



**HAL**  
open science

# Évolution des génomes des endosymbiotes chez les insectes phloémophages : le cas d'*Hamiltonella defensa* en interaction avec ses différents partenaires

Pierre-Antoine Rollat-Farnier

## ► To cite this version:

Pierre-Antoine Rollat-Farnier. Évolution des génomes des endosymbiotes chez les insectes phloémophages : le cas d'*Hamiltonella defensa* en interaction avec ses différents partenaires. Microbiologie et Parasitologie. Université Claude Bernard - Lyon I, 2014. Français. NNT : 2014LYO10289 . tel-01130037

**HAL Id: tel-01130037**

**<https://theses.hal.science/tel-01130037>**

Submitted on 11 Mar 2015

**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.



N° d'ordre

Année 2014

## THESE DE L'UNIVERSITE DE LYON

Délivrée par

L'UNIVERSITE CLAUDE BERNARD LYON 1

**Ecole doctorale :** *Evolution Ecosystèmes Microbiologie Modélisation*

Laboratoire de Biométrie et Biologie Evolutive, UMR CNRS

### DIPLOME DE DOCTORAT

(Arrêté du 7 août 2006)

Soutenue publiquement le 24 Novembre 2014 par

**Pierre-Antoine ROLLAT-FARNIER**

# **Évolution des génomes des endosymbiotes chez les insectes phloémophages : le cas d'*Hamiltonella defensa* en interaction avec ses différents partenaires**

### Composition du jury :

Fabrice VAVRE	DR CNRS	Co-directeur de thèse
Marie-France SAGOT	DR INRIA	Co-directrice de thèse
Laurence MOUTON	MC Université Lyon 1	Co-encadrante de thèse
Anastasia TSAGKARAKOU	DR NAGREF	Rapporteur
Claudine MEDIGUE	DR CNRS	Rapporteur
Pierre GREVE	PU Université de Poitiers	Rapporteur
Olivier DURON	CR CNRS	Examineur
Cristina VIEIRA-HEDDI	PU Université Lyon 1	Examineur



# UNIVERSITE CLAUDE BERNARD - LYON 1

## **Président de l'Université**

Vice-président du Conseil d'Administration  
Vice-président du Conseil des Etudes et de la Vie Universitaire  
Vice-président du Conseil Scientifique  
Directeur Général des Services

## **M. François-Noël GILLY**

M. le Professeur Hamda BEN HADID  
M. le Professeur Philippe LALLE  
M. le Professeur Germain GILLET  
M. Alain HELLEU

## **COMPOSANTES SANTE**

Faculté de Médecine Lyon Est – Claude Bernard  
Faculté de Médecine et de Maïeutique Lyon Sud – Charles  
Mérieux  
Faculté d'Odontologie  
Institut des Sciences Pharmaceutiques et Biologiques  
Institut des Sciences et Techniques de la Réadaptation  
Département de formation et Centre de Recherche en Biologie  
Humaine

Directeur : M. le Professeur J. ETIENNE  
Directeur : Mme la Professeure C. BURILLON  
  
Directeur : M. le Professeur D. BOURGEOIS  
Directeur : Mme la Professeure C. VINCIGUERRA  
Directeur : M. le Professeur Y. MATILLON  
Directeur : Mme. la Professeure A-M. SCHOTT

## **COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE**

Faculté des Sciences et Technologies  
Département Biologie  
Département Chimie Biochimie  
Département GEP  
Département Informatique  
Département Mathématiques  
Département Mécanique  
Département Physique  
UFR Sciences et Techniques des Activités Physiques et Sportives  
Observatoire des Sciences de l'Univers de Lyon  
Polytech Lyon  
Ecole Supérieure de Chimie Physique Electronique  
Institut Universitaire de Technologie de Lyon 1  
Ecole Supérieure du Professorat et de l'Education  
Institut de Science Financière et d'Assurances

Directeur : M. F. DE MARCHI  
Directeur : M. le Professeur F. FLEURY  
Directeur : Mme Caroline FELIX  
Directeur : M. Hassan HAMMOURI  
Directeur : M. le Professeur S. AKKOUCHE  
Directeur : M. le Professeur Georges TOMANOV  
Directeur : M. le Professeur H. BEN HADID  
Directeur : M. Jean-Claude PLENET  
Directeur : M. Y. VANPOULLE  
Directeur : M. B. GUIDERDONI  
Directeur : M. P. FOURNIER  
Directeur : M. G. PIGNAULT  
Directeur : M. le Professeur C. VITON  
Directeur : M. le Professeur A. MOUGNIOTTE  
Directeur : M. N. LEBOISNE



## Remerciements

---

Il y a plusieurs mois, lorsque je m'imaginai devant cette page blanche, je me visualisais apaisé et détendu, réfléchissant tranquillement à la liste des gens que je souhaitais remercier. Naïf que j'étais.

Me voilà donc pour de vrai devant cette page déjà un peu noircie, à quelques heures de l'envoi de ce manuscrit aux personnes qui vont juger mon travail. La transition étant toute trouvée, j'en profite pour remercier du fond du cœur ces personnes, tout en me réservant le droit de retirer ces remerciements en fonction de l'opinion qu'ils s'en feront (et toc !).

Je remercie bien évidemment mes directeurs et encadrants de thèse (mais au fait, qui est quoi ?!), qui par leurs nombreux commentaires ont permis la réalisation de ce manuscrit. Merci Marie-France pour ta gentillesse, ta sensibilité et ta confiance, toi qui m'auras permis de prolonger ma thèse dans de bonnes conditions – merci également donc à l'ERC AdG Sisyphe et l'INRIA pour m'avoir financé. Merci à Laurence pour ta sympathie, ta franchise également (c'est sincère !), qui pourrait presque te faire passer pour quelqu'un de bourr... quelqu'un bourré de qualités ! Merci à Fabrice, pour ta disponibilité, tes remontages de moral à quelques moments-clés, ton soutien énorme en cette fin de thèse.

Merci aux GEIHP et aux BAOBAB pour les bons moments passés, à Delphine pour m'avoir réorienté en Bio-Informatique, à Florence pour avoir tant pallié ma « phobie administrative ». Merci à ceux qui m'ont fourni un coup de main de dernière minute (relecture, impression, correction ...). Merci à tous les foteux et toutes les fouteuses, et ceux qui l'étaient moins mais qui nous ont suivis sur le pré.

Le laboratoire, et avant cela l'Université, m'ont permis de faire la connaissance de véritables amis, voire de frères. Aussi j'embrasse Sandrine, Aurélie-Papabis, Greg, Bamba, Lucas, Marine, Adrien, Maxime, David, Alex, Anna, Garance, ainsi que leurs conjoints et leurs enfants. Sans eux, cette aventure aurait été bien plus difficile ;-) Merci pour les soirées, les randos, les discussions philosophiques, politiques, d'Histoire, les parties d'Europa, de tarot, de loup-garou, les coups de main, les moments passés au Bryan's ...

Merci à tous mes amis de ma Bourgogne chérie. Ceux qui ont à un moment ou un autre été sur Lyon pendant mes études (10 ans...), et avec qui j'ai passé des moments... mémorables. Ceux que j'ai moins vus... Merci à Niol, mon vieux frère, à Meum pour ses blagues, sa bonne humeur permanente, Loïc, Benoît 4 bis, Romain 4 bibis, pour nos vacances je ne sais plus où, pour les soirées dans un lieu magique, pour Chez Paco, à Thomas et son guide du Poissard, à Mister Loupi pour tous les moments délirants que tu m'as fait vivre, à Anaïs ma sœur, pour les vacances à Mèze, à sa maman qui m'a accueilli à bras ouverts, à Philippe, mon amigo, pour l'abreu, à Pab, pour ta subjectivité totale devant les matchs de Lyon, à Titiste, à Florent, au Chel, à Mathilde, je suis si heureux qu'on se soit retrouvés, à Benji, à Anaïs bis, à Tadoss, à ton oncle pour le séjour en Ardèche, à Bélouga Boule-de-Gras, à Tripo... euh, à Clément. Merci également à l'OL, une « formidable raison d'être heureux ».

Et puis merci à la Famille. À mes cousins, cousines, oncles et tantes, parrain, marraine. À mes grands-parents, qui assisteront j'en suis sûr à ma soutenance. À ma belle-famille.

À ma mémé Yvonne, que j'ai toujours eu à cœur de rendre fière. À ma Bienne, pour tout ce que tu fais ou a fait pour nous. À Jean-Yves, pour tout ce que tu fais pour « elle », pour ce sourire que tu lui as redonné, pour avoir poussé les portes de notre famille réputée si difficile d'accès :-)

À ma Moumoune, à mes deux frères. Je ne pense pas que des mots puissent retranscrire l'amour que j'ai pour vous.

À « ma » Famille, mon Foyer, Émilie, que j'aurais connue grâce à cette thèse, et qui 4 ans plus tard est là à m'aider à y mettre un point final, des heures durant et malgré la douleur. À sa fille Lise, qui m'apporte tant de bonheur au quotidien. Et des rides aussi...^^

À ma puce, qui a attendu gentiment que papa finisse sa thèse avant de pointer le bout de son nez.

Je te dédie ce manuscrit mon ange, ainsi qu'à ta sœur, dans l'attente de ces moments géniaux que l'on s'apprête à vivre.







## Résumé

---

Les cellules des insectes hébergent de nombreuses bactéries endosymbiotiques, que l'on catégorise généralement selon leur importance pour l'hôte : les symbiotes dits « primaires » sont par définition obligatoires, et synthétisent des nutriments essentiels pour les insectes se nourrissant d'aliments pauvres ou déséquilibrés ; les symbiotes dits « secondaires » sont facultatifs, et utilisent des stratégies mutualistes et/ou la manipulation de la reproduction pour envahir et se maintenir dans les populations d'insectes.

*Hamiltonella defensa* est un endosymbiote secondaire ayant établi deux associations très distinctes chez les insectes phloémophages. Chez les pucerons, la bactérie protège l'hôte contre les parasitoïdes. Elle infecte de nombreux tissus dans l'hôte, et notamment l'hémolymphe, ce qui favoriserait le contact avec les œufs de parasitoïdes. Malgré ce phénotype protecteur, les coûts importants que sa présence inflige à son hôte empêchent sa fixation dans les populations. Chez l'aleurode *Bemisia tabaci*, on ne retrouve la bactérie que dans des cellules spécialisées dans l'hébergement des endosymbiotes, les bactériocytes. Elle s'y trouve entre autres en présence du symbiote primaire, *Portiera aleyrodidarum*, des conditions de vie propices aux échanges entre les deux symbiotes. Elle est fixée dans les populations d'insectes, ce qui suggère un rôle important pour le consortium, qui serait nutritif.

Dans le cadre de cette thèse, nous nous sommes intéressés aux spécificités de chacun de ces systèmes. Tout d'abord, au sein des bactériocytes de *B. tabaci*, nous avons révélé un partitionnement des capacités synthétiques des deux endosymbiotes *H. defensa* et *P. aleyrodidarum*, assorti à de potentielles complémentations métaboliques entre eux et avec leur hôte pour la synthèse d'acides aminés essentiels. Nous avons proposé un rôle nutritif primordial pour *H. defensa*, qui traduirait une transition vers un statut obligatoire pour l'hôte et justifierait sa fixation dans les populations.

Nous nous sommes également attardés sur l'évolution génomique du genre *Hamiltonella*, en comparant des souches infectant *B. tabaci* à une souche de puceron. Nous avons mis en évidence la spécialisation des symbiotes à leur hôte, et révélé que les génomes des endosymbiotes faisaient écho à leur écologie respective. La souche du puceron possède ainsi de nombreux facteurs de virulence et est associée à deux partenaires, un bactériophage et un plasmide de recombinaison. Ces systèmes, inactifs chez le symbiote de *B. tabaci*, sont directement en lien avec la protection et la course aux armements engagés contre les parasitoïdes. À l'inverse, l'avirulence présumée des symbiotes de l'aleurode est cohérente avec leur phénotype nutritif et une transition vers un statut obligatoire pour l'hôte.

Pour finir, nous nous sommes intéressés aux phénomènes d'accélération des taux de mutations chez *H. defensa*, comparativement à son espèce-sœur *Regiella insecticola*, également endosymbiotique et protectrice du puceron. Après avoir éliminé l'hypothèse selon laquelle la transition vers la vie intracellulaire aurait eu lieu indépendamment dans les deux lignées, nous avons tenté d'établir un lien entre ces différentiels d'évolvabilité chez les endosymbiotes et leur contenu en gènes, notamment ceux impliqués dans l'écologie et la réparation de l'ADN.

L'ensemble des résultats obtenus au cours de ce Doctorat ont permis de mieux comprendre l'évolution de l'espèce *H. defensa*, depuis le dernier ancêtre jusqu'aux espèces actuelles, en tâchant de faire le lien entre phénotype de la bactérie et évolution génomique.

## Mots-clés

---

*Hamiltonella defensa* ; évolution génomique ; génomique comparative ; écologie des endosymbiotes



## Abstract

---

Insect cells host many endosymbiotic bacteria, which are in general classified according to their importance for the host: "primary" symbionts are by definition mandatory and synthesize essential nutrients for the insects that feed on poor or unbalanced food sources, while "secondary" symbionts are optional and use mutualistic strategies and/or manipulation of reproduction to invade and persist within insect populations.

*Hamiltonella defensa* is a secondary endosymbiont that established two distinct associations with phloemophagous insects. In aphids, it protects the host against parasitoid attacks. Its ability to infect many host tissues, notably the hemolymph, could promote its contact with parasitoid eggs. Despite this protective phenotype, the high costs associated with its presence within the host prevent its fixation in the population. In the whitefly *Bemisia tabaci* however, this symbiont is found only in cells specialized in hosting endosymbionts, the bacteriocytes. In these cells, it cohabits with other symbiotic species, such as the primary symbiont *Portiera aleyrodidarum*, a proximity that favors potential exchanges between the two symbionts. It is fixed in populations of *B. tabaci*, which suggests an important role for the consortium, probably nutritious.

As part of this PhD thesis, we studied the specificities of each of these systems. First, in the bacteriocytes of *B. tabaci*, we identified a partitioning of the synthetic capacities of two endosymbionts, *H. defensa* and *P. aleyrodidarum*, in addition to a potential metabolic complementation between the symbionts and their host for the synthesis of essential amino acids. We proposed a key nutritive role for *H. defensa*, which would indicate a transition to a mandatory status in relation to the host and would explain its fixation in the population.

