of Sciences U S A*, 102: 826-831.
- Narita S., D. Kageyama, M. Nomura et T. Fukatsu. 2007. Unexpected mechanism of symbiont-induced reversal of insect sex: feminizing *Wolbachia* continuously acts on the butterfly *Eurema hecabe* during larval development. *Applied and Environmental Microbiology*, 73: 4332-4341.
- Negri I., M. Pellecchia, P.J. Mazzoglio, A. Patetta et A. Alma. 2006. Feminizing *Wolbachia* in *Zyginidia pullula* (Insecta, Hemiptera), a leafhopper with an XX/X0 sex-determination system. *Proceedings of the Royal Society B: Biological Sciences*, 273: 2409-2416.
- Nicosia A., M. Perugini, C. Franzini, M.C. Casagli, M.G. Borri, G. Antoni, M. Almoni, P. Neri, G. Ratti et R. Rappuoli. 1986. Cloning and sequencing of the pertussis toxin genes: operon structure and gene duplication. *Proceedings of the National Academy of Sciences U S A*, 83: 4631-4635.
- Niu H., Y. Rikihisa, M. Yamaguchi et N. Ohashi. 2006. Differential expression of VirB9 and VirB6 during the life cycle of *Anaplasma phagocytophilum* in human leucocytes is associated with differential binding and avoidance of lysosome pathway. *Cellular Microbiology*, 8: 523-534.
- Noh S.M., K.A. Brayton, W.C. Brown, J. Norimine, G.R. Munske, C.M. Davitt et G.H. Palmer. 2008. Composition of the surface proteome of *Anaplasma marginale* and its role in protective immunity induced by outer membrane immunization. *Infection and Immunity*, 76: 2219-2226.
- O'Neill S.L. et T.L. Karr. 1990. Bidirectional incompatibility between conspecific populations of *Drosophila simulans*. *Nature*, 348: 178-180.
- O'Neill S.L., R. Giordano, A.M. Colbert, T.L. Karr et H.M. Robertson. 1992. 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. *Proceedings of the National Academy of Sciences U S A*, 89: 2699-2702.
- O'Neill S.L., A.A. Hoffmann et J.H. Werren. 1997. Influential passengers: inherited microorganisms and arthropod reproduction. Oxford University Press, New York. 232 pages.

- Oehler S. et K. Bourtzis. 2000. First International *Wolbachia* Conference: *Wolbachia* 2000. *Symbiosis*, 29: 151-161.
- Ohashi N., N. Zhi, Q. Lin et Y. Rikihisa. 2002. Characterization and transcriptional analysis of gene clusters for a type IV secretion machinery in human granulocytic and monocytic ehrlichiosis agents. *Infection and Immunity*, 70: 2128-2138.
- Okuno A., Y. Hasegawa, T. Ohira, Y. Katakura et H. Nagasawa. 1999. Characterization and cDNA cloning of androgenic gland hormone of the terrestrial isopod *Armadillidium vulgare*. *Biochemical and Biophysical Research Communications*, 264: 419-423.
- Orihuela C.J., J.N. Radin, J.E. Sublett, G. Gao, D. Kaushal et E.I. Tuomanen. 2004. Microarray analysis of pneumococcal gene expression during invasive disease. *Infection and Immunity*, 72: 5582-5596.
- Palmer G.H. et K.A. Brayton. 2007. Gene conversion is a convergent strategy for pathogen antigenic variation. *Trends in Parasitology*, 23: 408-413.
- Pan X., A. Luhrmann, A. Satoh, M.A. Laskowski-Arce et C.R. Roy. 2008. Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science*, 320: 1651-1654.
- Panteleev D., I.I. Goriacheva, B.V. Andrianov, N.L. Reznik, O.E. Lazebnyi et A.M. Kulikov. 2007. The endosymbiotic bacterium *Wolbachia* enhances the nonspecific resistance to insect pathogens and alters behavior of *Drosophila melanogaster*. *Genetika*, 43: 1277-1280.
- Park J., K.J. Kim, K.S. Choi, D.J. Grab et J.S. Dumler. 2004. *Anaplasma phagocytophilum* AnkA binds to granulocyte DNA and nuclear proteins. *Cellular Microbiology*, 6: 743-751.
- Paschos A., G. Patey, D. Sivanesan, C. Gao, R. Bayliss, G. Waksman, D. O'Callaghan et C. Baron. 2006. Dimerization and interactions of *Brucella suis* VirB8 with VirB4 and VirB10 are required for its biological activity. *Proceedings of the National Academy of Sciences U S A*, 103: 7252-7257.
- Pichon S., D. Bouchon, R. Cordaux, L. Chen, R.A. Garrett et P. Grève. 2009. Conservation of the Type IV Secretion System throughout *Wolbachia* evolution. *Biochemical and Biophysical Research Communications*, 385: 557-562.
- Poinsot D., K. Bourtzis, G. Markakis, C. Savakis et H. Merçot. 1998. *Wolbachia* transfer from *Drosophila melanogaster* into *D. simulans*: Host effect and cytoplasmic incompatibility relationships. *Genetics*, 150: 227-237.
- Rancès E., D. Voronin, V. Tran-Van et P. Mavingui. 2008. Genetic and functional characterization of the type IV secretion system in *Wolbachia*. *Journal of Bacteriology*, 190: 5020-5030.
- Rasgon J.L. 2008. Using predictive models to optimize *Wolbachia*-based strategies for vector-borne disease control. *Advances in Experimental Medicine and Biology*, 627: 114-125.

- Rashkova S., G.M. Spudich et P.J. Christie. 1997. Characterization of membrane and protein interaction determinants of the *Agrobacterium tumefaciens* VirB11 ATPase. *Journal of Bacteriology*, 179: 583-591.
- Remaut H., C. Tang, N.S. Henderson, J.S. Pinkner, T. Wang, S.J. Hultgren, D.G. Thanassi, G. Waksman et H. Li. 2008. Fiber formation across the bacterial outer membrane by the chaperone/usher pathway. *Cell*, 133: 640-652.
- Renesto P., C. Rovey, J. Schrenzel, Q. Leroy, A. Huyghe, W. Li, H. Lepidi, P. Francois et D. Raoult. 2008. *Rickettsia conorii* transcriptional response within inoculation eschar. *PLoS One*, 3: e3681.
- Rigaud T., C. Souty-Grosset, R. Raimond, J.P. Mocquard et P. Juchault. 1991. Feminizing endocytobiosis in the terrestrial crustacean *Armadillidium vulgare* Latr. (Isopoda): recent acquisitions. *Endocytobiology and Cell Research*, 7: 259-273.
- Rigaud T. et P. Juchault. 1995. Success and failure of horizontal transfers of feminizing *Wolbachia* endosymbionts in woodlice. *Journal of Evolutionary Biology*, 8: 249-255.
- Rigaud T. et F. Rousset. 1996. What explains the diversity of *Wolbachia*-arthropod interactions? *Biodiversity and Conservation*, 5: 999-1013.
- Rigaud T., P. Juchault et J.P. Mocquard. 1997. The evolution of sex determination in isopod crustaceans. *BioEssays*, 19: 409-416.
- Rigaud T. 1999. Further endosymbiont diversity: a tree hiding in the forest? *Trends in Ecology and Evolution*, 14: 212-213.
- Rigaud T., J. Moreau et P. Juchault. 1999. *Wolbachia* infection in the terrestrial isopod *Oniscus asellus*: sex ratio distortion and effect on fecundity. *Heredity*, 83: 469-475.
- Rigaud T., P.S. Pennings et P. Juchault. 2001. *Wolbachia* bacteria effects after experimental interspecific transfers in terrestrial isopods. *Journal of Invertebrate Pathology*, 77: 251-257.
- Rikihisa Y., M. Lin, H. Niu et Z. Cheng. 2009. Type IV secretion system of *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*. *Annals of the New York Academy of Sciences*, 1166: 106-111.
- Rousset F., D. Bouchon, B. Pintureau, P. Juchault et M. Solignac. 1992. *Wolbachia* endosymbionts responsible for various alterations of sexuality in arthropods. *Proceedings of the Royal Society B: Biological Sciences*, 250: 91-98.
- Rowley S.M., R.J. Raven et E.A. McGraw. 2004. *Wolbachia pipientis* in Australian spiders. *Current Microbiology*, 49: 208-214.
- Ruang-Areerate T. et P. Kittayapong. 2006. *Wolbachia* transinfection in *Aedes aegypti*: a potential gene driver of dengue vectors. *Proceedings of the National Academy of Sciences U S A*, 103: 12534-12539.