We also focused on the genomic evolution of the genus *Hamiltonella*, by comparing the strains infecting *B. tabaci* with a strain infecting the aphids. We highlighted the specialization of the symbionts to their hosts, and found that the genomes of the endosymbionts reflected their respective ecology. The aphid strain thus possesses many virulence factors and is associated with two partners, a bacteriophage and a recombination plasmid. These systems, inactive in the symbiont of *B. tabaci*, are directly related to the protection against and arms race with parasitoids. Conversely, the presumed avirulence of whitefly endosymbionts is consistent with their nutritional phenotype and a transition to a mandatory status to the host.

Finally, we studied the phenomenon of 'accelerated mutation rate' in *H. defensa*, compared to its sister species *Regiella insecticola*, which is also a clade of protective endosymbionts of aphids. After excluding the assumption that the transition to the intracellular life occurred independently in the two lineages, we tried to establish a link between these differences in terms of evolvability in the endosymbionts and of their gene contents, particularly for genes involved in ecology and DNA repair. All the results obtained during this PhD have provided insight into the evolution of the species *H. defensa*, since the last ancestor to the present species, by establishing a link between bacterial phenotype and genomic evolution.

## Keywords

---

*Hamiltonella defensa* ; genomic evolution ; comparative genomics ; endosymbiont ecology



# SOMMAIRE

---

<b>INTRODUCTION</b>	<b>19</b>
<b>I. La Symbiose</b>	<b>20</b>
1. Naissance de la Théorie symbiotique	20
2. Symbiose et innovation	20
3. La vie à deux, la vie en rose ?	22
4. Dynamique des relations hôtes-symbiotes	24
a) Mode de transmission et alignement des intérêts	25
b) Transmission verticale et stratégies égoïstes	27
c) Sélection de partenaires et sanction des tricheurs	28
5. Redessiner les contours de l'individu	29
6. L' holobionte, une entité plurielle aux intérêts convergents ?	31
<b>II. L'endosymbiose chez les insectes</b>	<b>34</b>
1. Pourquoi l'endosymbiose ?	34
2. Symbiose obligatoire chez les insectes	35
a) Régime alimentaire carencé et symbiose obligatoire	35
b) Une relation intime	38
c) Métabolites « symbiotiques »	40
3. Des associations fugaces : les endosymbiotes secondaires	43
a) Modes de transmission	43
b) Tropisme tissulaire des endosymbiotes facultatifs	46
4. Diversité des associations	47
a) Manipulateurs de la reproduction	49
b) Mutualistes	49
c) Mutualisme et conditions environnementales	51
d) Interactions génotypiques	52
5. Interactions entre différentes espèces de symbiotes au sein d'un même hôte	54
a) Maintien des multi-infections	54
b) Tragédie des biens communs	55
c) Régulations et évolution des multi-infections	56
6. Un système simplifié bien complexe...	57
<b>III. Évolution génomique des endosymbiotes</b>	<b>58</b>
1. Symbiogenèse	58
2. Évolution des génomes endosymbiotes	59
a) Écologie des endosymbiotes	59
b) Les premiers stades de l'endosymbiose	61
c) Symbioses anciennes	63
<b>PROBLEMATIQUE</b>	<b>69</b>
<b>MODELE BIOLOGIQUE</b>	<b>71</b>
<b>I. <i>Hamiltonella-Regiella</i> : Un clade d'endosymbiotes protecteurs</b>	<b>71</b>
<b>II. La symbiose puceron-<i>Hamiltonella defensa</i></b>	<b>71</b>
1. Un symbiote versatile	71

2. Un phénotype protecteur répandu	73
3. Des coûts constitutifs	74
4. Bases génétiques et génomiques de cette protection	74
<b>III. Bemisia tabaci et l'endosymbiose</b>	<b>76</b>
1. Ennemi d'états	79
2. Le compartiment endosymbiotique de <i>Bemisia tabaci</i>	79
a) Un endosymbiote obligatoire : <i>Portiera aleyrodidarum</i>	80
b) L'endosymbiose secondaire chez <i>Bemisia tabaci</i>	80
c) Endosymbiotes secondaires à forte prévalence chez l'aleurode	82
<b>ARTICLE PREMIER : REDUCTION GENOMIQUE ET COMPLEMENTATIONS METABOLIQUES POTENTIELLES ENTRE DEUX ENDOSYMBIOTES DE L'ALEURODE BEMISIA TABACI</b>	<b>89</b>
Contexte	89
Attentes et objectifs	90
Abstract	91
<b>I. Background</b>	<b>93</b>
<b>II. Results and Discussion</b>	<b>97</b>
1. Isolation of endosymbionts and DNA amplification	97
2. Sequencing and general features of the <i>Portiera</i> and <i>Hamiltonella</i> genomes	97
3. Functional annotation of the genomes of <i>Portiera</i> and <i>Hamiltonella</i> and metabolic reconstruction	100
4. Essential amino acid synthesis depends on within-pathway complementation between symbionts	103
5. <i>Portiera</i> and <i>Hamiltonella</i> synthesize different vitamins and cofactors	107
6. Transport capabilities in link with exchanged metabolites	108
<b>III. Conclusions</b>	<b>109</b>
<b>IV. Methods</b>	<b>110</b>
1. Whitefly rearing	110
2. Fluorescence in situ hybridization (FISH)	110
3. Purification and amplification of endosymbionts DNA from a single bacteriocyte	111
4. Diagnostic PCR	111
5. Genome sequencing, assembly and annotation	111
6. KEGG (Kyoto Encyclopedia of Genes and Genome) analysis	112
7. Metabolic complementation analysis	112
8. <i>Portiera</i> transporter analysis	113
9. Nucleotide sequence accession numbers	113
Reference	114
Bilan	120
<b>ARTICLE SECOND : DEUX CLADES D'HOTES, DEUX ARSENAUX BACTERIENS : EVOLUTION PAR PERTE DE GENES CHEZ LES ENDOSYMBIOTES FACULTATIFS</b>	<b>123</b>
Contexte	123
Attentes et objectifs	123
Abstract	125
<b>I. Introduction</b>	<b>127</b>
<b>II. Results</b>	<b>130</b>
1. Genome features of hBtab_MEAM1	130

2.	Phylogenetic reconstruction and infection history	131
3.	Gene content evolution in <i>H. defensa</i>	133
4.	Functional losses in the hBtab clade	134
	a) Extra-chromosomal elements	134
	b) Type 3 secretion systems (T3SS)	135
	c) Environmental perception	137
	d) Metabolism of cofactors	138
5.	Functional losses in hApis_5AT	139
	a) Type 2 Secretion System (T2SS)	139
	b) Alpha-fimbriae synthesis	140
	c) Siderophore synthesis	140
	d) Metabolism of amino acids and Isoprenoids	141
6.	Common conservation and loss of virulence systems	141
7.	Rapid evolution and divergence of cell wall components	142
<b>III.</b>	<b>Discussion</b>	<b>145</b>
1.	A <i>Yersinia</i> -like arsenal for a pathogenic ancestor	146
2.	Divergence through gene losses	146
3.	Two bacteria, two arsenals	147
4.	Versatility and self-sufficiency	149
5.	Mutualism and the evolution of evolvability	149
<b>IV.</b>	<b>Conclusions</b>	<b>150</b>
<b>V.</b>	<b>Material and methods</b>	<b>151</b>
1.	Whitefly strains	151
2.	DNA extraction and amplification	151
3.	Assembly and annotation	152
4.	Comparative genomics and outgroup genome	153
5.	Gene families	154
6.	Reconstruction of the ancestor of the genus <i>Hamiltonella</i>	154
7.	Phylogeny of the genus <i>Hamiltonella</i>	155
	Reference	157
	Bilan	164

## **ARTICLE TROISIEME : DIFFERENTIEL DE TAUX D'EVOLUTION CHEZ LES ENDOSYMBIOTES : INFLUENCE DU CONTENU GENIQUE ET DE L'ÉCOLOGIE SUR L'ÉVOLUTION GENOMIQUE**

		<b>167</b>
	Contexte	167
	Attentes et objectifs	167
<b>I.</b>	<b>Introduction</b>	<b>169</b>
<b>II.</b>	<b>Material and methods</b>	<b>173</b>
1.	Choice of an outgroup	173
2.	Genome acquisition	174
3.	Ortholog relationship and ancestor reconstruction	175
4.	Free-living and endosymbiotic core genomes	176
5.	Determination of the substitution rates	176
6.	Gene properties	177
7.	Ancestor reconstruction	177
8.	GO term enrichment	177



<b>III. Results</b>	<b>178</b>
1. Phylogeny reconstruction and substitution rates	178
2. Convergences in gene losses	179
3. Repartition of the losses	180
4. DNA repair genes	181
5. AT enrichment	182
6. Rate of gene losses	183
7. Ecology of the endosymbionts	184
8. Selective pressures	185
<b>IV. Discussion</b>	<b>187</b>
1. Parsimony and comparative genomics	187
2. From an ancestor to another	188
3. Mutation rates in the two lineages	189
4. Host selection level	189
<b>V. Conclusion</b>	<b>191</b>
Reference	192
Bilan	195
<b>DISCUSSION</b>	<b>197</b>
<b>I. Complémentation métabolique entre les endosymbiotes de <i>Bemisia tabaci</i></b>	<b>198</b>
1. Complémentation inter-voies	198
2. Complémentations intra-voies	199
3. Transaminases et plasticité enzymatique	200
<b>II. <i>Hamiltonella defensa</i> : un endosymbiote co-primaire de l'aleurode ?</b>	<b>202</b>
1. Pertes des capacités biosynthétiques chez <i>Portiera aleyrodidarum</i>	202
2. Source de lysine dans l'holobionte ancestral	203
3. Synthèse de la lysine chez <i>Bemisia tabaci</i> ASL	205
<b>III. Évolution de la dépendance</b>	<b>209</b>
1. Redondance et dépendance	209
2. Fixation du nouveau consortium	209
3. Mutualisme apparent	210
4. Symbiose et remplacements	211
a) Remplacements du symbiote néo-primaire chez <i>Bemisia tabaci</i>	211
b) Les remplacements chez les Hémiptères	213
<b>IV. Conséquences de la fixation d'un endosymbiote néo-primaire</b>	<b>218</b>
1. Phénotype étendu : les holobiontes <i>Bemisia tabaci</i> MED et MEAM1	218
2. Phénotype d' <i>Hamiltonella defensa</i> chez <i>Bemisia tabaci</i>	219
<b>V. Niveaux de sélection chez les endosymbiotes</b>	<b>221</b>
1. Rétention de gènes et modalités de la relation mutualiste	221
2. Sélection de clades endosymbiotiques	222
a) Généralités	222
b) Conditions pour que la sélection de clade opère	223
c) « Capture » d'un endosymbiote secondaire	224
d) Dernier ancêtre commun	225
<b>PERSPECTIVES</b>	<b>227</b>
<b>I. Interactions métaboliques chez <i>Bemisia tabaci</i></b>	<b>227</b>

1. Limites de la génomique	227
2. Symbiotes néo-primaires	228
3. Tri infections	229
<b>II. <i>Hamiltonella defensa</i> et sélection de clades</b>	<b>230</b>
<b>CONCLUSION</b>	<b>233</b>
<b>REFERENCES BIBLIOGRAPHIQUES</b>	<b>235</b>



# INTRODUCTION

---

Mon travail de thèse s'est focalisé sur l'évolution d'organismes particuliers : les bactéries colonisant les cellules des insectes. Nous avons utilisé le génome de ces symbiotes comme un véritable témoin de leur histoire évolutive. Ce travail trouve donc sa place dans ce domaine passionnant qu'est la biologie de la symbiose, et a pu se faire grâce aux outils modernes de séquençage des génomes bactériens et de génomique comparative.

Dans un premier temps, nous allons nous attarder quelque peu sur la symbiose. Sa découverte, son caractère ubiquiste, mais surtout sa dynamique et le comportement égoïste des partenaires, sont autant de thèmes qui vont nous intéresser afin de bien comprendre les forces évolutives mises en jeu.

Nous réduirons ensuite notre champ d'étude aux insectes, qui sont des partenaires privilégiés de symbiotes bactériens intracellulaires. Nous aborderons les spécificités de ce modèle insectes-bactéries, nous plaçant du point de vue du procaryote. Nous verrons comment un modèle simple de prime abord, peut cacher en réalité une myriade d'associations toutes plus originales les unes que les autres.

Dans une troisième partie, nous nous intéresserons à un des aspects les plus intéressants de ces bactéries : leur génome, et son évolution. Ces derniers, appartenant parfois à des organismes phylogénétiquement éloignés, présentent en effet des convergences impressionnantes, reflet des conditions de vie subies par les organismes intracellulaires.

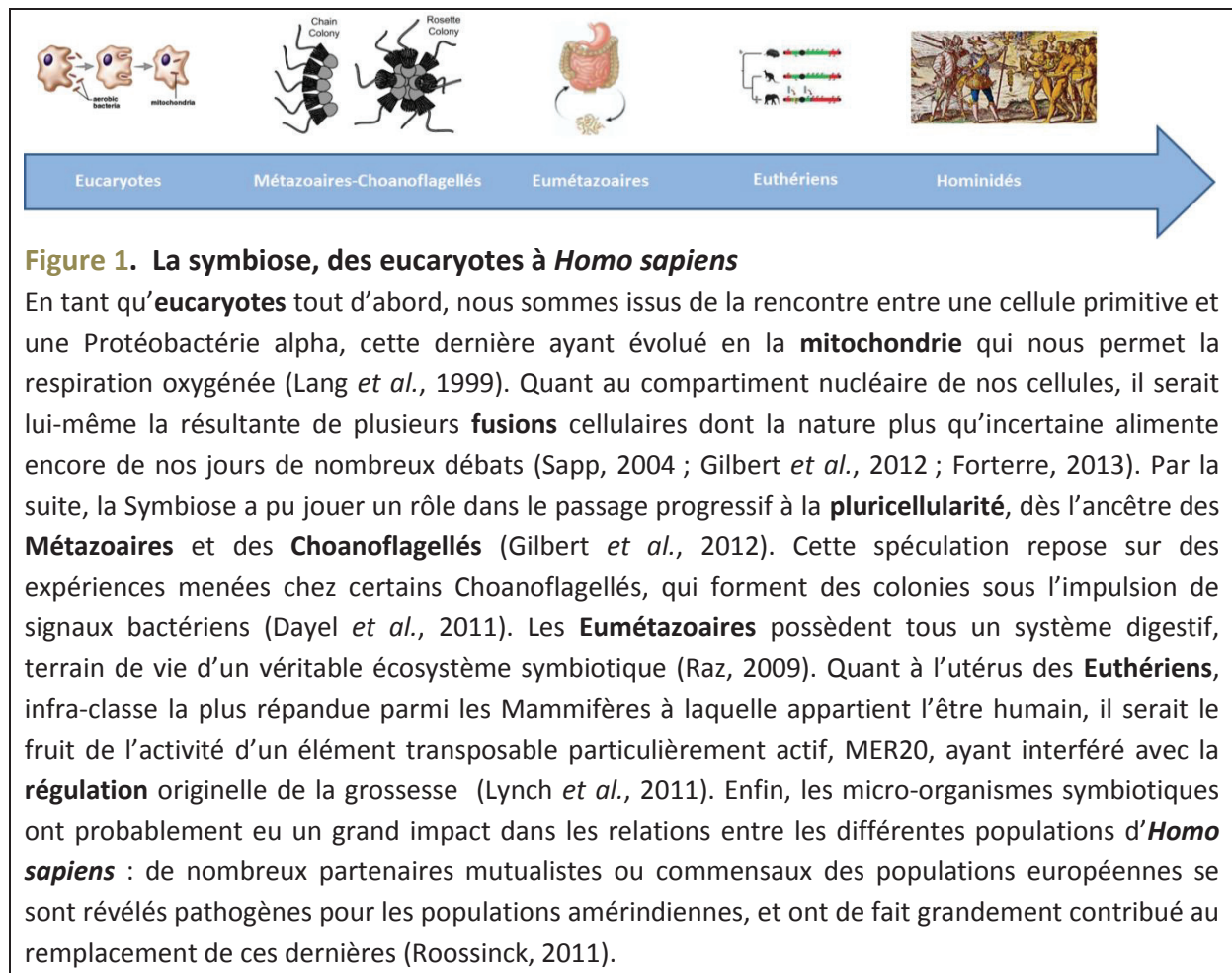
# I. La Symbiose

## 1. Naissance de la théorie symbiotique

Le milieu du XIX<sup>ème</sup> siècle allait marquer un tournant primordial dans la conception du Monde Vivant et des relations entre les espèces. Jusqu'alors, les interactions interspécifiques recensées ne dépassaient guère le cadre quelque peu brutal de la prédation, et il était inconcevable pour l'esprit humain que deux espèces puissent physiquement et intimement coexister. Deux découvertes majeures, œuvres de mycologues de renom, allaient fournir le terreau nécessaire à une évolution des mentalités à ce propos. En premier lieu, une réinterprétation de différentes maladies végétales, imputées au préalable à une dégénérescence de la plante elle-même, permit la découverte de différents champignons pathogènes (Berkeley, 1846). À la suite de cela, le 10 septembre 1867, le récemment promu Professeur Schwendener profita d'un colloque pour offrir courageusement au public ses conclusions quant à ses observations microscopiques des tissus des lichens (revue-hommage dans Honegger, 2000). Désormais, il conviendrait selon lui de considérer ces derniers, non plus en tant que plantes autonomes, mais comme la résultante d'une association entre deux espèces distinctes, un champignon et une algue. Si elle n'a pas encore reçu son nom de baptême, et s'il lui faudra affronter plusieurs décennies de critiques et d'attaques répétées avant d'être pleinement adoptée par la communauté scientifique, la « symbiose » vient pourtant bel et bien de naître en tant que théorie pouvant transcender le caractère monospécifique de l'individu.

## 2. Symbiose et innovation

La symbiose est mère d'innovation évolutive (Margulis et Fester, 1991). Vingt années ne s'étaient pas écoulées depuis les travaux de Schwendener, que déjà de nombreuses symbioses étaient mises à nu : les ectomycorrhizes, les associations entre zoochlorellae ou zooxanthellae d'une part, et protistes ou invertébrés d'autre part, ou encore les nodules racinaires des Fabales (revue dans Honegger, 2000). Les siècles qui suivirent précisèrent son omniprésence, elle qui concerne – pratiquement – tous les clades du vivant (Moya *et al.*, 2008). Par exemple, de nombreuses associations symbiotiques ont gouverné l'évolution tout au long de la branche conduisant à l'espèce humaine, dont quelques exemples sont donnés en **Figure 1**.



La symbiose serait à l'origine d'impressionnantes radiations, notamment chez les végétaux et les insectes. Le chloroplaste, cyanobactérie photosynthétique, est à l'origine du succès évolutif des algues et des plantes, leur permettant de fixer le carbone en utilisant l'énergie solaire (Wernegreen, 2012). La terrestrialisation des végétaux aurait été permise quant à elle par une association avec des champignons mutualistes (Jeffrey, 1962), dont les bénéfices seraient multiples : acquisition de nutriments, protection contre les pathogènes, production d'hormones de croissance etc. (revue dans Taylor et Osborn, 1996). Parmi les plantes terrestres d'ailleurs, les Leguminosae ont élaboré des structures bien connues, les nodules, dans lesquelles elles hébergent des bactéries fixatrices d'azote à qui elles fournissent du carbone en contrepartie (revue dans van Rhijn et Vanderleyden, 1995). Les Gymnospermes et les Angiospermes utilisent la zoochorie, et notamment l'entomochorie, pour la dispersion de leurs graines et propagules diverses (Howe, 1986 ; Kevan et Baker, 1983). Les insectes, en plus d'être le vecteur majeur des propagules d'Angiospermes, ont établi des associations avec nombre d'organismes procaryotes depuis près de 300 millions

d'années (Moran *et al.*, 2008). Ces symbiotes, qui vont nous intéresser tout particulièrement dans le cadre de ce manuscrit, ont permis à leurs hôtes, de par leurs capacités biosynthétiques, d'étendre leur niche à des milieux pauvres en certains nutriments essentiels, tels que le phloème, le xylème, le bois ou le sang, conduisant à d'impressionnantes radiations.

Pour finir, même si la symbiose mitochondriale est qualifiée parfois de « plus vieille symbiose » (Steinert *et al.*, 2000), dans le sens où elle perdure depuis peut-être plus de 2 milliards d'années (Margulis, 1981), il est intéressant d'imaginer les diverses associations qu'ont pu former les formes de vies qui existaient avant l'apparition des eucaryotes. Ainsi, certaines formes de symbioses observées aujourd'hui ont probablement vu le jour dès les premiers balbutiements du vivant. Notamment, les eubactéries ont pu former des relations mutualistes avec certains phages, utilisant la phase lytique ou les toxines de ces derniers pour éliminer les bactéries compétitrices (Bossi *et al.*, 2003 ; Brown *et al.*, 2006) – les virus pourraient en effet être présents sur Terre depuis la soupe prébiontique (Forterre et Prangishvili, 2009). Les archaebactéries sont elles aussi connues pour leurs relations symbiotiques, principalement avec des eucaryotes, mais également avec des eubactéries, une quinzaine de familles d'archévirus, ainsi que d'autres archaebactéries – bien que ce dernier cas semble exceptionnel, une seule association ayant été décrite à ce jour (revue dans Wrede *et al.*, 2012). Quant aux virus eux-mêmes, ce sont obligatoirement des êtres symbiotiques, aussi bien parasites que mutualistes. Cette notion de « symbiotes parasites » nous invite d'ailleurs à proposer une définition claire de ce que nous appellerons « symbiose » dans le cadre de ce manuscrit. Tout cela laisse en tous les cas rêveur sur l'impact potentiel de la symbiose sur l'évolution des micro-organismes « primitifs ».

### **3. La vie à deux, la vie en rose ?**

Donner une définition précise de ce qu'est la symbiose, même après un siècle et demi d'avancées, n'est pas chose aisée. La faute en incombe principalement au caractère trop mélioratif du terme dans le langage courant. Paradoxalement, le terme de « symbiose » [*Symbiotismus*], proposé par Frank dix ans après l'allocution de Schwendener, a justement été créé dans le but de s'affranchir des bénéfices retirés par les deux partenaires (Frank, 1877). Son étymologie se voulait d'ailleurs sans équivoque, puisque « vivre » (grec *βίωσις*)

« ensemble » (grec *σύν*) ne signifie pas nécessairement « vivre en harmonie ». Pour Schwendener lui-même (1869), l'association qu'il observait sous son microscope était de type parasitique. Ainsi, un lichen mettrait aux prises un champignon « parasite, bien que doté de la sagesse des hommes d'état », exploitant ses « Hilotes » d'algues – terme faisant référence aux esclaves de la Sparte antique. Suivant la même logique que Frank (1877), seule la notion d'intimité entre deux espèces distinctes prévaut dans la première définition concrète de la symbiose émise par de Bary (1879). Malgré cela, de nos jours et dans le monde scientifique, deux définitions de la symbiose sont encore en compétition. Pour les partisans d'une définition « généraliste », comme pour de Bary en 1879, la « symbiose » serait toute association interspécifique intime sur le plan physique. Cette définition admet donc le parasitisme, le commensalisme ou le mutualisme comme autant de déclinaisons possibles de la symbiose (voir **Encadré 1**). Une seconde définition restreint celle-ci à toute association de laquelle les deux partenaires retirent un bénéfice. Cette seconde définition correspond donc uniquement au cas du « mutualisme », tel que défini dans l'**Encadré 1**.

#### **Encadré 1. Continuum mutualisme – parasitisme**

Dans le cadre de ce manuscrit, nous considérerons que les symbiotes tirent bénéfice de leur hôte, qui constituera en quelque sorte leur niche écologique. Cet avis ne fait pas l'unanimité, certains évoquant une « domestication » du symbiote par l'hôte, et remettant en cause le bénéfice retiré par la bactérie (exemple dans Baumann, 2005). En ce qui nous concerne, seul le phénotype du symbiote pour l'hôte différera, qui variera du parasitisme (impact **négatif** de la présence du symbiote sur la valeur adaptative de l'hôte) au mutualisme (impact **positif** de la présence du symbiote), en passant par le commensalisme (la présence du symbiote n'a aucune conséquence sur la valeur adaptative de l'hôte). Les triangles rouge et vert indiquent qu'au-delà de ces grandes catégories, toute une gamme existe dans la balance entre coûts et bénéfices : une bactérie commensale puise en réalité dans les ressources de l'hôte et pourrait être considérée en ce cas comme un parasite à la virulence réduite. Parfois, sa présence peut faire barrière à des bactéries pathogènes, de sorte qu'elle soit en ce cas un partenaire mutualiste.

Parasitisme

Commensalisme

Mutualisme

Valeur adaptative de l'hôte

#### **Définitions**

**Valeur adaptative** – ce terme est celui que j'ai choisi afin de traduire le terme anglais *fitness*. Il correspond à la capacité d'un organisme à transmettre ses gènes.

**Balance coûts-bénéfices** – balance entre les coûts imposés à l'hôte par le symbiote, et les bénéfices conférés par ce dernier.



Dans le cadre de cette thèse, la définition de la symbiose que je retiendrai sera celle énoncée par de Bary en 1879. Par commodité, les protagonistes seront qualifiés « d'hôte », organisme hôteur, et de « symbiotes », généralement des microorganismes que celui-ci héberge. Ce choix dans la définition résulte aussi bien d'une « tradition » bien ancrée au sein de mes équipes d'accueil que de ma perception personnelle de la chose. Tout d'abord, vouloir à tout prix classer les différentes relations s'avère extrêmement réducteur, tant les interactions entre « symbiotes » et « hôtes » sont complexes et instables (nous y reviendrons). En outre, la balance des coûts et des bénéfices retirés par les différents partenaires n'est pas toujours évidente à évaluer. S'en suit alors un certain embarras d'ordre sémantique si l'on se refuse à utiliser le terme de « symbiote » pour un partenaire non mutualiste. Deux phrases contiguës dans un article de de Caralt et ses collaborateurs (2007) illustrent parfaitement la confusion que je souhaite éviter. Ainsi, l'introduction de cet article nous explique que de nombreuses levures colonisent les bactériosponges, mais qu'il n'a pas encore été déterminé si ces organismes étaient des parasites ou des symbiotes - cet antagonisme entre « symbiose » et « parasitisme » nous place donc dans la définition restreinte de la **symbiose**, synonyme ici de **mutualisme**. La phrase suivante évoque alors l'abondance de « symbiotes » chez ces éponges, tout en admettant que leur phénotype demeure inconnu - revenant par conséquent de manière inconsciente à une définition plus lâche et permissive de la symbiose.

#### **4. Dynamique des relations hôtes-symbiotes**

En plus d'être difficiles à caractériser, les relations symbiotiques sont extrêmement dynamiques, et ne vont pas être figées le long de ce continuum (Ewald, 1987 ; Sachs *et al.*, 2011). Nous allons nous intéresser aux facteurs qui vont favoriser les transitions évolutives entre mutualisme et parasitisme, dans quelque sens que ce soit. Pour illustrer nos propos, nous allons prendre le point de vue des symbiotes, majoritairement procaryotes, bien que quelques eucaryotes unicellulaires apparaissent de manière sporadique. En effet, dans l'ensemble des symbioses abordées au fil des pages, l'hôte sera toujours considéré comme étant bénéfique pour le symbiote. Il en constituera la ressource principale, de manière directe mais également par l'intermédiaire de son régime alimentaire.