- Salzberg S.L., J.C. Hotopp, A.L. Delcher, M. Pop, D.R. Smith, M.B. Eisen et W.C. Nelson. 2005. Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. *Genome Biology*, 6: R23.
- Salzberg S.L., D. Puiu, D.D. Sommer, V. Nene et N.H. Lee. 2009. Genome sequence of the *Wolbachia* endosymbiont of *Culex quinquefasciatus* JHB. *Journal of Bacteriology*, 191: 1725.
- Sandkvist M. 2001a. Type II secretion and pathogenesis. *Infection and Immunity*, 69: 3523-3535.
- Sandkvist M. 2001b. Biology of type II secretion. *Molecular Microbiology*, 40: 271-283.
- Savvides S.N., H.J. Yeo, M.R. Beck, F. Blaesing, R. Lurz, E. Lanka, R. Buhrdorf, W. Fischer, R. Haas et G. Waksman. 2003. VirB11 ATPases are dynamic hexameric assemblies: new insights into bacterial type IV secretion. *The EMBO Journal*, 22: 1969-1980.
- Schilthuisen M. et R. Stouthamer. 1997. Horizontal transmission of parthenogenesis-inducing microbes in *Trichogramma* wasps. *Proceedings of the Royal Society B: Biological Sciences*, 264: 361-366.
- Schmitt W. et R. Haas. 1994. Genetic analysis of the *Helicobacter pylori* vacuolating cytotoxin: structural similarities with the IgA protease type of exported protein. *Molecular Microbiology*, 12: 307-319.
- Schulein R., P. Guye, T.A. Rhomberg, M.C. Schmid, G. Schroder, A.C. Vergunst, I. Carena et C. Dehio. 2005. A bipartite signal mediates the transfer of type IV secretion substrates of *Bartonella henselae* into human cells. *Proceedings of the National Academy of Sciences U S A*, 102: 856-861.
- Sedgwick S.G. et S.J. Smerdon. 1999. The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends in Biochemical Sciences*, 24: 311-316.
- Segal G., J.J. Russo et H.A. Shuman. 1999. Relationships between a new type IV secretion system and the Icm/Dot virulence system of *Legionella pneumophila*. *Molecular Microbiology*, 34: 799-809.
- Seshadri R., I.T. Paulsen, J.A. Eisen, T.D. Read, K.E. Nelson, W.C. Nelson, N.L. Ward, H. Tettelin, T.M. Davidsen, M.J. Beanan, R.T. Deboy, S.C. Daugherty, L.M. Brinkac, R. Madupu, R.J. Dodson, H.M. Khouri, K.H. Lee, H.A. Carty, D. Scanlan, R.A. Heinzen, H.A. Thompson, J.E. Samuel, C.M. Fraser et J.F. Heidelberg. 2003. Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proceedings of the National Academy of Sciences U S A*, 100: 5455-5460.
- Seubert A., R. Hiestand, F. de la Cruz et C. Dehio. 2003. A bacterial conjugation machinery recruited for pathogenesis. *Molecular Microbiology*, 49: 1253-1266.
- Sexton J.A. et J.P. Vogel. 2002. Type IVB secretion by intracellular pathogens. *Traffic*, 3: 178-185.

- Shrivastava R. et J.F. Miller. 2009. Virulence factor secretion and translocation by *Bordetella* species. *Current Opinion in Microbiology*, 12: 88-93.
- Sinkins S.P., T. Walker, A.R. Lynd, A.R. Steven, B.L. Makepeace, H.C. Godfray et J. Parkhill. 2005. *Wolbachia* variability and host effects on crossing type in *Culex* mosquitoes. *Nature*, 436: 257-260.
- Stathopoulos C., D.R. Hendrixson, D.G. Thanassi, S.J. Hultgren, J.W. St Geme, 3rd et R. Curtiss, 3rd. 2000. Secretion of virulence determinants by the general secretory pathway in Gram-negative pathogens: an evolving story. *Microbes and Infection*, 2: 1061-1072.
- Stephens K.M., C. Roush et E. Nester. 1995. *Agrobacterium tumefaciens* VirB11 protein requires a consensus nucleotide-binding site for function in virulence. *Journal of Bacteriology*, 177: 27-36.
- Stouthamer R., J.A. Breeuwer, R.F. Luck et J.H. Werren. 1993. Molecular identification of microorganisms associated with parthenogenesis. *Nature*, 361: 66-68.
- Taminiau B., M. Daykin, S. Swift, M.L. Boschiroli, A. Tibor, P. Lestrade, X. De Bolle, D. O'Callaghan, P. Williams et J.J. Letesson. 2002. Identification of a quorum-sensing signal molecule in the facultative intracellular pathogen *Brucella melitensis*. *Infection and Immunity*, 70: 3004-3011.
- Tanaka K., S. Furukawa, N. Nikoh, T. Sasaki et T. Fukatsu. 2009. Complete WO phage sequences reveal their dynamic evolutionary trajectories and putative functional elements required for integration into the *Wolbachia* genome. *Applied and Environmental Microbiology*, 75: 5676-5686.
- Taylor M.J. et A. Hoerauf. 1999. *Wolbachia* bacteria of filarial nematodes. *Parasitology Today*, 15: 437-442.
- Taylor M.J., C. Bandi et A. Hoerauf. 2005. *Wolbachia* bacterial endosymbionts of filarial nematodes. *Advances in Parasitology*, 60: 245-284.
- Teixeira L., A. Ferreira et M. Ashburner. 2008. The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLoS Biology*, 6: e2.
- Thanassi D.G., C. Stathopoulos, A. Karkal et H. Li. 2005. Protein secretion in the absence of ATP: the autotransporter, two-partner secretion and chaperone/usher pathways of Gram-negative bacteria. *Molecular Membrane Biology*, 22: 63-72.
- Tram U. et W. Sullivan. 2002. Role of delayed nuclear envelope breakdown and mitosis in *Wolbachia*-induced cytoplasmic incompatibility. *Science*, 296: 1124-1126.
- Tram U., P.M. Ferree et W. Sullivan. 2003. Identification of *Wolbachia*-host interacting factors through cytological analysis. *Microbes and Infection*, 5: 999-1011.
- Tsagkarakou A., T. Guillemaud, F. Rousset et M. Navajas. 1996. Molecular identification of a *Wolbachia* endosymbiont in a *Tetranychus urticae* strain (Acari: Tetranychidae). *Insect Molecular Biology*, 5: 217-221.

- Tsai K.H., C.G. Huang, L.C. Wang, Y.W. Yu, W.J. Wu et W.J. Chen. 2007. Molecular evidence for the endosymbiont *Wolbachia* in a non-filaroid nematode, *Angiostrongylus cantonensis*. *Journal of Biomedical Science*, 14: 607-615.
- Tseng T.T., B.M. Tyler et J.C. Setubal. 2009. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiology*, 9: S2.
- Tummuru M.K., S.A. Sharma et M.J. Blaser. 1995. *Helicobacter pylori* picB, a homologue of the *Bordetella pertussis* toxin secretion protein, is required for induction of IL-8 in gastric epithelial cells. *Molecular Microbiology*, 18: 867-876.
- Turley A.P., L.A. Moreira, S.L. O'Neill et E.A. McGraw. 2009. *Wolbachia* infection reduces blood-feeding success in the Dengue Fever mosquito, *Aedes aegypti*. *PLoS Neglected Tropical Diseases*, 3: e516.
- Van Duyne G.D. 2001. A structural view of cre-loxp site-specific recombination. *Annual Review of Biophysics and Biomolecular Structure*, 30: 87-104.
- Van Valen L.M. 1973. A new evolutionary law. *Evolution Theory*, 1: 1-30.
- Vandel A. 1941. Recherches sur la génétique et la sexualité des isopodes terrestres. VI: Les phénomènes de monogénie chez les Oniscoïdes. *Bulletin biologique de la France et de la Belgique*, 75: 316-363.
- Vavre F., R. Allemand, F. Fleury, P. Fouillet et M. Boulétreau. 1999. A new cytoplasmic incompatibility type due to *Wolbachia* in haplodiploid insects. *Annales De La Societe Entomologique De France*, 35: 133-135.
- Veneti Z., M.E. Clark, S. Zabalou, T.L. Karr, C. Savakis et K. Bourtzis. 2003. Cytoplasmic incompatibility and sperm cyst infection in different *Drosophila-Wolbachia* associations. *Genetics*, 164: 545-552.
- Vergunst A.C., B. Schrammeijer, A. den Dulk-Ras, C.M. de Vlaam, T.J. Regensburg-Tuink et P.J. Hooykaas. 2000. VirB/D4-dependent protein translocation from *Agrobacterium* into plant cells. *Science*, 290: 979-982.
- Vergunst A.C., M.C. van Lier, A. den Dulk-Ras, T.A. Stuve, A. Ouwehand et P.J. Hooykaas. 2005. Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of *Agrobacterium*. *Proceedings of the National Academy of Sciences U S A*, 102: 832-837.
- Verne S. 2007. Diversité et impact des *Wolbachia* féminisantes sur la structuration génétique d'*Armadillidium vulgare*. Thèse de doctorat d'Etat, Université de Poitiers, 199 pages.
- Verne S., M. Johnson, D. Bouchon et F. Grandjean. 2007. Evidence for recombination between feminizing *Wolbachia* in the isopod genus *Armadillidium*. *Gene*, 397: 58-66.

- Vidal M. et P. Legrain. 1999. Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Research*, 27: 919-929.
- Voth D.E., D. Howe, P.A. Beare, J.P. Vogel, N. Unsworth, J.E. Samuel et R.A. Heinzen. 2009. The *Coxiella burnetii* ankyrin repeat domain-containing protein family is heterogeneous, with C-terminal truncations that influence Dot/Icm-mediated secretion. *Journal of Bacteriology*, 191: 4232-4242.
- Wade M.J. et N.W. Chang. 1995. Increased male fertility in *Tribolium confusum* beetles after infection with the intracellular parasite *Wolbachia*. *Nature*, 373: 72-74.
- Walker T., L. Klasson, M. Sebahia, M.J. Sanders, N.R. Thomson, J. Parkhill et S.P. Sinkins. 2007. Ankyrin repeat domain-encoding genes in the wPip strain of *Wolbachia* from the *Culex pipiens* group. *BMC Biology*, 5: 39.
- Wang X., Z. Cheng, C. Zhang, T. Kikuchi et Y. Rikihisa. 2007. *Anaplasma phagocytophilum* p44 mRNA expression is differentially regulated in mammalian and tick host cells: involvement of the DNA binding protein ApxR. *Journal of Bacteriology*, 189: 8651-8659.
- Ward D.V., O. Draper, J.R. Zupan et P.C. Zambryski. 2002. Peptide linkage mapping of the *Agrobacterium tumefaciens* vir-encoded type IV secretion system reveals protein subassemblies. *Proceedings of the National Academy of Sciences U S A*, 99: 11493-11500.
- Watarai M., S. Makino et T. Shirahata. 2002. An essential virulence protein of *Brucella abortus*, VirB4, requires an intact nucleoside-triphosphate-binding domain. *Microbiology*, 148: 1439-1446.
- Werren J.H., W. Zhang et L.R. Guo. 1995. Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. *Proceedings of the Royal Society B: Biological Sciences*, 261: 55-63.
- Werren J.H. 1997. Biology of *Wolbachia*. *Annual Review of Entomology*, 42: 587-609.
- Werren J.H., L. Baldo et M.E. Clark. 2008. *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology*, 6: 741-751.
- Williams K.P., B.W. Sobral et A.W. Dickerman. 2007. A robust species tree for the alphaproteobacteria. *Journal of Bacteriology*, 189: 4578-4586.
- Winans S.C., N.J. Mantis, C.Y. Chen, C.H. Chang et D.C. Han. 1994. Host recognition by the VirA-VirG two-component regulatory proteins of *Agrobacterium tumefaciens*. *Research in Microbiology*, 145: 461-473.
- Wu H.Y., P.C. Chung, H.W. Shih, S.R. Wen et E.M. Lai. 2008. Secretome analysis uncovers an Hcp-family protein secreted *via* a type VI secretion system in *Agrobacterium tumefaciens*. *Journal of Bacteriology*, 190: 2841-2850.

Wu M., L.V. Sun, J. Vamathevan, M. Riegler, R. Deboy, J.C. Brownlie, E.A. McGraw, W. Martin, C. Esser, N. Ahmadinejad, C. Wiegand, R. Madupu, M.J. Beanan, L.M. Brinkac, S.C. Daugherty, A.S. Durkin, J.F. Kolonay, W.C. Nelson, Y. Mohamoud, P. Lee, K. Berry, M.B. Young, T. Utterback, J. Weidman, W.C. Nierman, I.T. Paulsen, K.E. Nelson, H. Tettelin, S.L. O'Neill et J.A. Eisen. 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biology, 2: e69.