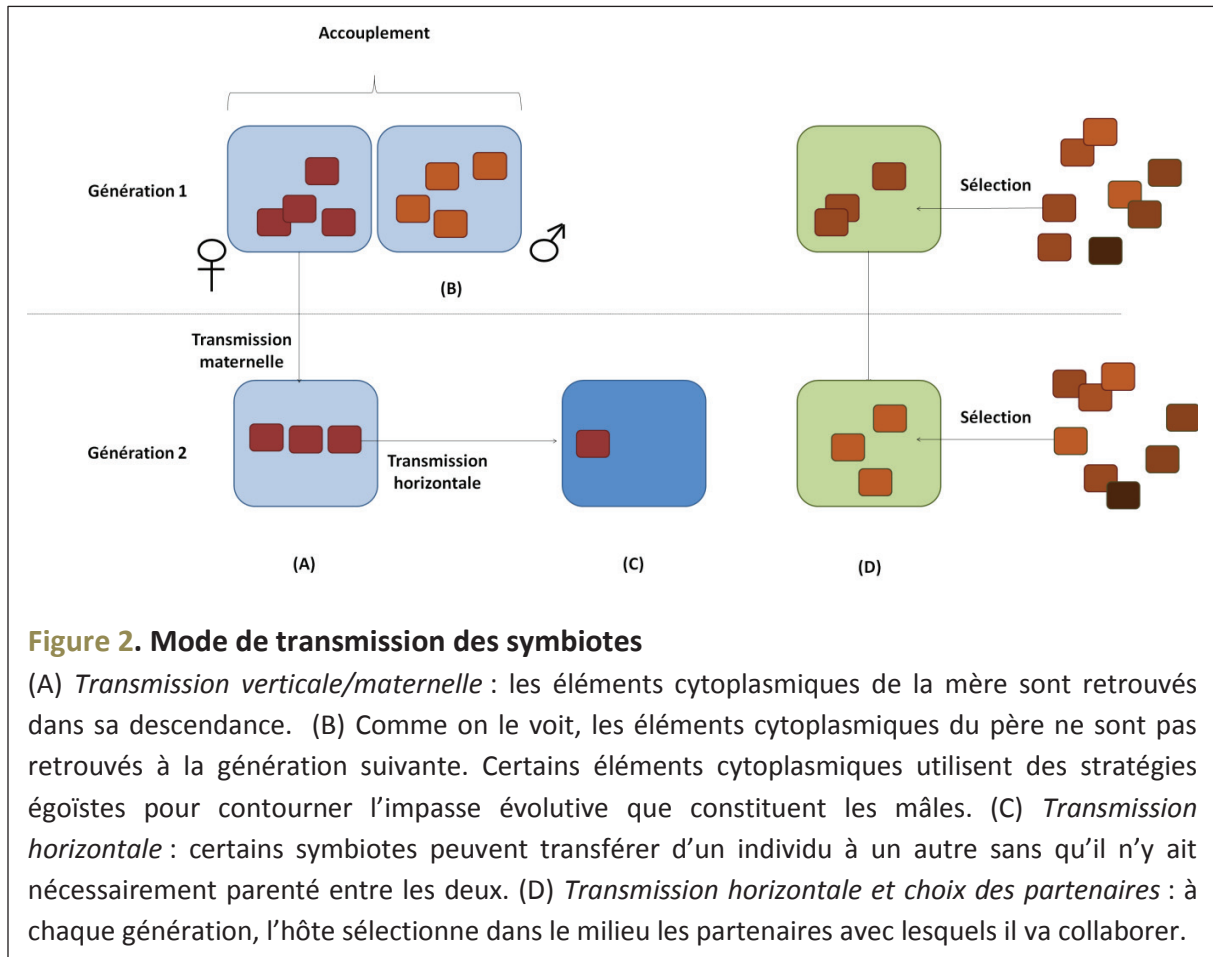
À l'inverse, tout symbiote a forcément un coût physiologique pour son hôte : il consomme une partie de ses nutriments (Shigenobu *et al.*, 2000), malmène son homéostasie (Kremer *et al.*, 2009), ou encore son système immunitaire (Fytrou *et al.*, 2006). Ces coûts vont réduire la valeur adaptative de l'organisme qui les héberge, affectant aussi bien sa fécondité que sa durée de vie (Nachappa *et al.*, 2013 ; Dykstra *et al.*, 2014). En contrepartie, certains symbiotes peuvent conférer des bénéfices à leur hôte, ainsi que nous l'avons abordé avec la respiration assurée par la mitochondrie. Il y a donc une réelle balance entre coûts et bénéfices, dont la valeur définira la nature de la relation.

#### a) Mode de transmission et alignement des intérêts

Le mode de transmission du symbiote est un des déterminants principaux de l'évolution du type de relation qu'il va établir avec son hôte (Ewald, 1987). D'un côté, on parle de symbiote à transmission « verticale » lorsque celle-ci s'effectue depuis l'hôte jusque la descendance de celui-ci (**Figure 2A**). La valeur adaptative du symbiote va ici dépendre de celle de son hôte, ce qui entraîne par conséquent une convergence de leurs intérêts (Sachs et Simms, 2006). Du point de vue du symbiote, cela implique que toute atteinte à la valeur adaptative de l'hôte va se faire en retour à son propre détriment : la virulence du symbiote sera ainsi contre-sélectionnée (Sachs *et al.*, 2014). À l'inverse, toute amélioration de cette valeur adaptative permettra au symbiote de se propager. La transmission **verticale** va donc favoriser la mise en place de stratégies mutualistes (Ewald, 1987). Le cas des organelles est un exemple typique de tels symbiotes, mais nous verrons dans ce manuscrit que les insectes – et plus généralement les Arthropodes – s'associent volontiers avec différentes bactéries mutualistes et à transmission verticale (Moran *et al.*, 2008 ; Baumann, 2005).

Les symbiotes peuvent également transiter entre deux hôtes non-apparentés, voire appartenant à des espèces différentes. Cette transmission, qualifiée d'**horizontale**, ou encore de **latérale**, ne fait pas coïncider les intérêts des deux partenaires. L'hôte constituant une quantité importante de ressources, le symbiote va pouvoir l'exploiter, se répliquer à son détriment, maximisant de fait sa propre valeur adaptative avant de transiter vers un nouvel hôte. La transmission horizontale va donc favoriser les stratégies parasitiques de la part des symbiotes (Fine, 1975). Le cas le plus évocateur concerne les pathogènes tels que *Yersinia pestis*, l'agent de la peste noire, dont le coût est énorme puisque la présence de ces

bactéries va être fatale pour l'eucaryote (Perry et Fetherston, 1997). De manière intéressante, en obligeant la transmission horizontale de *Symbiodinium microadriaticum*, qui est dans la nature une algue unicellulaire mutualiste de la méduse, Sachs et Wilcox (2006) ont révélé à quel point le parasitisme pouvait être « tentant » lorsque disparaissait la transmission verticale.



**Figure 2. Mode de transmission des symbiotes**

(A) *Transmission verticale/maternelle* : les éléments cytoplasmiques de la mère sont retrouvés dans sa descendance. (B) Comme on le voit, les éléments cytoplasmiques du père ne sont pas retrouvés à la génération suivante. Certains éléments cytoplasmiques utilisent des stratégies égoïstes pour contourner l'impasse évolutive que constituent les mâles. (C) *Transmission horizontale* : certains symbiotes peuvent transférer d'un individu à un autre sans qu'il n'y ait nécessairement parenté entre les deux. (D) *Transmission horizontale et choix des partenaires* : à chaque génération, l'hôte sélectionne dans le milieu les partenaires avec lesquels il va collaborer.

Le degré de virulence des pathogènes varie énormément d'une espèce à une autre. Il y aurait pour simplifier une virulence « optimale », qui correspond à une exploitation de l'hôte suffisante pour permettre la réplication du symbiote, mais qui ne sacrifie pas trop rapidement celui-ci (Anderson et May, 1982 ; Ewald, 1983). Si cet optimum varie entre les espèces, c'est qu'il dépend de très nombreux paramètres, entre autres, de l'efficacité du système immunitaire de l'hôte, de sa durée de vie, de la densité en hôtes à infecter, etc. Le mode de transmission horizontal va lui aussi influencer cet optimum : les pathogènes

sexuellement transmissibles ou se propageant par voies aériennes sont généralement moins virulents que les pathogènes transmis par un vecteur (Ebert et Bull, 2003).

### **b) Transmission verticale et stratégies égoïstes**

Nous avons caricaturé quelque peu la corrélation existant entre le mode de transmission et la place de la relation symbiotique sur le continuum mutualisme-parasitisme. Il convient désormais de nuancer cette vision dichotomique de la réalité. Il existe tout d'abord des stratégies égoïstes chez les symbiotes à transmission verticale. Elles concernent des bactéries fascinantes, les manipulatrices de la reproduction (revues dans Werren, 1997 ; Werren *et al.*, 2008 ; Engelstädter et Hurst, 2009). Celles-ci peuvent utiliser deux grands types de stratégies, chacune ayant pour conséquence une augmentation de la proportion de femelles infectées dans la population.

En effet, dans les espèces anisogamiques, la transmission est généralement **maternelle**, les mâles représentant des « impasses évolutives » pour tout élément cytoplasmique, et donc pour ces bactéries manipulatrices (Zeh, 2004). Ils sont par conséquent les victimes désignées de ces bactéries, dont une première stratégie va être de biaiser le sexe-ratio de l'insecte en faveur des femelles, et ce de diverses manières. L'exemple le plus extrême est le masculicide (« *male-killing* » dans la littérature scientifique anglophone). Tuer les mâles peut être une manière détournée de réallouer l'énergie à leurs sœurs, notamment en cas de fortes compétitions à l'intérieur d'une même fratrie et/ou si les femelles sont capables de consommer les cadavres de leurs frères (Engelstädter et Hurst, 2009). Les manipulateurs peuvent également induire la féminisation des mâles génétiques, qui vont dès lors pouvoir transmettre le manipulateur (Bouchon *et al.*, 2008). Cette féminisation peut également avoir lieu chez les hyménoptères haplo-diploïdes (Werren, 1997). Lorsqu'elles ne sont pas fécondées, les femelles (diploïdes) de ces insectes produisent uniquement des descendants mâles, qui sont haploïdes. En cas de présence de manipulateurs, les femelles peuvent produire des femelles diploïdes sans reproduction sexuée (l'on parle de parthénogenèse thélytoque), et qui sont donc cette fois-ci des femelles génétiques et non-plus des mâles (Werren, 1997).

Le second grand type de stratégie est appelé « incompatibilité cytoplasmique » (IC) (Werren, 1997 ; Werren *et al.*, 2008). Lorsqu'infectés par des bactéries générant de l'IC, les mâles génétiques restent des mâles phénotypiques, et sont tout à fait à même de produire une descendance viable avec des femelles infectées. En revanche, leurs accouplements avec des femelles non-infectées ne donnent naissance à aucune descendance. Les lignées matrilineaires infectées auront donc une valeur adaptative supérieure aux lignées non-infectées, du fait de leur capacité à produire une descendance viable avec tous les mâles de la population, quand les femelles non-infectées ne peuvent se reproduire qu'avec les mâles non-infectés.

Ces redoutables phénotypes ont conduit au grand succès des bactéries manipulatrices. Le genre *Wolbachia*, un symbiote de l'Ordre des Rickettsiales et découverte en 1924 par Hertig et Wolbach, pourrait infecter jusqu'à 40% des espèces d'arthropodes selon Zug et Hammerstein (2012), un chiffre « récent » qui est venu tempérer les 66% annoncés dans une précédente étude statistique (Hilgenboecker *et al.*, 2008). Des bactéries des genres *Cardinium*, *Rickettsia*, *Arsenophonus* et *Spiroplasma* présentent elles-aussi des phénotypes manipulateurs (Duron *et al.*, 2008), genres que nous retrouverons en aval dans ce document, et qui appartiennent à des clades très éloignés d'eubactéries. Les manipulateurs fascinent les évolutionnistes, qui s'interrogent par exemple sur l'impact que peuvent avoir de telles bactéries sur l'apparition des nouvelles espèces, notamment par l'intermédiaire de l'incompatibilité cytoplasmique et de l'isolement reproducteur qui l'accompagne (Bordenstein *et al.*, 2001). Elles présentent également un intérêt appliqué, notamment dans le contrôle des populations naturelles d'insectes ravageurs (Zabalou *et al.*, 2004).

### c) Sélection de partenaires et sanction des tricheurs

Nous venons de voir que le mutualisme ne prévenait pas totalement l'apparition de stratégies égoïstes. À l'inverse, la coopération n'est pas incompatible lorsqu'il y a transmission horizontale des symbiotes. Prenons l'exemple de calmars associés à la bactérie *Vibrio fischeri*, dont la phosphorescence permet le camouflage par « contre-illumination » de son hôte (McFall-Ngai, 1990), ou les Rhizobiales fixatrices d'azote qui s'associent avec des plantes Légumineuses (Tamm, 1991). Dans les deux cas, la relation **mutualiste** est pourtant

réinitialisée à chaque génération d'hôtes, ceux-ci ne transmettant pas la bactérie à leur descendance (Wei et Young, 1989 ; Dowling et Broughton, 1986). Il existe toutefois une forme de contrôle de la part de ces hôtes, qui vont discriminer dans les populations libres les futurs partenaires mutualistes (**Figure 2D**), phénomène que l'on qualifie simplement de « choix du partenaire » (Sachs *et al.*, 2004). Ce contrôle sous-tend une communication entre les deux espèces : par exemple, les Légumineuses synthétisent de nombreux flavonoïdes qui vont permettre ou non l'association avec les différentes souches de Rhizobiales, tandis que ces derniers initient le dialogue par l'intermédiaire de systèmes de sécrétion ou de leur paroi cellulaire (revue dans Cooper, 2007).

L'hôte peut effectuer cette recherche des meilleurs partenaires, même une fois ceux-ci hébergés en son sein (Denison et Kiers, 2004). En effet, il peut exister de la variabilité génétique dans les populations symbiotiques, et ce quel que soit leur mode de transmission. Cela signifie que leur bénéfice pour l'hôte est inégal, certains apportant une contribution moindre à la valeur adaptative de celui-ci (Smith et Smith, 1996). Ces bactéries commensales voire parasites sont appelées « **tricheuses** », qui vont bénéficier des ressources de l'hôte tout en s'affranchissant des coûts liés à la rétribution de celui-ci. Cela va leur conférer un avantage sur celles qui jouent le jeu du mutualisme, de sorte qu'elles vont proliférer dans la population symbiotique – au détriment bien évidemment de l'hôte (West *et al.*, 2002). L'hôte peut toutefois discriminer les symbiotes sur la base de leur attitude mutualiste et arrêter de coopérer, voire sanctionner lesdits tricheurs. Par exemple, les Légumineuses peuvent ajuster la disponibilité en oxygène dans les nodules, et sanctionner les bactéries ne fixant pas suffisamment d'azote (Denison et Kiers, 2004).