Yen J.H. et A.R. Barr. 1971. New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens*. Nature, 232: 657-658.

ANNEXES

ANNEXE 1 : LE DOUBLE-HYBRIDE.

Dissection et extraction d'ADN.

Les ovaires prélevés à partir de femelles dont le statut d'infection est inconnu ont été broyés à l'aide d'un piston en polypropylène (Merck Eurolab) dans 400 µl de tampon d'extraction (Tris 100 mM ; EDTA 10 mM ; NaCl 100 mM ; SDS 0,1% ; DTT 50 mM, pH 8) contenant 0,1 mg de protéinase K (20 mg/ml) et incubés pendant 4 h à 37°C. Le surnageant contenant l'ADN est extrait à l'aide d'un ½ volume de phénol et un ½ volume de chloroforme isoamyle (1 :24), puis centrifugé à 12000 rpm pendant 8 min à température ambiante. Cette étape est répétée deux fois, la dernière étant effectuée avec un volume de chloroforme uniquement. L'ADN est ensuite précipité par addition d'1/10ème de volume d'acétate de sodium (3 M, pH 5,2) et d'un volume d'isopropanol pendant une nuit à -20°C. L'échantillon est enfin centrifugé à 15000 rpm pendant 30 min à 4°C. Le culot d'ADN est rincé à l'éthanol 70°, séché et repris dans 50 µl d'eau bidistillée, puis conservé à 4°C. Le dosage de l'ADN est réalisé à l'aide d'un spectrophotomètre par mesure de la DO à 260 nm, et la pureté de l'échantillon est estimée grâce au rapport de DO 260/280.

Préparation des inserts et des vecteurs.

Les ADN complets codant VirD4, VirB4 et VirB11 et ceux codant les différents paralogues de *pk1* et *pk2* ont été obtenus par PCR puis purifiés en utilisant des amorces contenant des sites de restriction compatibles avec l'insertion dans les vecteurs pACT2 (aimablement fourni par le Dr E. Huguet) ou pGBKT7 (Clontech). Le plasmide pACT2 a été linéarisé par coupure enzymatique avec *Bam*HI et *Xho*I, et le plasmide pGBKT7 avec *Eco*RI et *Sal*I ou *Eco*RI et *Bam*HI. La réaction est réalisée dans un volume de 20 µL, à 37°C pendant 1 nuit, en présence du plasmide ou de l'insert, du tampon (10X) approprié à l'enzyme utilisée et de 10 U d'enzyme de restriction. La réaction de ligature est réalisée à 16°C pendant 16h, dans un volume final de 10 µL contenant l'insert et le vecteur linéarisé (dans un rapport 3/1), le tampon de ligation 5X et 3 U de ligase T4 (New England Biolabs).

Clonage des inserts et des vecteurs.

La solution de ligation est mélangée avec 100 µL de bactéries compétentes (XL1-Blue MRF' supercompetent cells, Stratagene) et incubée sur de la glace pendant 30 min. Après un choc thermique à 42°C (90 s), les bactéries sont transformées avec le plasmide recombinant et rapidement placées dans 500 µL de milieu LB (Tryptone 10 g/L, Extrait de levure 5 g/L, NaCl 5 g/L) et cultivées à 37°C pendant 1h pour permettre l'expression phénotypique du marqueur de sélection. Les bactéries sont ensuite étalées sur des boîtes de culture (LB agar, 20 g/L) contenant l'antibiotique de sélection (50 µg/ml de Kanamycine ou 50 µg/ml d'Ampicilline). La sélection des clones positifs a été effectuée par PCR sur les colonies obtenues en utilisant des amorces spécifiques du vecteur. Le programme de PCR comportait 35 cycles de 2 min à 94°C, 1 min à 55°C, 4 min à 72°C et une étape finale à 72°C pendant 5 min. ont été séquencés dans les 2 directions par un séquenceur d'ADN automatisé (ABI Prism model 377, Perkin-Elmer).

Minipréparations d'ADN.

L'extraction d'ADN plasmidique a été réalisée avec le kit Promega Wizard® plus, SV minipreps. Une culture de bactéries transformées en milieu LB liquide (2 ml) est centrifugée durant 5 min à 10000 rpm. Le culot obtenu est resuspendu dans 250 µL de solution 1 auxquels sont ajoutés 250 µL de solution de lyse et 10 µL de solution protéase alcaline. Après 5 min d'incubation à température ambiante, 350 µL de la solution de neutralisation sont ajoutées. Chaque tube est ensuite centrifugé durant 10 minutes à 13000 rpm à température ambiante. Le surnageant obtenu est placé sur une colonne d'affinité et centrifugé durant 1 minute à 13000 rpm. La colonne est rincée par centrifugation (13000 rpm, 1 min) successivement avec 750 µL puis 250 µL de la solution de lavage du kit. L'élution de l'ADN fixé à la colonne est réalisée avec 50 µL d'eau ultrapure.

Séquençage.

Une réaction de PCR comme indiquée précédemment a été réalisée avec les plasmides extraits par minipréparation comme matrice. Les fragments de PCR sont séquencés à l'aide du kit « ABI PRISM TM » (Applied Biosystems). L'amorce de séquençage (10 pmoles) et l'ADN purifié (Kit Jet Pure, Genomed) (~ 100 ng) sont ajoutés au milieu réactionnel (1/30 de Big Dye Terminator (BDT) et du tampon 5X). Une réaction de PCR incorporant des didéoxynucléotides fluorescents est réalisée avec 25 cycles de 10 s à 95°C, 5 s à 50°C et 4 min à 60°C après une dénaturation initiale de 2 min à 95°C.

Suit alors une étape de précipitation où sont ajoutés dans chaque tube 50 µl d'éthanol 100% et 2,2 µl d'acétate de sodium 3M (pH 5,2). Après centrifugation (12000 rpm, 20 min), les culots sont rincés à l'éthanol 70%, séchés puis repris dans 15 µl de formamide désionisé, et dénaturés à 95°C pendant 3 min. L'analyse est réalisée sur un séquenceur ABI PRISM 3130 de Perkin Elmer®, au Pôle Biologie Santé de Poitiers. Les séquences obtenues ont été analysées à l'aide de BLASTN et de TBLASTN (Altschul *et al.* 1990) pour vérifier leur caractère spécifique à *wVulC*.

Transformation des levures par les plasmides.

Un autre protocole de minipréparation d'ADN plasmidique a été utilisé pour la transformation de levures, permettant notamment de préserver les ARN utilisés comme carrier lors de la transformation. Cette fois, les bactéries sont resuspendues au vortex dans 100 µL de solution 1 (glucose 0,9% (w/v) ; EDTA 10 mM, pH 8 ; Tris HCL 25 mM, pH8), auxquels sont ajoutés 200 µL de solution 2 (NaOH 0,2N ; SDS 1% (5w/v)). Après 10 min d'incubation à température ambiante, 150 µL de solution 3 (Na acétate 3M, pH 5,2) sont ajoutés et la solution est incubée à nouveau 10 min sur de la glace. L'ensemble est ensuite centrifugé durant 5 min à 14000 rpm à 4°C. Deux volumes d'éthanol 100° sont ajoutés au surnageant et la solution obtenue est centrifugée durant 10 minutes à 14000 rpm à 4°C. Le culot d'ADN est resuspendu dans 100 µL de solution 4 (EDTA 1 mM, pH 8 ; Tris HCL 10 mM, pH 7.5 ; Na acétate 0.15 M, pH 5.2) et 200 µL d'éthanol absolu. Cette solution est centrifugée durant 10 minutes à 14000 rpm à 4°C. Le culot obtenu est séché et resuspendu dans 25 µL d'eau distillée.