## 5. Redessiner les contours de l'individu

« *Que seraient les Sciences de la Vie si la symbiose était vue comme la règle, non comme l'exception ?* » (Gilbert *et al.*, 2012)

Certains chiffres sont évocateurs. Près de la moitié du volume d'une éponge est d'origine bactérienne (Vacelet, 1975 ; Taylor *et al.*, 2007). Quinze pour cent des espèces d'insectes dépendent de bactéries intracellulaires pour leur développement et survie (Douglas, 1998 ; Baumann, 2005). Le corail reçoit de la part de ses symbiotes jusqu'à 90% de ses besoins énergétiques (Muscatine et Porter, 1977). Dans un corps humain, 90% des

cellules appartiennent en réalité au Règne des Eubacteria (Bäckhed *et al.*, 2005; Ley *et al.*, 2006) ! Quant aux cellules « Eucaryotes » composant un corps humain, n'oublions pas qu'elles hébergent une Protéobactérie alpha, la susmentionnée mitochondrie, et que le génome nucléaire qu'elles abritent est lui-même composé à 43% d'éléments viraux plus ou moins dégradés (Ryan, 2009). Ainsi, lorsque l'on s'intéresse à la symbiose, les sciences de la vie viennent tutoyer la philosophie, en témoigne cet article de Gilbert et ses collaborateurs (2012) qui pourrait faire un intéressant sujet à l'épreuve du Baccalauréat : « *Nous n'avons jamais été des individus* ».

Dans cette revue, différentes définitions historiques de l'individu animal sont détaillées, puis malmenées sous le prisme de la symbiose. Ainsi, nous ne sommes pas des individus **anatomiquement** parlant (qu'est-ce qu'une vache sans bactéries intestinales ?), ni d'un point de vue du **développement** (le bon fonctionnement de celui-ci étant sous le contrôle de signaux bactériens) ou de la **physiologie** (chez les Vertébrés, une partie du métabolisme des lipides et des vitamines sont de la responsabilité des bactéries). Nous ne sommes pas non plus des individus **génétiques**, les génomes symbiotiques conférant au système une seconde – et importante ! – forme de variabilité ; ni **immunitaires**, les symbiotes étant impliqués à la fois dans le développement de nos systèmes de défense, mais également directement dans la lutte contre les pathogènes. Bref, comme le conclut cette revue, et dans un hommage à l'organisme ayant enfanté la théorie symbiotique, « *nous sommes tous des lichens* ».

Cette notion de l'individu n'est pas importante uniquement d'un point de vue philosophique, mais en ce qui nous concerne, parce qu'elle interroge sur l'entité sur laquelle va agir la sélection naturelle. Ce vieux débat est bien connu. Un termite peut difficilement être considéré comme un individu d'un point de vue sélectif, la colonie entière constituant l'unité reproductive. Mais qu'en est-il de la définition de cette unité si nous projetons d'y inclure les symbiotes nécessaires à la digestion de la cellulose ? Éliminez-les, ou bien éliminez les mitochondries, et « l'individu » eucaryote meurt, qui perd donc du même coup son statut d'individu au profit d'une structure plus complexe. Cette pluralité de l'individu a débouché sur la création d'un terme nouveau, « l'holobionte » (Rosenberg *et al.*, 2007),

regroupant sous la même entité l'hôte et ses symbiotes, et qui serait l'unité principale sélectionnée.

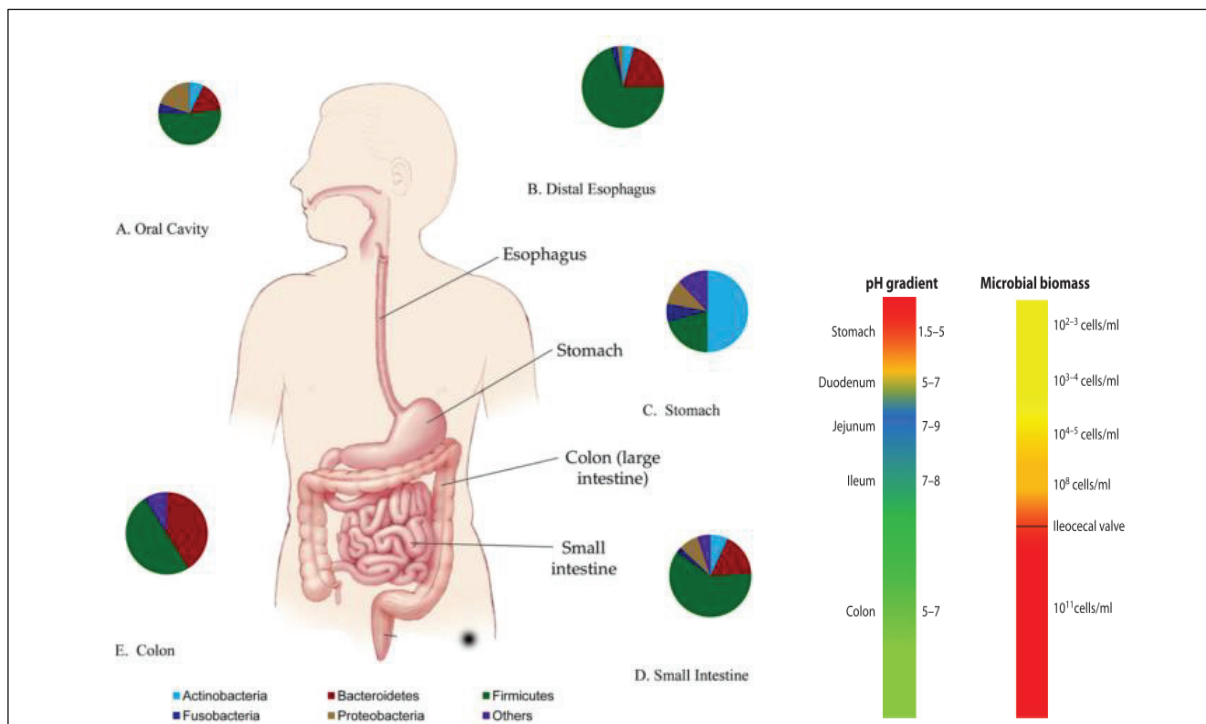
## 6. L'holobionte, une entité plurielle aux intérêts convergents ?

La création du terme « holobionte » dans le but d'impliquer les symbiotes dans le devenir évolutif de l'hôte ne suffit pas pour autant à rendre plus aisée la compréhension des mécanismes de coopération/conflits existant entre les partenaires, et donc de ce qui va constituer l'unité sélectionnée. Pour le moment, j'ai illustré mes propos par des exemples simples, du cas par cas. Cela a permis de donner une première mesure de la diversité et de la complexité des relations symbiotiques, de leur richesse, et de leur dynamisme. En réalité, la relation s'établit rarement entre un hôte et un seul symbiote, mais plutôt entre un hôte et une multitude d'espèces symbiotiques.

Ainsi, dans leur théorie hologénomique de l'évolution, Rosenberg et ses collaborateurs (2007) pointent du doigt le caractère dynamique de cet holobionte : les individus d'une même population n'hébergent pas les mêmes symbiotes, et cette composition symbiotique peut également varier au cours de la vie d'un même individu. Cela sous-tend que des mutations, des pertes et des acquisitions de symbiotes peuvent avoir lieu, mettant quelque peu à mal l'idée d'un devenir partagé entre l'hôte et les symbiotes, et par là même l'unité de sélection que formerait l'holobionte. Peut-on parler d'holobionte lorsque les symbiotes sont majoritairement acquis horizontalement ? Peut-on parler d'holobionte lorsqu'un symbiote aura pour « intérêt » à ce que l'hôte meure dans le but de favoriser sa propre réplication ?

Pour nous convaincre de la complexité de la question, prenons pour exemple un système parmi les plus impressionnants qui soient, le tractus gastro-intestinal de l'« holobionte » vertébré. Ce tube digestif recèle très certainement le **microbiome** le plus riche qui soit (Qin *et al.*, 2010) : de la bouche à l'anus, les différents segments forment autant de niches qui hébergent chacune des quantités considérables de bactéries (**Figure 3**).





**Figure 3. Tube digestif et microbiome humain** (d'après Dave *et al.*, 2012, à gauche, et Walter et Ley, 2011, à droite)

Plus de 400 espèces peuvent composer cet holobionte, bien que de grands groupes dominent largement cette biodiversité, tels les Firmicutes dans l'intestin. On estime que le métagénome microbiotique contient 150 fois plus de gènes que le génome eucaryote (Qin *et al.*, 2010).

Dans pareil écosystème, il est très difficile de distinguer les symbiotes « autochtones » des « allochtones », les permanents des transitoires, les parasites des mutualistes. De même, leur mode de transmission demeure pour grande partie inconnu. Si cet écosystème pris comme un tout à l'air de se comporter en parfait partenaire mutualiste, les phénotypes des symbiotes occupent en réalité l'ensemble du continuum susmentionné. De plus, l'effet de chaque symbiote est en interaction avec celui de l'ensemble du microbiome : tel commensal protège en réalité contre certains parasites, quand à l'inverse, tel mutualiste peut devenir parasite quand d'autres pathogènes sont présents. Il y a donc un nombre dantesque de combinaisons de phénotypes, qui reste difficile à appréhender (lire par exemple Wilks *et al.*, 2013 pour une revue sur les nombreuses associations entre bactéries commensales et virus).

Au regard de cette complexité extrême, nous allons nous focaliser sur un système bien plus simple, mais non moins intéressant. Ce système nous permettra d'appréhender au mieux les relations de coopération et de compétition s'exerçant entre les différents partenaires, que ce soit les symbiotes entre eux, ou avec l'hôte. Surtout, les modes de transmissions des symbiotes seront connus et principalement verticaux, favorisant de ce fait la convergence des intérêts et autorisant d'autant plus à considérer le tout comme un super-organisme tel que celui décrit par Rosenberg et ses collaborateurs (2007). Ainsi, nous avons choisi d'étudier l'évolution de la symbiose entre les insectes phloémophages et des bactéries particulières, les symbiotes intracellulaires.

## II. L'endosymbiose chez les insectes

### 1. Pourquoi l'endosymbiose ?

Nous allons désormais nous focaliser sur les bactéries colonisant les cellules des insectes. Le terme français « étymologiquement correct » pour qualifier ces symbiotes serait « **endocytobiotés** », la racine « cyto » (grec *κύτος*) faisant explicitement référence à leur mode de vie intracellulaire (Nardon et Heddi, 2013). À ce terme, la littérature scientifique préfère toutefois celui plus vague d'« endosymbiotes », qui qualifie plus généralement tout symbiote vivant à l'intérieur de son hôte. J'utiliserai moi-même ce dernier terme tout au long du manuscrit, par habitude et pour ne pas déroger à l'usage. En outre, les articles que j'ai rédigés l'utilisent sans vergogne, et je ne voudrais pas me contredire.

La caractéristique principale des bactéries qui vont nous intéresser est donc leur environnement (principalement) intracellulaire, et en tous les cas un spectre tissulaire réduit. Par opposition aux résidents du tube digestif qui sont confrontés à un environnement variable, la cellule représente un environnement confiné, uniforme, « homéostatique » (Shigenobu *et al.*, 2000). L'espace y est limitant et ne permet pas l'hébergement d'un nombre incalculable d'espèces et phylotypes bactériens. Tout au plus a-t-on observé cinq espèces différentes colonisant la même cellule, à ma connaissance en tous les cas (Subandiyah *et al.*, 2000), auxquelles il convient d'ajouter les mitochondries. Pour une espèce donnée, les individus se comptent en milliers par cellule (Mira et Moran, 2002), bien loin des chiffres édifiants des populations intestinales.

De plus, les endosymbiotes que nous allons étudier (voir l'**Encadré 2**) présentent une autre caractéristique importante : bien que certains soient capables de transférer horizontalement, la composante principale de leur transmission est verticale, c'est-à-dire qu'elle s'effectue majoritairement de la mère vers ses descendants (Baumann, 2005 ; Moran *et al.*, 2008). Cela va évidemment avoir un énorme impact sur la nature de la relation qu'ils entretiennent avec leur hôte, et va favoriser le mutualisme comme nous l'avons déjà vu, bien qu'il ne faille pas oublier le nombre impressionnant d'espèces d'insectes infectées par des bactéries manipulatrices de la reproduction.

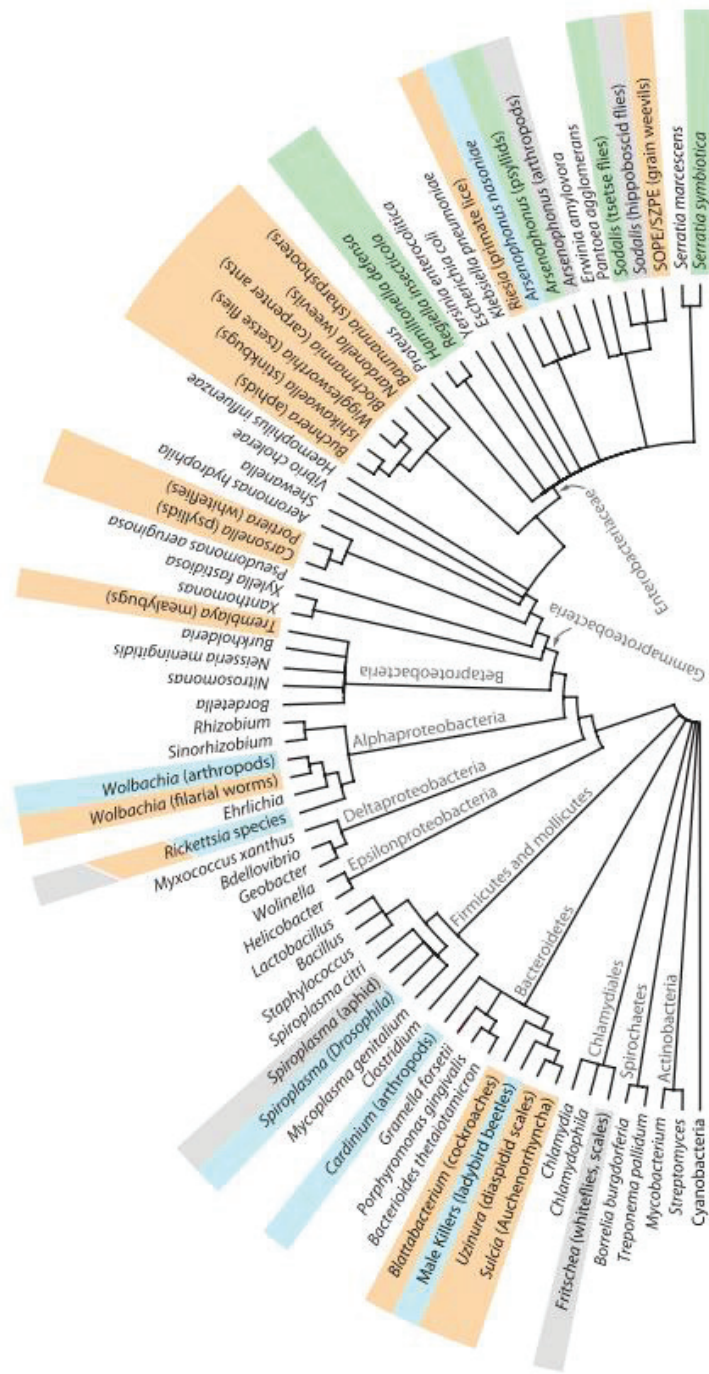
## Encadré 2. La nomenclature des endosymbiotes

La quasi-totalité des bactéries endosymbiotiques n'étant pas cultivables, elles sont nommées sur la base d'un terme introduit par Murray et Schleifer en 1994 : « *Candidatus* ». Ainsi le symbiote primaire du puceron répond-il au joli nom de « *Candidatus Buchnera aphidicola* », le genre (*Buchnera*) et l'espèce (*aphidicola*) ne devant pas figurer en italique lorsque précédés de ce terme. Or, dans l'usage, cette même bactérie pourra être nommée de différentes manières : selon la nomenclature recommandée par Murray et Schleifer (abrégée conventionnellement en « *Ca. Buchnera aphidicola* »), selon la nomenclature linnéenne classique (« *Buchnera aphidicola* », abrégé en « *B. aphidicola* »), voire seulement par son genre (« *Buchnera* »), et ainsi être privée du nom de son espèce – cette situation est très fréquente dans la littérature ayant trait à l'endosymbiose. Très rarement, ce choix est clairement mentionné (Rosas-Pérez *et al.*, 2014 ; Gruwell *et al.*, 2012). Pour ma part, devant cet imbroglio, j'ai opté pour la solution médiane, c'est-à-dire la nomenclature linnéenne (excepté lorsque le nom de l'espèce n'aura pas été référé dans la publication citée). Quant au terme « *Candidatus* », il a sûrement sa raison d'être, mais à dire vrai celle-ci m'échappe. Pour déposer officiellement une nouvelle espèce bactérienne, et non un « *Candidatus* », il faut en effet relever différents aspects relatifs à ses capacités de culture sur milieu artificiel. Or, devant l'explosion des techniques de séquençage et d'assemblages des génomes, nous allons nous retrouver avec un pourcentage très important de bactéries non-cultivées (voire non-cultivables). Pourtant, elles n'en seront pas moins caractérisées par une position taxonomique claire, garantie par la puissance d'analyses phylogénomiques – pouvant utiliser non seulement l'ARN ribosomique 16S, mais également des centaines de gènes/protéines. Il me semble qu'une bactérie dont le génome a été séquencé est tout aussi appréhensible par la communauté scientifique qu'une bactérie cultivée et dont seuls sont connus la séquence de l'ADN ribosomique 16S et le comportement sur milieu gélosé...

## 2. Symbiose obligatoire chez les insectes

### a) Régime alimentaire carencé et symbiose obligatoire

Si l'insecte s'impose comme un modèle d'étude idéal, c'est parce qu'il semble être un hôte idéal pour les endosymbiotes (**Figure 4**). On estime ainsi à plus de 50% le nombre d'espèces d'insectes infectées par de tels partenaires (Van Dijk, 2008) (j'exclurai désormais moi aussi les mitochondries de la classe très prisée des endosymbiotes bactériens). Cependant, à l'intérieur même de la Classe des Insectes se trouvent des « points-chauds symbiotiques ». C'est le cas des insectes dont le régime alimentaire ne subvient pas à tous leurs besoins (sang, sève végétale), et qui dépendent obligatoirement de ce fait de symbiotes nutritifs (Buchner, 1965 ; Nogge, 1982 ; Baumann, 2005 ; Moran *et al.*, 2008). Ce dernier est conventionnellement appelé symbiote « obligatoire » ou « primaire ».



**Figure 4. Phylogénie des endosymbiotes d'insectes** (d'après Moran *et al.*, 2008)

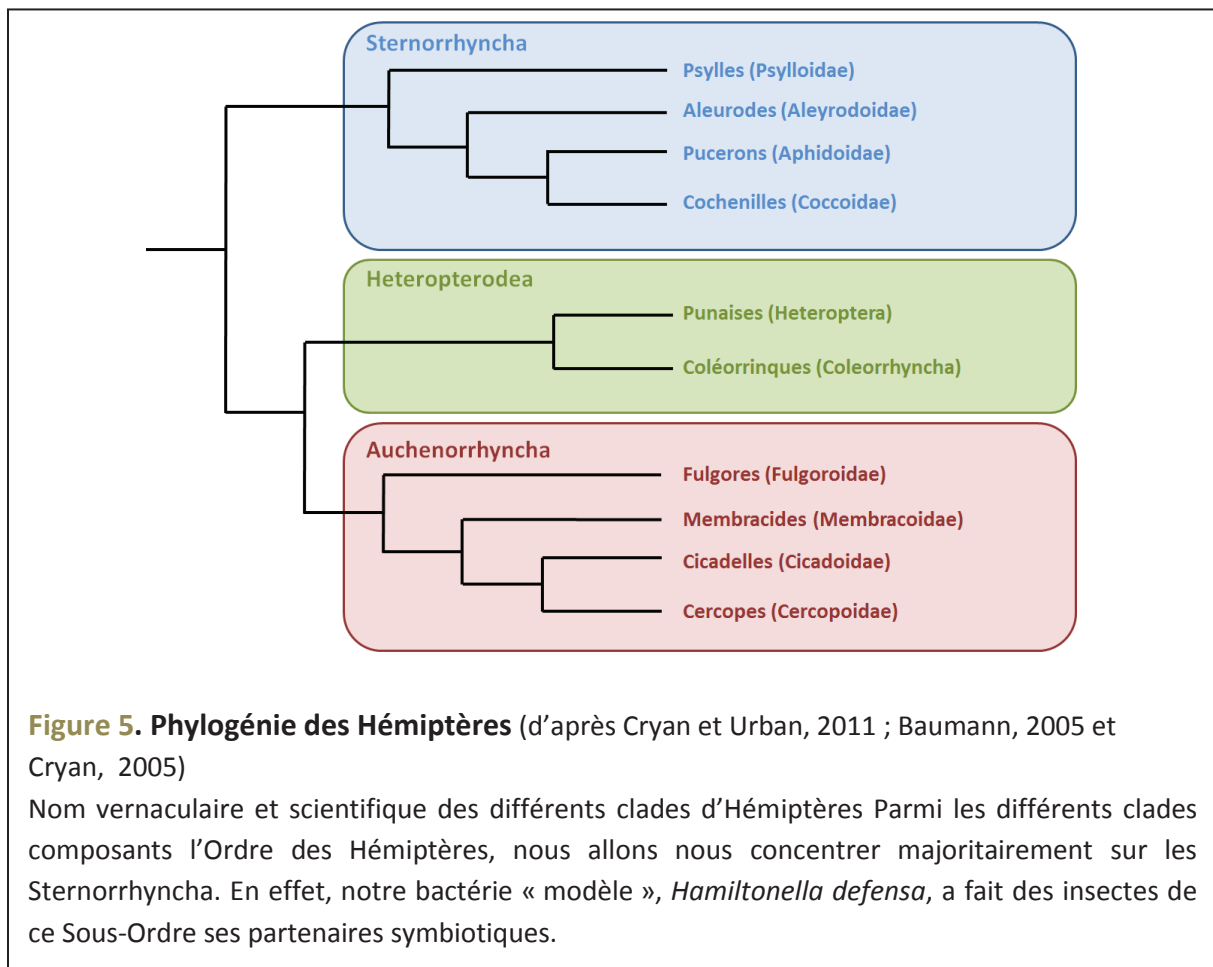
Les symbiotes obligatoires sont indiqués en orange, les facultatifs en vert (mutualistes), bleu (parasites), et gris (phénotype inconnu).

Le phloème, composant de la sève, est justement un exemple de régime alimentaire déséquilibré (Douglas, 1993). Associé au xylème, il constitue le réseau vascularisé des Trachéophytes, impliqué dans le transport des photoassimilats depuis les feuilles matures jusqu'aux racines (Schulz, 1998). S'il est riche en métabolites organiques et surtout en sucres qui en font une cible de choix pour les insectes ravageurs (Brodbeck *et al.*, 1993), il ne prodigue pas en quantité satisfaisante l'ensemble des **acides aminés essentiels** requis (**Tableau 1**).

<u>Acide aminé</u>	<u>Abbréviation</u>
Arginine	Arg
Histidine	His
Isoleucine	Ile
Leucine	Leu
Lysine	Lys
Méthionine	Met
Phénylalanine	Phe
Threonine	Thr
Tryptophane	Try
Valine	Val

**Tableau 1. Acides aminés essentiels chez les insectes**  
 Les insectes, comme tous Métazoaires, ne sont pas capables de synthétiser le squelette carboné de tous les acides aminés qui entrent dans la composition des protéines. Ces acides aminés sont qualifiés d'« essentiels », et doivent par conséquent impérativement être présents en quantité suffisante dans le régime alimentaire de l'animal. Leur définition est propre à chaque espèce selon ses capacités métaboliques, et peut même changer au cours de la vie de l'animal.

Dans toute l'histoire connue du vivant, un seul groupe de Métazoaires a réussi à en faire son alimentation unique : certains insectes de l'ordre des Hémiptères (**Figure 5** ; Douglas, 2006). Dès 1965, Buchner imagina la symbiose comme une réponse probable à cette impasse métabolique. En effet, ces insectes ont pour particularité d'héberger obligatoirement en leur sein un endosymbiote dont ils dépendent, responsable de la synthèse de métabolites essentiels présents en faible quantité dans la sève (Buchner, 1965 ; Baumann, 2005 ; Moran *et al.*, 2008). Cette extension de la niche primordiale de l'insecte est en parfaite adéquation avec l'idée selon laquelle la symbiose est une source formidable d'innovation évolutive.



Les associations que nous allons détailler vont principalement concerner ces insectes se nourrissant de phloème, ce qui nous permettra de nous familiariser avec les notions qui seront importantes chez notre modèle d'étude. Toutefois, certaines associations entre endosymbiotes et insectes **non** phloémophages seront parfois proposées si elles s'avèrent les plus à mêmes d'illustrer mes propos.

### b) Une relation intime

Le symbiote primaire est par définition nécessaire à la survie de son hôte, et est par conséquent obligatoirement transmis de manière verticale à la descendance de l'insecte. La congruence entre les phylogénies des deux partenaires est parfaite – c'est-à-dire que l'on peut superposer les deux arbres – ce qui suggère un événement unique d'infection de l'hôte par le symbiote, suivi d'une co-cladogénèse entre les espèces. L'association entre eux peut ainsi être amenée à perdurer : les symbioses les plus vieilles chez les Hémiptères pourraient avoir été initiées il y a près de 270 millions d'années (**Tableau 2**).

Clade de l'hôte	Hôte	Symbiote primaire	Clade du symbiote	Âge de l'association
Hémiptère	Puceron	<i>Buchnera</i>	Protéobactéries Gamma	180 Ma
	Aleurode	<i>Portiera</i>	Protéobactéries Gamma	180 Ma
	Psylle	<i>Carsonella</i>	Protéobactéries Gamma	120 Ma
	Cochenille	<i>Tremblaya</i>	Protéobactéries Béta	40 Ma
	Auchenorrhyncha	<i>Sulcia</i>	Bacteroidetes	270 Ma
	Punaise	<i>Ishikawaella</i>	Protéobactéries Gamma	Indéterminé
Diptère	Mouche tsétsé	<i>Wigglesworthia</i>	Protéobactéries Gamma	40 Ma
Hyménoptère	<i>Camponatus</i> spp.	<i>Blochmannia</i>	Protéobactéries Gamma	50 Ma
Coléoptère	Charançon	<i>Nardonella</i>	Protéobactéries Gamma	Indéterminé
Blattoptère	Blatte	<i>Blattabacterium</i>	Bacteroidetes	150 Ma

**Tableau 2. Symbioses obligatoires anciennes (d'après Moran *et al.*, 2008)**

Ce tableau ne présente que les symbioses obligatoires les plus anciennes dans les grands clades d'hôtes cités, et est loin de retranscrire la diversité de ces symbiotes obligatoires. Chez les Auchenorrhyncha par exemple, on retrouve des Protéobactéries Béta des genres *Vidania*, *Zinderia*, *Nasuia*, des Protéobactéries Gamma du genre *Baumannia*, et même des pseudo-levures symbiotiques chez les Delphacidea et les Deltocephalinae (Bennett et Moran, 2013). On retrouve ces pseudo-levures chez les pucerons de la sous-famille des Cerataphidinae (Vogel et Moran, 2013), et des Hormaphidinae (Fukatsu et Ishikawa, 1996). Certains pucerons sont également infectés par des souches obligatoires de *Serratia symbiotica*, une Protéobactérie Gamma (Lamelas *et al.*, 2011).

Le tropisme tissulaire des endosymbiotes primaires est extrêmement réduit. Témoin de la pérennité de la relation intime entre les insectes et ces bactéries, la quasi-totalité des espèces dispose d'un organe alloué à l'hébergement de ce dernier et à l'échange de métabolites : le bactériome (Buchner, 1965). Chez les insectes phloémophages, il s'agit d'un organe généralement bilobé accolé au tube digestif, qui peut varier quelque peu dans sa morphologie et les différents types de cellules qui le composent (Douglas, 1989). Les cellules qui hébergent le symbiote primaire sont appelées « bactériocytes », et sont qualifiées de « primaires » lorsque d'autres bactériocytes existent et n'hébergent pas ce symbiote. Si l'on



sort des insectes phloémophages, l'on retrouve un organe bilobé entourant le tube digestif de la mouche tsétsé (Aksoy, 1995), alors que chez les fourmis du genre *Camponotus*, les bactériocytes s'intercalent entre les cellules intestinales, et ne forment pas un organe distinct (Schröder *et al.*, 1996). Bien que présentant des différences dans leur localisation et leur organisation, l'apparition de telles cellules symbiotiques dans des clades éloignés d'insectes représente tout de même un bel exemple de convergence évolutive.