Moins de 24h avant la transformation, une colonie de la souche de levure à transformer est utilisée pour ensemencer une préculture de 10 ml de milieu YPD (Extrait de levure 10 g/L ; Peptone 10 g/L ; Glucose 20 g/L ; Agar 20 g/L) et agitée une nuit à 30°C. Cette préculture est ensuite diluée dans le milieu YPD pour obtenir une D0600 de 0,2. Après 4h d'agitation à 30°C, les cellules se trouvent dans une phase de croissance exponentielle, ce qui correspond à une D0600 comprise entre 0,4 et 0,6. Les cellules sont alors centrifugées 5 min à 3000 g. Le culot est suspendu dans 100 µL de solution A (Tris HCL 10 mM, pH 7.5 ; EDTA 1mM, pH 8 ; acétate de lithium 0,1M ; sorbitol 1M) et incubée pendant 45 min à 30°C. Après une centrifugation d'une minute à 9000 X g, le culot de cellules est resuspendu soigneusement avec 5 µL d'une solution contenant l'ADN plasmidique puis incubé pendant 10 min à température ambiante. Cent microlitres de solution B (Tris HCL 10 mM, pH7,5 ; EDTA 1 mM, pH8 ; acétate de lithium 0,1M ; PEG 4000 40% (w/v)) et 1 à 2 µg de plasmide sont ajoutés et la solution est incubée pendant 30 min à 30°C.

Après un choc thermique de 10 min à 42°C, les cellules sont centrifugées et resuspendues dans 100 µL de solution C (Tris HCL 10 mM, pH7.5 ; EDTA 1 mM, pH 8 ; sorbitol 0,6 M). Les levures sont finalement étalées sur le milieu sélectif approprié aux plasmides transformés pendant 3 à 5 jours.

Tests d'interaction.

Tous les tests d'interaction ont été effectués avec un premier vecteur exprimant le domaine d'activation de Gal4 (pACT2) et portant le gène marqueur d'auxotrophie LEU2 qui permet sa sélection sur un milieu dépourvu de leucine. Le second vecteur du système exprime le domaine de fixation Gal4 (pGBKT7) et porte le gène marqueur d'auxotrophie TRP1 qui permet sa sélection sur un milieu dépourvu de tryptophane. Dans tous les cas, les 2 gènes rapporteurs utilisés correspondent au gène HIS3 permettant une croissance sur un milieu dépourvu d'histidine et le gène lacZ permettant une coloration bleue en présence d'X-Gal. Les deux vecteurs du système sont ensuite introduits au sein d'une même levure, soit par co-transformation, soit par croisement sexuel. La première méthode consiste à transformer une souche de levure par les deux plasmides simultanément. La seconde fait appel à deux souches haploïdes de signe sexuel opposé, qui sont transformées par chacun des plasmides. Le croisement sexuel de ces souches permet d'obtenir des cellules diploïdes contenant les deux constructions. La méthode de co-transformation requière autant de co-transformations qu'il y a d'interaction à tester (par exemple 10 partenaires à tester entre-eux représentent 100 co-transformations à réaliser pour tester 100 interactions), alors que la méthode de croisement sexuel permet de réduire le nombre de transformation aux nombres de partenaires à tester (dans ce cas, 10 transformations suffisent à tester 100 interactions).

Différentes souches de levure ont été utilisées. Elles possèdent toutes les gènes rapporteurs HIS3 et lacZ insérés dans leurs chromosomes sous le contrôle des séquences promotrices reconnues spécifiquement par Gal4 (souches CG1945 ou Y187). Ces souches sont toutes déficientes pour les gènes d'auxotrophie TRP1, LEU2 et HIS3 afin de permettre la sélection des plasmides transformés et de pouvoir tester les interactions protéine/protéine. De plus, les souches utilisées dans le système Gal4 sont délétées du gène GAL4 endogène. Après la mise en présence des deux plasmides, les levures sont étalées sur des milieux sélectifs permettant d'une part de sélectionner les cellules possédant les deux plasmides et de vérifier ainsi l'efficacité de la transformation ou du croisement (milieu -LW), et d'autre part de sélectionner les cellules dans lesquelles une interaction protéine/protéine a lieu par l'activation du premier gène rapporteur HIS3 (milieu -LWH). L'activation du deuxième gène rapporteur lacZ est, quant à elle, visualisée par un test X-Gal.

- Milieu sélectif -LW pour un litre : Yeast Nitrogen Base 6,7 g, do-LW 1,54 g, NaOH 5 N 0,8 ml, glucose 2% final (autoclavé séparément). Pour obtenir un milieu sélectif -L, ajouter 76 mg de W. Pour obtenir un milieu sélectif -W, ajouter 380 mg de L. Pour obtenir un milieu solide, ajouter 20 g/L de Bacto-agar.
- Milieu sélectif -LWH pour un litre : Yeast Nitrogen Base 6,7 g, do-LWH 1,46 g, NaOH 5 N 0,8 ml. Le glucose est autoclavé séparément et ajouté à une concentration finale de 2%. Pour obtenir un milieu sélectif -H, ajouter 76 mg de W et 380 mg de L. Pour obtenir un milieu sélectif supplémenté en 3-AT, à partir d'un stock 1 M filtré, ajouter le volume requis après refroidissement du milieu jusqu'à environ 50°C.

Les milieux à tester sont fraîchement préparés avant chaque test de façon à obtenir des concentrations finales de 25 g.L⁻¹ pour L-Alanine (A), L-Histidine (H), L-Uracile (U), L-Tryptophane (W) et 50 g.L⁻¹ pour L-Leucine (L). Au préalable, trois solutions d'acides aminés (H+A+U, H+A+U+L, H+A+U+W) sont préparées à partir de poudre pure de chaque acide aminé (L-amino acids, Sigma-Aldrich) dissoutes dans de l'eau bidistillée et filtrée à 0,45 µm. Toutes les solutions utilisées dans cette étude sont préparées avec de l'eau distillée traitée au DEPC afin de prévenir une contamination par des RNases.

Tests en goutte.

Le croisement sexuel a été plus spécifiquement utilisé pour obtenir les résultats d'interaction présentés dans la partie IV. Le croisement sexuel entre deux souches de signe sexuel opposé (Y187 et CG1945) permet d'obtenir des levures diploïdes. Cette méthode est une alternative à la co-transformation pour mettre en présence dans la même cellule les deux plasmides du système double-hybride et présente l'avantage d'être beaucoup plus efficace. Les deux partenaires du croisement à réaliser sont étalés sur le milieu sélectif approprié aux marqueurs de sélection qu'ils contiennent et incubés de 3 à 5 jours à 30°C. Après croissance, les 2 partenaires sont mélangés sur milieu YPD solide et incubés une nuit à 30°C. Le lendemain, ces étalements sont repris dans de l'eau et sont déposés selon des dilutions (1 à 10⁻⁴) sur boîtes -LW et -LWH en goutte de 10 µl.