Le rôle d'organe réceptacle de la symbiose du bactériome est couplé à une action protectrice : l'homéostasie à l'intérieur des cellules bactériocytaires confine le symbiote dans un environnement stable, et le préserve du système immunitaire eucaryote (Braendle *et al.*, 2003). Dans certains cas, le bactériome peut agir comme le vecteur de la transmission verticale des endosymbiotes. Chez l'adulte femelle, des bactériocytes vont ainsi migrer dans les oocytes nouvellement formés, assurant de cette manière une transmission à sa descendance (Costa *et al.*, 1996 ; Coombs *et al.*, 2007). Cependant, on retrouve là encore une diversité dans le mode de transmission du symbiote primaire. Chez la mouche tsétsé, des glandes à lait transfèrent le symbiote à la larve lors de sa maturation intra-utérine (Attardo *et al.*, 2008). Dans ce cas, le symbiote primaire infecte autant les bactériocytes que ces glandes nourricières chez les femelles, les premiers afin d'assurer son rôle nutritif, les seconds sa transmission (Balmand *et al.*, 2013). Dans d'autres cas, cette transmission est permise par une migration du symbiote lui-même dans le nouvel embryon, tel que c'est le cas chez le puceron avec le symbiote primaire *B. aphidicola* (Koga *et al.*, 2012).

### **c) Métabolites « symbiotiques »**

Outre l'hébergement et la transmission du symbiote primaire, le bactériome va être le siège d'intenses échanges entre l'hôte et la bactérie (The International Aphid Genomics Consortium, 2010). Le symbiote primaire est en effet dépendant de nutriments que l'hôte va mettre à sa disposition, et va synthétiser en échange des métabolites essentiels absents – ou du moins rares – dans le phloème. Nous allons pour notre part nous focaliser sur l'apport de la bactérie à son insecte d'hôte. Parmi les nombreuses associations que l'on peut trouver chez les Hémiptères, un système en particulier a monopolisé l'attention des chercheurs, l'association puceron/*Buchnera aphidicola* (Douglas, 1998). Le nom de baptême du symbiote est un hommage à Paul Buchner, du fait de son rôle important dans l'étude de la Symbiose

chez les insectes, et notamment les Hémiptères (Munson *et al.*, 1991). On retrouve ce symbiote chez pratiquement toutes les espèces de pucerons (Shigenobu *et al.*, 2000). Cette ubiquité d'un symbiote primaire met bien en évidence son caractère primordial pour son hôte. Pour preuve, au niveau de l'individu, des insectes **aposymbiotiques** générés par prise d'antibiotiques présentent une fécondité drastiquement réduite, sinon nulle (Griffiths et Beck, 1974). Différentes expériences combinant aussi bien des complémentations alimentaires que des analyses métaboliques, moléculaires et génomiques ont permis de prouver le caractère nutritif de ce mutualisme chez le puceron (revue dans Douglas, 1998).

#### *Acides aminés essentiels*

Une première manipulation consiste à nourrir les insectes aposymbiotiques au moyen de régimes artificiels, afin d'essayer de compenser la perte du symbiote. Ainsi, plusieurs expériences ont démontré que les acides aminés essentiels étaient au cœur de la symbiose primaire des Hémiptères (revue dans Douglas, 1998). Chez le puceron *Myzus persicae*, des insectes aposymbiotiques – c'est-à-dire privés de *B. aphidicola* – ont une survie et une croissance pratiquement nulle lorsque nourris sur des régimes dépourvus de n'importe lequel d'entre les dix acides aminés essentiels (Mittler, 1971). Quant aux insectes symbiotiques, ils sont affectés par des carences en seulement trois acides-aminés, ce qui suppose que les populations naturelles les trouvent en quantité suffisante dans leur régime alimentaire (Dadd et Krieger, 1968).

Pour synthétiser ces acides aminés, le symbiote primaire *B. aphidicola* utiliserait un acide aminé non-essentiel présent dans des concentrations importantes dans le phloème des plantes : le glutamate chez le puceron *A. pisum*, l'asparagine chez *M. persicae* (Jiang *et al.*, 2013). Il n'est cependant pas capable d'utiliser les déchets azotés produits par son hôte (Wilkinson et Douglas, 1995), au contraire de certains symbiotes primaires. Par exemple, chez les cochenilles non-excrétrices de miellat de la famille des Diaspididae, il est possible que le symbiote *Uzinura diaspidicola* soit capable de recycler l'azote à partir de l'acide urique afin qu'il soit réutilisé dans la synthèse des acides aminés (Gruwell *et al.*, 2012). Les bactéries du genre *Sulcia* pourraient également recycler les déchets des insectes Auchenorrhyncha, et jouer un rôle central dans l'assimilation de l'azote (McCutcheon et Moran, 2007).

Les techniques moléculaires ont étayé ces premiers résultats. Les techniques de clonage/séquençage et les analyses fonctionnelles ont permis de révéler les nombreux gènes intervenant dans la synthèse d'acides aminés essentiels chez *B. aphidicola* (Kolibachuk et Baumann, 1995). Elles ont également permis de détecter deux plasmides impliqués respectivement dans la synthèse de la leucine et du tryptophane (Lai *et al.*, 1994). Enfin, le séquençage et l'assemblage du génome complet de *B. aphidicola* a révélé une bactérie au métabolisme pratiquement entièrement dédié à la synthèse d'acides aminés essentiels (Shigenobu *et al.*, 2000). Depuis, les assemblages des génomes de différents symbiotes primaires sont arrivés à la même conclusion, quel que soit le clade d'appartenance du symbiote (Nakabachi *et al.*, 2006 ; Sloan et Moran, 2012 ; Rosas-Pérez *et al.*, 2014).

#### *Autres métabolites*

Outre les acides aminés essentiels, d'autres nutriments pourraient être synthétisés par la bactérie, et notamment des vitamines et cofacteurs. Chez les mouches tsétsés hématophages, la vitamine B est par exemple un composant essentiel fourni par le symbiote primaire *Wigglesworthia glossinidia* (Nogge, 1982). Chez les insectes phloémophages, il existe une grande variabilité dans leur capacité à synthétiser ces éléments. Les différentes souches de *B. aphidicola* varient ainsi énormément dans leur contenu génique en rapport à leur synthèse (Lamelas *et al.*, 2011). De même, le symbiote primaire *Baumannia cicadellincola* (symbiote de cicadelles) possède 83 gènes impliqués dans la synthèse de ces éléments (Wu *et al.*, 2006), quand certaines souches de *Tremblaya princeps* (symbiotes de cochenilles) ne possèdent pas les capacités de synthétiser quelque vitamine que ce soit (López-Madrugal *et al.*, 2011). Outre les preuves génomiques, la manipulation du régime alimentaire a démontré que *B. aphidicola* était impliqué dans la synthèse de la riboflavine (vitamine B2) chez *A. pisum*, alors que le puceron *M. persicae* acquiert ce métabolite à travers son alimentation (Nakabachi et Ishikawa, 1999).

Pour finir, les gènes codant la synthèse d'autres métabolites ont été rarement identifiés dans les génomes des endosymbiotes primaires d'insectes phloémophages. Toutefois, les génomes de certaines souches de *B. aphidicola* possèdent les gènes codant pour la synthèse de la queuosine, un nucléoside rare que les eucaryotes doivent normalement acquérir depuis leur régime alimentaire (Jiang *et al.*, 2013).

### 3. Des associations fugaces : les endosymbiotes secondaires

Outre le symbiote primaire, de nombreux autres partenaires transitoires nichent dans les cellules des Hémiptères et autres insectes, et sont qualifiés de « secondaires ». Ils présentent des originalités dans leur mode de transmission, dans leur localisation dans l'hôte, mais également dans la nature de leur relation avec celui-ci, qui contrairement à celle qu'entretient le primaire, oscille entre parasitisme et mutualisme (Moran *et al.*, 2008). Nous allons consacrer plusieurs sections à l'écologie de ces endosymbiotes, qui constitueront notre modèle d'étude principal.

#### a) Modes de transmission

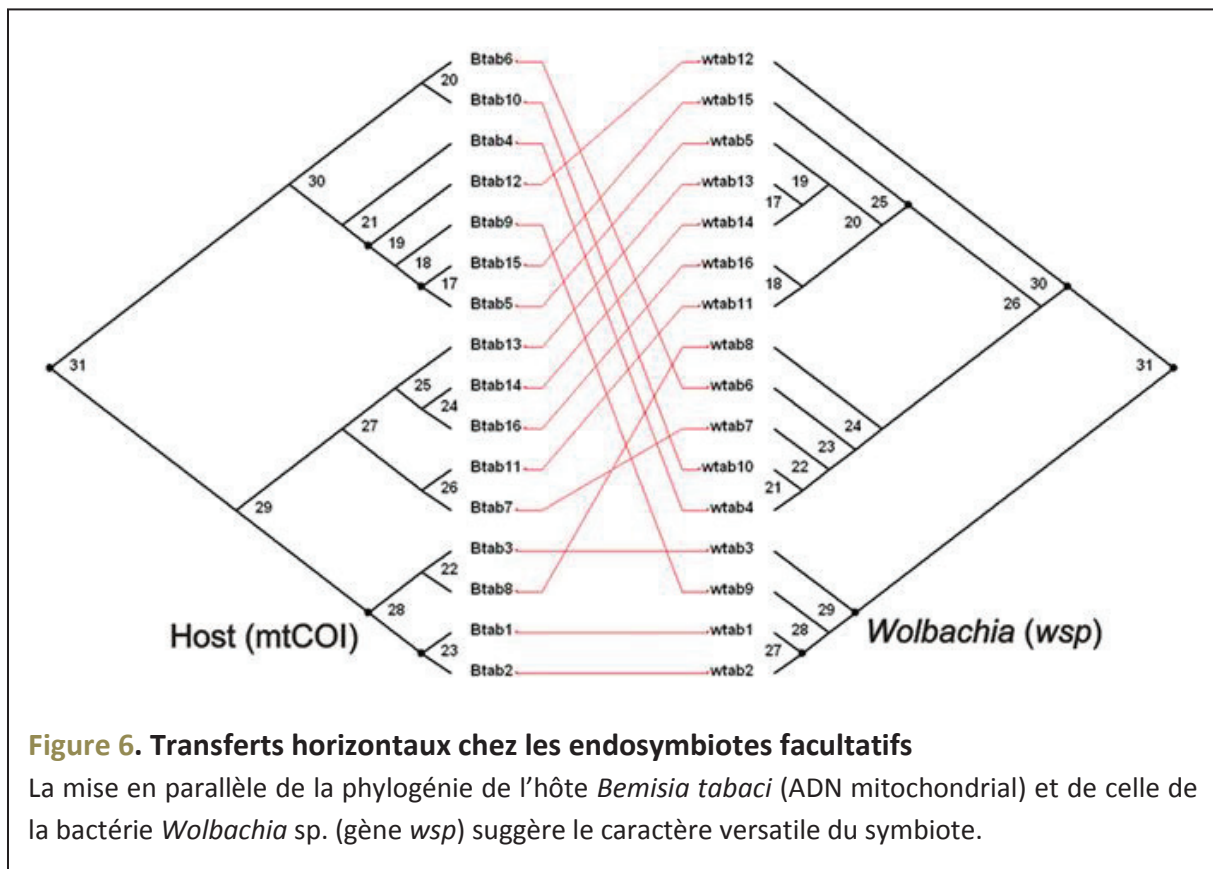
##### *Transmission verticale incomplète*

Comme les symbiotes primaires, ces endosymbiotes ont une transmission s'effectuant principalement de la mère à sa descendance. L'utilisation du terme « fugace », sans doute quelque peu excessive, a pour but de les différencier clairement des endosymbiotes primaires. Contrairement à ces derniers, une femelle infectée peut produire des descendants non-infectés, révélant une transmission verticale imparfaite de ces symbiotes (Chen et Purcell, 1997). Bien sûr, un endosymbiote primaire peut « accidentellement » ne pas être transmis à la descendance. Or, alors que ce dernier cas est immédiatement contre-sélectionné, les descendants dépourvus de symbiotes secondaires sont eux parfaitement viables. Aussi qualifie-t-on ces derniers de « facultatifs », par opposition à l'endosymbiote primaire « obligatoire ». Différents paramètres vont influencer sur cette capacité à transférer verticalement, notamment la densité symbiotique à l'intérieur de l'hôte, la localisation dans certains organes, les conditions environnementales etc. (Anbutsu *et al.*, 2008).

##### *Modalités de transfert horizontal*

Une seconde spécificité des symbiotes secondaires par rapport aux primaires, assez évidente à « détecter » par des analyses phylogénétiques, est qu'ils peuvent être transférés horizontalement, depuis leur hôte vers un conspécifique ou un hôte d'une autre espèce (**Figure 6**) (Baumann, 2005 ; Moran *et al.*, 2008). Du fait de ces transferts horizontaux et des pertes de symbiotes, la phylogénie des bactéries et celle des hôtes présentent une congruence limitée, voire totalement inexistante, de sorte qu'on ne retrouve pas chez les

endosymbiotes facultatifs le phénomène de co-cladogenèse unissant symbiotes primaires et insectes (Baumann, 2005 ; Moran *et al.*, 2008).



Ainsi, pour la quasi-totalité des bactéries que nous allons évoquer dans ce manuscrit, des souches de bactéries très proches peuvent infecter des hôtes appartenant aux différents clades d'insectes phloémophages, voire d'autres Ordres d'insectes (Mouton *et al.*, 2012 ; Telesnicki *et al.*, 2012). Les taux de transmission horizontale demeurent mal connus, mais les données suggèrent qu'ils ont lieu à des temps « évolutifs » plutôt qu'« écologiques » (Duron et Hurst, 2013), à des taux cependant très variables (Ferrari et Vavre, 2011).

Différentes hypothèses sont avancées quant aux « voies de transmission » horizontale préférentielles de ces bactéries. Certaines ont d'ailleurs été confirmées par l'expérimentation, et posent la question de la fréquence de ces transferts chez les espèces concernées.