Name	Genotype
Yeast strains	
CG1945	MAT _a , ura3-52, his 3-200, lys2-801, ase2-101, trp1-901, leu2-3, 112, Gal4-542, Gal80-538, cyh-2, Lys2::Gal1 _{UAS} -Gal1 _{TATA} -HIS3; URA3::GAL4 _{17mers(X3)} -CyC1 _{TATA} -lacZ
Y187	MAT _α , ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112, gal4Δ, met ⁻ , gal80Δ, MEL1, URA3::GAL1 _{UAS} -GAL1 _{TATA} -lacZ

ANNEXE 2 : LA METHODE CRAFT.

La technique de CRAFT a été mise au point par P.J.J Hooykaas et A. Vergunst (Sylvius Laboratory de, Université de Leiden, Pays-Bas) pour démontrer la sécrétion d'effecteurs de bactéries intracellulaires *via* le T4SS d'*Agrobacterium tumefaciens*. Nous avons alors réalisé le clonage des différents gènes cibles (*pk1*, *pk2* et *phos*) dans le vecteur pSDM3197 fourni par A. den Dulk-Ras. Pour cela nous avons procédé de la même façon qu'avec le clonage d'inserts dans les vecteurs utilisés en double-hybride. Nous avons amplifiés les inserts par PCR avec une polymérase très fidèle à partir d'extraits d'ADN d'animaux *Armadillidium vulgare* infectés par *wVulC*. Ces produits ont été purifiés par précipitation, coupés avec le couple d'enzymes de restriction adapté, purifiés de nouveau par précipitation puis liés au plasmide pSDM3197 prédigéré par le même couple. Les produits de ligature ont servi à transformer la souche DH5 α d'*E. coli* (Invitrogen). Les clones positifs étaient screenés sur des boîtes de pétri contenant un milieu gélosé LB et l'antibiotique gentamicine (40 mg/L final). Une PCR colonie à confirmer les clones positifs et une PCR de séquençage à partir des minipréparations de plasmides nous ont permis de vérifier la séquence des gènes d'intérêt ainsi que le cadre de lecture en fusion avec la séquence codant la Cre. Les bactéries transformées et ensemencées sur milieu LC ont ensuite été envoyées par courrier aux collègues hollandais.

ANNEXE 3 : L'EXTRACTION ET LA PURIFICATION D'ARN.

Animaux et prélèvement des ovaires.

Les animaux utilisés appartiennent à l'espèce *Armadillidium vulgare* Latreille (1804) hébergeant *wVulC*, souche de *Wolbachia* qui a pour origine le site de Celles-sur-Belle (79, France). Cette population est entretenue au laboratoire depuis plusieurs années dans des salles à 20°C, en photopériode naturelle ou en photopériode stimulante: 18 heures de lumière et 6 heures d'obscurité (18/6 : L/D). Les animaux sont maintenus dans des boîtes (10 x 30 cm) contenant du terreau humide et sont nourris ad libitum (feuilles de tilleuls séchées et carottes). Des femelles infectées par *Wolbachia* ont été sélectionnées selon leur âge, leur taille et le stade de développement de leurs ovocytes de façon à obtenir des échantillons les plus homogènes et reproductibles possibles. Ces ovocytes ont été choisis en fin de phase de vitellogenèse car ils présentent une densité maximale de bactérie (P. Grève, comm. personnelle). Lors du repos sexuel hivernal, les femelles ont été placées plusieurs semaines en photopériode stimulante (L/D : 18/6) afin d'obtenir des ovocytes de taille optimale.

Extraction classique d'ARN totaux.

Les ovaires prélevés par dissection (n=30) sont immédiatement déposés dans des tubes Eppendorf plongés dans l'azote liquide en attendant l'extraction d'ARN. Celle-ci a été réalisée à l'aide du kit RNeasy® Mini Kit (Qiagen) selon les instructions du fournisseur. Les gonades ont été broyées et homogénéisées à l'aide d'un sonicateur dans 350 µL de tampon RLT 10X contenant 1% de β-mercaptoéthanol et 44 U d'inhibiteur de RNAses (Invitrogen). Après centrifugation (12000 x g, 3 min), la phase aqueuse contenant l'ARN est mélangée à 350 µL d'éthanol 70 % par plusieurs pipetages. Le volume total est déposé sur une colonne d'affinité. Après centrifugation (12000 x g, 15 s), la colonne est lavée avec 700 µL de tampon RW1 et 2 fois avec 500 µL de tampon RPE (la composition des tampons RLT, RW1 et RPE n'est pas précisée par le fabricant). Les ARN sont ensuite élués de la colonne par 27 µL d'eau ultrapure et stockés à -80°C. L'ADN contaminant est digéré avec 3 à 5 U de DNase (Ambion) pendant une demi-heure à 37°C. Le dosage de la solution d'ARN est réalisé à l'aide d'un spectrophotomètre par mesure de l'absorbance à 260 nm (A260), et la pureté de l'échantillon est estimée grâce au rapport de A260/A280 à l'aide d'un Nanodrop.

Extraction d'ARN enrichis en ARN bactériens.

Les ovaires de femelles infectées (n=30) sont prélevés puis directement déposés dans 500 µl d'une solution iso-osmotique de PBS-sucrose-glutamine (NaCl 137 mM, Na₂HPO₄ 7.8 mM, KCl 2.7 mM, KH₂PO₄ 1.47 mM pH 7.4, L-glutamine 5 mM, Saccharose 0,25 M) additionnée de 44 U d'inhibiteur de RNAses (Invitrogen) et maintenue sur de la glace. Les ovaires sont ensuite homogénéisés à l'aide d'un broyeur de tissus Potter de Thomas, composé d'un piston en Téflon®, d'un cylindre en verre, et d'une tête de piston à embout rond strié (espace piston-cylindre compris entre 130 et 180 µm). L'homogénat est ensuite centrifugé à 200 X g pendant 15 min, à 4°C, afin de se débarrasser des noyaux et autres débris cellulaires. Le surnageant est prélevé et soumis à deux centrifugations successives dans les mêmes conditions. Le surnageant final est ensuite centrifugé à 4100 g, pendant 15 min à 4°C, sans frein. Le culot résultant correspond à un culot enrichi en bactéries qui est homogénéisé au sonicateur dans une solution de kit Ambion RiboPure. Les ARNs sont alors précipités dans de l'éthanol 100% froid (2,5 V) et de l'acétate de sodium 3 M pH 6,0 (1/9 Vol.). Après centrifugation et séchage à l'air libre, les ARNs sont repris dans de l'eau DEPC et conservés à - 80°C. L'ADN contaminant est digéré avec 3 à 5 U de DNase (Ambion) pendant 30 min à 37°C. La concentration des ARNs totaux est déterminée comme précédemment.

Transcription inverse (Reverse Transcription, RT).

- Moloney-Murine Leukemia Reverse Transcriptase (M-MLV RT)

Les ARNs totaux (1 µg) et 50 pmoles d'amorces non spécifiques (solution d'hexadésoxyribonucléotides de séquence aléatoire, 30 unités de DO/µl) sont mélangés dans 6 µl d'eau bidistillée, puis dénaturés à 65°C pendant 5 min et placés immédiatement sur de la glace. A cette solution, sont ajoutés 4 µl de tampon RT 5X (KCl 375 mM ; MgCl₂ 15 mM ; Tris-HCl 250 mM, pH 8,3), 2 µl de DTT (20 mM), 5 µl de dNTP (2 mM chacun) et 1 µl de M-MLV RT (200 U/µl, Gibco-BRL). Après 1 h d'incubation à 42°C, la solution d'ADNc est stockée à -20°C.

- SuperScript III

La réaction est effectuée dans un volume final de 20 µl contenant 0,2 à 5 µg d'ARN total, préalablement dénaturé à 65°C pendant 5 min, 50 ng d'une solution d'hexadésoxyribonucléotides (amorces aléatoires), 50 pmoles de mélange de désoxyribonucléotides triphosphates (dNTP, 10 µM), du tampon de RT (composition non précisée par le fabriquant), 0,1 M de DTT, 25 mM de MgCl₂ et 200 U d'enzyme SuperScript III RT.

Après addition de l'enzyme, le mélange réactionnel est incubé pendant 10 min à 25°C, 50 min à 50°C et 5 min à 85°C. Enfin, les échantillons sont traités avec 2 U de RNase H et incubés 20 min à 37°C. La solution d'ADNc est conservée à -20°C.

RT-PCR et analyse des produits d'amplification.

L'amplification est réalisée dans un volume de 25 µl, contenant 0,5 µl de solution d'ADNc, 25 pmol de chaque amorce (sens et antisens), du tampon GoTaq 4X (KCl 50 mM ; MgCl₂ 1,5 mM ; Tris-HCl 10 mM, pH 9,0 ; 0,1% de Triton X-100), des dNTP (2,5 mM chacun) et 1,25 U de GoTaq polymérase (Promega). L'amplification est effectuée par 35 cycles correspondant à une dénaturation de l'ADN à 95°C pendant 1 min, une hybridation des amorces à différentes températures (de 47 à 62°C) pendant 1 min, et une élongation à 72°C pendant 2 min. Cette amplification est suivie d'une extension finale de 5 min à 72°C. Les produits d'amplification sont analysés sur un gel d'agarose 1,5% dans du tampon TAE 0,5X (Tris 20 mM ; acétate de sodium 10 mM ; EDTA 0,5mM, pH 8,3) et visualisés grâce au Bromure d'éthidium (BET) sous UV. Leur taille est vérifiée à l'aide d'un marqueur de masse moléculaire approprié.

Northern Blotting.

- Electrophorèse

Les échantillons d'ARN (0,5-20 µg) sont séchés au concentrateur sous vide réfrigéré (Speed Vac, Savant®) pendant une dizaine de minutes à -40°C puis resuspendus dans 10 µL de tampon de charge (formamide 48 % ; formaldéhyde 6,4% ; bleu de bromophénol 0,5 % ; glycérol 6,6 % dans du tampon MOPS 1X (3[N-morpholino]propane sulfonic acid, pH 7,0) (MOPS 10X= MOPS 0,2 M ; acétate de sodium 0,05 M ; EDTA 0,01 M, pH 5,5), et séparés par électrophorèse en conditions dénaturantes sur gel d'agarose 0,8% dans du tampon MOPS 1X (3[N-morpholino]propane sulfonic acid, pH 7,0 / MOPS 5X : MOPS 0,1 M ; acétate de sodium 40 mM ; EDTA 1 mM), auquel sont ajoutés du formaldéhyde (17 %) et 0,57 µg/mL de bromure d'éthidium. La migration est réalisée dans du tampon MOPS 1X sous une tension de 10 V/cm et la visualisation des ARN est réalisée sous lumière ultraviolette. Après migration, le gel est lavé par 2 bains successifs d'eau bidistillée afin d'éliminer le formaldéhyde qui est préjudiciable à la qualité du transfert. Le gel est équilibré pendant 15 min dans un tampon SSC 20X (NaCl 3 M ; citrate de sodium 3 M ; pH 7,0) puis transféré sur une membrane de nylon (Q-Biogene®) dans le même tampon.

- Transfert sur membrane

Le gel d'ARN, après migration, est transféré sous vide pendant 2h à l'aide du VacuGene™ XL (Pharmacia LKB Biotechnology) sur une membrane de Nylon® (Hybond-N+ nylon membrane QBiogene®) dans du tampon SSC 20 X (NaCl 3 M ; citrate de sodium 3 M ; pH 7,0). Les ARN sont ensuite fixés sur la membrane par séchage à 80°C pendant 2h. Les membranes de nylon, sur lesquelles ont été transférés les ARN, ont été réhydratées dans une solution d'EDTA 2 mM pH 8, SDS 0,1 %. Les sites d'adsorption non spécifiques ont été saturés pendant 4 h à 42°C, avec 5 ml de tampon d'hybridation [formamide 40 %, Dextran sulfate 10 %, SSC 4 X, Denhardt 1X (polyvinylpyrrolidone 10 mg/mL, ficoll 400 10 mg/mL), Tris-HCl 20 mM, pH 7,4], contenant 0,3 mg/ml d'ADN de sperme de saumon fraîchement dénaturé. La sonde d'ADNc marquée au 32P a été dénaturée pendant 5 min à 95°C, puis mélangée à 5 ml de tampon d'hybridation. Les membranes sont incubées dans ce mélange pendant 16 à 20 h à 42°C. Les membranes de nylon sont ensuite rincées successivement, sous agitation, dans du SSC 2X, 0,1 % SDS à température ambiante (2 x 30 min), et dans du SSC 0,1X, 0,1% SDS à 52°C (2 x 30 min). La radioactivité sur les membranes a été révélée par exposition au phosphorimager « STORM » (Molecular Dynamics®, CA., USA).

- Marquage radioactif de sondes d'ADNc par amorçage aléatoire (random priming)

Deux sondes, obtenues par amplification par PCR à partir de matrices d'ADN génomique de femelles infectées par *Wolbachia wVulC*, ont été utilisées pour cette étude. Une sonde correspondant au produit de PCR de 900 pb obtenu par amplification de l'ADNc de 16S rDNA de *Wolbachia wVulC*, et une sonde correspondant à l'ADNc partiel codant l'actine d'*Armadillidium vulgare*. Chaque sonde est marquée par la méthode d'amorçage aléatoire. Les sondes sont préparées à partir de 100 ng de produits de PCR purifiés à l'aide du kit Jet Pure (Genomed) selon les instructions du fournisseur. Les fragments d'ADNc sont ensuite dénaturés par la chaleur (95°C pendant 5 min), puis refroidi sur de la glace. La solution est mélangée à 5 µl d'une solution d'hexadésoxyribonucléotides (Pharmacia), 20 µl de tampon de réaction (N-[2-hydroxyéthyl]piperazine-N'-[2-éthanesulfonate] (HEPES) 0.5 M, pH 6.6, MgCl₂ 12 mM, β-mercaptoéthanol 25 mM, Tris-HCl 12 mmol, pH 8), contenant les désoxyribonucléotides dATP, dTTP, dGTP (0.1 mM chacun), 2 µl de la solution de nucléotide radioactif [α-32P]-dCTP, (3 Ci/µmole, 10 µCi/µl, Amersham) et 5 U d'une solution d'ADN polymérase (fragment de Klenow, Gibco-BRL) (Feinberg et Vogelstein, 1983). Le mélange réactionnel est incubé pendant 1 h à 37°C, et la réaction arrêtée par l'addition de 20 µl d'une solution d'EDTA 0.5 M (pH 8). Une fraction de 0,5 µl est prélevée avant l'arrêt de la réaction, et précipitée dans 5 ml d'acide trichloroacétique 10%. Après filtration sur membrane Whatman GF/C, la radioactivité incorporée est mesurée à l'aide d'un compteur à scintillation.

ANNEXE 4 : LA MICROSCOPIE CONFOCALE.

Par ses capacités de réaliser des coupes optiques verticales ou obliques, point par point, la microscopie confocale représente sûrement le développement le plus important de la microscopie optique au cours des 10 dernières années. Grâce à son principe de sectionnement optique, cette technique améliore de façon spectaculaire la résolution dans les plans XY et XZ et permet l'analyse d'échantillons semi transparents jusqu'à 150µm d'épaisseur sans nécessiter de coupe mécanique préalable. Le traitement informatique associé permet d'afficher des images provenant du signal enregistré par un détecteur très sensible et de reconstituer la distribution tridimensionnelle de la fluorescence. L'expérimentation a été réalisée dans le cadre d'une formation d'une semaine proposée par l'école doctorale de Poitiers et avait pour objectif de mettre au point un protocole servant à localiser les bactéries dans les ovocytes hôtes. Les ovaires de femelles d'*A. vulgare* (population ZN) infectées par la souche de *Wolbachia wVulC* ont été prélevés par dissection puis placés sur une lame de verre. Afin de perméabiliser la membrane des ovocytes, les ovaires ont été soumis à une exposition à différents détergents avant et/ou après fixation au PFA (paraformaldéhyde, 0,25% final) :

Temps Détergent	15 min		15 min	30 min
Témoin			0.1%	
Triton X-100	0.1%	Fixation PFA	0.1%	
				0.1%
			0.2%	
NP40			0.1%	
NaBH4			0.1%	

Les noyaux ont ensuite été marqués avec du TOPRO-3 (1:1000, Interchim). Les observations ont été effectuées avec l'aide d'Anne Cantereau (Ingénieur de recherche CNRS dans l'Institut de Physiologie et Biologie Cellulaires (IPBC)) sur le microscope confocal Biorad MRC 1024 de la Plateforme Image de l'Université de Poitiers basée au Pôle Biologie Santé. Les résultats ont été encourageants mais ont révélé une coloration insuffisante des noyaux et de l'actine. Etant donnée la grande taille des ovocytes, la possibilité de faire des micro-injections de marqueurs à l'intérieur même des ovocytes a été avancée.

Système de sécrétion de type IV et protéines à domaines ankyrines dans les interactions *Wolbachia*-arthropodes.

Wolbachia est une bactérie Gram-négative intracellulaire modifiant la reproduction de nombreux arthropodes. Chez l'isopode *Armadillidium vulgare*, la souche *wVulC* entraîne la féminisation des mâles. Nous avons caractérisé deux opérons *vir* s'exprimant dans tous les tissus hôtes et codant un système de sécrétion de type IV (T4SS) pouvant permettre d'exporter des effecteurs bactériens vers le cytoplasme de l'hôte. La comparaison des séquences et de l'organisation des gènes de 37 souches de *Wolbachia* a révélé la forte conservation des deux opérons *vir* suggérant l'importance du T4SS dans la biologie de la bactérie. Nous avons également identifié, dans le génome en cours de séquençage de *wVulC*, 66 gènes codant des protéines à domaines ankyrines. Ces motifs forment des sites d'interactions protéine-protéine chez les eucaryotes et sont supposés être impliqués chez *Wolbachia* dans l'interaction avec des protéines de l'hôte. Nous avons montré qu'une des trois copies du gène *pk2* de *wVulC*, n'est exprimée que chez des souches féminisantes mais chez aucune des 3 souches induisant l'incompatibilité cytoplasmique chez les isopodes terrestres. Ce produit du gène *pk2* pourrait être impliqué dans la féminisation de l'hôte. Toutefois, nous avons réalisé des tests d'interaction par double-hybride en levures et par la méthode CRAFT (Cre-recombinase Reporter Assay for Translocation) entre les protéines du T4SS et cinq protéines à domaines ankyrines dont Pk2 afin de savoir si ces dernières étaient sécrétées par ce système. Les résultats montrent qu'aucun des cinq produits de gènes *ank* testés n'est sécrété par la bactérie mais se révèlent encourageants pour identifier les effecteurs de *Wolbachia*.

Type IV secretion system and ankyrin domain-containing proteins in *Wolbachia*-arthropods interactions.

Wolbachia are intracellular Gram-negative bacteria that are reproductive manipulators of many arthropods. In the isopod *Armadillidium vulgare*, the *Wolbachia wVulC* strain induces male feminization. Here, we characterized two *vir* operons which are expressed in all host tissues and which encode a type IV secretion system (T4SS) used to translocate bacterial effectors into host cytoplasm. Gene organization and sequence comparison in 37 *Wolbachia* strains highlighted the high conservation of both *vir* operons and their importance for the biology of the bacteria. We also identified in the on-going assembly of the *wVulC* genome, 66 ankyrin domain-encoding genes. Ankyrin motifs are known to mediate protein-protein interactions in eukaryotic organisms and thus are suggested to mediate in *Wolbachia* the interaction with host molecules. We showed that one of the three copies of the *wVulC pk2* gene is only expressed in feminizing strains but not in the three strains inducing cytoplasmic incompatibility in terrestrial isopods. The associated Pk2 protein could be involved in male feminization. We thus tested the interaction between three T4SS proteins and five ankyrins (including Pk2) *via* the yeast two-hybrid and CRAFT (Cre-recombinase Reporter Assay for Translocation) methods. None of the five ankyrin proteins were revealed to be secreted by the *wVulC* strain. Nevertheless, this promising approach may enable us to identify *Wolbachia* effectors.

Discipline : Biologie évolutive

Mots-clés : Interaction hôte-parasite, *Wolbachia*, Système de sécrétion, Protéines à domaines ankyrines, Féminisation, Incompatibilité cytoplasmique, Effecteurs bactériens, Crustacés isopodes.

Key words : Host-parasite interaction, *Wolbachia*, Secretion System, Ankyrin domain-containing proteins, Feminization, Cytoplasmic incompatibility, Bacterial effectors, Crustacean isopod.

Intitulé et adresse du laboratoire

Laboratoire Ecologie, Evolution, Symbiose - UMR CNRS 6556
Bâtiment Sciences Naturelles, 1er étage, Université de Poitiers
40, avenue du Recteur Pineau - 86022 Poitiers Cedex – FRANCE.