L'accouplement et la promiscuité physique qui l'accompagne constitueraient de manière logique une source importante de transferts. Chez *A. pisum*, des femelles aposymbiotiques peuvent être infectées lors de l'accouplement, puis transmettre ces symbiotes à leur descendance (Moran et Dunbar, 2006). Il y a donc transfert horizontal depuis le mâle vers la femelle, suivi d'un transfert vertical jusque la descendance de cette dernière. Certains symbiotes présentent d'importantes populations dans l'appareil génital mâle, de sorte que cette composante paternelle de la transmission pourrait être importante chez *A. pisum*. Cette transmission sexuelle pourrait également être interspécifique, certains mâles n'étant pas très regardant quant à l'identité de leur partenaire (Moran et Dunbar, 2006). Ces cas de transmissions sexuelles ne sont pas restreints aux pucerons. Chez les aleurodes, de tels transferts pourraient avoir lieu entre chacun des deux sexes pendant l'accouplement (Brumin *et al.*, 2012). Chez les psylles, seul le mâle semble pouvoir infecter la femelle (Mann *et al.*, 2011).

Le fait de partager une même niche peut favoriser les transferts entre les phloémophages, conspécifiques ou non. Chez les Aleyrodidae, *Rickettsia* pourrait se retrouver hors de l'hôte, coloniser le phloème des plantes et ainsi réinitialiser l'infection avec de nouveaux hôtes (Caspi-Fluger *et al.*, 2011). Ce transfert depuis des surfaces contaminées pourrait permettre un transfert des symbiotes vers des hôtes très divers (Frago *et al.*, 2012). Il est envisageable que ce transfert ne soit pas « passif » : le symbiote pourrait utiliser la plante comme réservoir et s'y répliquer, augmentant ainsi sa capacité à se propager à un nouvel hôte, comme c'est le cas de certaines bactéries phytopathogènes transportées par les insectes (Frago *et al.*, 2012). Les sécrétions des Hémiptères pourraient également favoriser la transmission horizontale des symbiotes. On a retrouvé *Hamiltonella defensa* dans le miellat des insectes ainsi que dans la cire sécrétée par les cornicules (des tubes présents sur le dernier segment dorsal chez le puceron), qui est un liquide de défense de l'insecte (Darby et Douglas, 2003), ces bactéries se trouvant donc en contact direct avec les plantes et ceux qui s'en nourrissent.

Enfin, les transferts horizontaux peuvent avoir lieu entre différents niveaux trophiques, et notamment entre les Hémiptères et leurs ennemis naturels, prédateurs et parasitoïdes. Le genre *Cardinium* comprend des souches très proches, telles associées avec

la cochenille *Aspidiotus nerii*, telles avec les guêpes du genre *Aphytis* qui parasitent cet Hémiptère (Weeks *et al.*, 2003). Les bactéries du genre *Spiroplasma* infectent aussi bien les pucerons que leur terrible prédateur tant adulé par les jardiniers, la coccinelle (Fukatsu *et al.*, 2001). Une telle hypothèse a été validée expérimentalement en ce qui concerne les bactéries *Hamiltonella defensa* et *Regiella insecticola*, toutes deux transmises au puceron *Aphis fabae* depuis divers ennemis : des acariens de la famille des Trombididae et des parasitoïdes appartenant aux espèces *Aphidius colemani* et *Lysiphlebus fabarum* (Gehrer et Vorburger, 2012). Les spectres d'hôtes de ces différents prédateurs peuvent être larges, ce qui favorise les transmissions horizontales interspécifiques. Les sécrétions dont nous avons déjà parlé pourraient jouer un rôle dans ces échanges entre prédateurs et proies. Par exemple, la cire de cornicule de *Rhopalosiphum padi* peut jouer le rôle de kairomones chez la guêpe Aphidiidae *Lysiphlebus testaceipes* (Grasswitz et Paine, 1992).

#### **b) Tropisme tissulaire des endosymbiotes facultatifs**

De manière intéressante, les symbiotes secondaires présentent une grande diversité dans le spectre de tissus qu'ils colonisent, contrairement aux endosymbiotes primaires et leur localisation restreinte généralement aux seuls bactériocytes. Ce tropisme tissulaire est directement relié à leur mode de transmission.

Chez les Hémiptères, on retrouve de nombreux endosymbiotes secondaires dans les cellules du bactériome. Toutefois, le type cellulaire infecté dépend du clade des insectes. Chez le puceron, ils infectent généralement des bactériocytes différents de ceux du symbiote primaire, que l'on appelle bactériocytes secondaires, ou l'enveloppe cellulaire du bactériome (Tsuchida *et al.*, 2005 ; Moran *et al.*, 2005a). Chez les psylles et certaines cochenilles, les symbiotes secondaires sont également séparés des symbiotes primaires, et sont contenus dans les cellules syncytiales du bactériome (Subandiyah *et al.*, 2000 ; Thao *et al.*, 2000 ; Matsuura *et al.*, 2009). Chez les aleurodes en revanche, on peut retrouver l'ensemble des symbiotes secondaires dans les mêmes bactériocytes que le primaire (Goettlieb *et al.*, 2008). Certains endosymbiotes secondaires ne sont pas retrouvés dans les bactériocytes, et utilisent directement les systèmes reproducteurs des insectes. Chez les aleurodes par exemple, la présence de *Rickettsia* sp. dans le système reproducteur des femelles assurerait sa transmission verticale via les cellules folliculaires des oocytes (Brumin *et al.*, 2012). L'on

retrouve ainsi des symbiotes dans les appareils génitaux mâles – testicules, glandes accessoires etc. – des pucerons, des psylles et des aleurodes (Moran et Dunbar, 2006 ; Mann *et al.*, 2011 ; Caspi-Fluger *et al.*, 2011), ce qui est à mettre en lien avec la transmission paternelle supposée de ces symbiotes.

Le tropisme tissulaire des endosymbiotes peut également être mis en lien avec leur capacité à transférer horizontalement. Les glandes salivaires et le stylet de l'insecte, infectés par différentes bactéries, seraient par exemple en relation directe avec le milieu externe (Brumin *et al.*, 2012 ; Ammar *et al.*, 2011 ; Caspi-Fluger *et al.*, 2012). D'ailleurs, nous l'avons vu, certains endosymbiotes peuvent être retrouvés hors de leur hôte. Ainsi et comme l'indique leur nom, les phytopathogènes peuvent proliférer dans les plantes consommées par l'insecte (voir l'**Encadré 3** qui leur est consacré). *Rickettsia* sp. pourrait également utiliser la plante comme un réservoir (Caspi-Fluger *et al.*, 2012). On retrouve *H. defensa* dans le miellat, la cire des cornicules mais également sur les feuilles (Darby et Douglas, 2003). Cependant, à ma connaissance, on ne connaît pas leur capacité de survie dans ces environnements. Cela pose en tous les cas la question de la définition d'un endosymbiote, si celui-ci peut avoir un cycle de vie partiellement extracellulaire (**Encadré 3**).

#### **4. Diversité des associations**

Comme nous l'avons déjà évoqué, les symbiotes secondaires arrivent à se maintenir dans les populations malgré une transmission maternelle incomplète. Cela signifie que la valeur adaptative des compartiments cytoplasmiques infectés dépasse celle des compartiments cytoplasmiques non infectés (Vautrin et Vavre, 2009). De nombreux endosymbiotes des insectes sont en effet des partenaires mutualistes, tandis que d'autres manipulent la reproduction de leur hôte.



### Encadré 3. Phytopathogènes

Une catégorie d'organismes dits « phytopathogènes » infecte les Hémiptères. Ces bactéries ont chez l'insecte un mode de vie intracellulaire et ont une composante verticale dans leur transmission, de sorte qu'elles semblent remplir toutes les conditions requises pour être évoquées dans ce manuscrit (Hansen *et al.*, 2008). Leur particularité est d'être en outre transmises à une plante cible dans laquelle elles vont proliférer, colonisant les tissus dont va se nourrir l'insecte pour ainsi réinitialiser le cycle (Hansen *et al.*, 2008). Nous avons donc à faire à un cas très intéressant de symbiote facultatif qui diffère des mutualistes précédemment évoqués par un taux de transferts horizontaux *a priori* supérieur, ceux-ci ayant lieu de manière cyclique au cours de la saison – rien à voir donc avec les quelques transferts par millions d'années que laisse imaginer la phylogénie des autres endosymbiotes facultatifs.

L'hôte insecte est ici qualifié de « vecteur » par l'Homme, du fait des maladies que son endosymbiote transmet aux plantes cultivées par celui-ci. Par exemple, les Cixiidae (Auchenorrhyncha) hébergent des espèces phytopathogènes d'*Arsenophonus* telles que la bien-nommée *A. phytopathogenicus*, responsable du syndrome de « Basses richesses » décelé en Bourgogne dans les années 1990 (revue dans Bressan, 2014). Quant aux psylles, elles sont les vecteurs de bactéries du genre *Liberibacter* (Protéobactérie Alpha) : *Diaphorina citri* est le vecteur de *L. asiaticus* et *L. americanus*, responsables de la maladie du Dragon Jaune (黃龍病) chez les agrumes (do Carmo Teixeira *et al.*, 2005) ; *L. solanacearum* est transmis par *Bactericera cockerelli* (Šulc) et infecte le phloème des Solanaceae, chez qui il est responsable de maladies telle que la « chips zébrée » chez la pomme de terre (Hansen *et al.*, 2008).

Généralement, le phénotype des phytopathogènes pour l'insecte hôte demeure mal connu, l'Homme s'étant davantage intéressé aux conséquences agronomiques. La transmission verticale du symbiote devrait favoriser le mutualisme, de même que son mode de transmission horizontale particulier, puisque la bactérie sera contre-sélectionnée si elle atténue la capacité vectrice de l'insecte. Or, la présence de *L. solanacearum* est néfaste pour la fécondité de son hôte (Nachappa *et al.*, 2012 ; Nachappa *et al.*, 2014). Le bénéfice pour l'insecte pourrait toutefois être indirect, qui viendrait contrebalancer le coût à l'infection. Il a été démontré que *L. psyllauros* perturbait les systèmes de défense de la plante, ce qui pourrait permettre une meilleure exploitation de celle-ci par l'insecte (Casteel, 2012).

Nous avons vu que d'autres endosymbiotes *a priori* non phytopathogènes présentaient toutefois des similarités avec ceux-ci. Ainsi, *Rickettsia* sp. peut être transmise depuis l'aleurode à une plante de la même façon que les bactéries du genre *Liberibacter* (Caspi-Fluger *et al.*, 2012), bien que rien ne soit connu quant à son interaction avec l'hôte végétal. Il est possible que tout un continuum existe entre les deux catégories bien distinctes d'« endosymbiotes d'insectes » et de « bactéries phytopathogènes vectorisées par les insectes ». D'ailleurs, le genre *Arsenophonus* présente de nombreuses transitions entre endosymbiose stricte et organismes phytopathogènes (Bressan, 2014).

### **a) Manipulateurs de la reproduction**

Nous avons déjà abordé la notion de manipulation en amont dans ce manuscrit et n'allons pas nous étendre outre mesure. Chez les phloémophages, ce sont les cochenilles qui semblent les victimes les plus fréquentes de ces symbiotes. Ainsi, les transitions entre les nombreux systèmes de reproduction de ces insectes pourraient être le fait de manipulateurs de la reproduction (Normark, 2004) bien que cette hypothèse ne soit corroborée que par une analyse corrélative (Normark, 2004 ; Ross *et al.*, 2012). D'autres études ont démontré une association entre les haplotypes asexués du Diaspididae *Aspidiotus nerii* et la bactérie manipulatrice *Cardinium* (Provencher *et al.*, 2005).

Chez les pucerons, les aleurodes et les psylles, peu de cas de manipulations de la reproduction ont été mis en évidence. On retrouve deux Super-Groupes du genre *Wolbachia* endémiques du puceron, les clades M et N (Augustinos *et al.*, 2011), et un Super-Groupe endémique des aleurodes, le clade O, mais leurs phénotypes sont inconnus (Bing *et al.*, 2014). Chez *A. pisum*, les Firmicutes *Spiroplasma* pourraient être impliqués dans la mortalité importante des mâles du genre *A. pisum* (Simon *et al.*, 2011), suggérant un phénotype masculicide dans cet insecte. *R. insecticola* pourrait influencer le mode de reproduction du puceron, son élimination engendrant une diminution des morphes sexués (Leonardo et Mondor, 2006).

### **b) Mutualistes**

En revanche, de nombreuses associations mutualistes ont été décrites entre les Hémiptères et les endosymbiotes secondaires. Ces symbiotes peuvent favoriser la résistance de l'hôlobionte à différents stress abiotiques : la chaleur par exemple peut être extrêmement dommageable pour l'hôte du fait de son ectothermie, mais peut également nuire au symbiote primaire (Ohtaka et Ishikawa, 1991). La présence de symbiotes facultatifs peut en atténuer les effets : *S. symbiotica* et *H. defensa* protégeraient le puceron *A. pisum* contre les stress thermiques (Oliver *et al.*, 2003 ; Russel et Moran, 2006), de même que *S. symbiotica* qui préserve de la destruction les bactériocytes soumis à de fortes chaleurs (Montllor *et al.*, 2002). Les symbiotes pourraient également intervenir dans la détoxification et la résistance aux insecticides (Kikuchi *et al.*, 2012).

Le stress peut également se présenter sous la forme d'un ennemi naturel, tels que des champignons, des virus ou des guêpes parasitoïdes. Tout ennemi de l'hôte affectant la valeur adaptative des éléments cytoplasmiques que ce dernier héberge, des stratégies de défense peuvent émerger chez les endosymbiotes. L'exemple le plus fameux concerne une souche manipulatrice du genre *Wolbachia*, wMel, qui protège la drosophile contre différents virus (Teixeira *et al.*, 2008). Ce faisant, cette bactérie confère un double avantage aux compartiments cytoplasmiques infectés, combinant action protectrice et manipulation de la reproduction. La protection bactérienne va m'intéresser tout particulièrement dans le cadre de ce manuscrit, puisqu'elle fait partie de la « panoplie » phénotypique de l'organisme qui m'a servi de modèle, *Hamiltonella defensa*. Ainsi, plusieurs Protéobactéries Gamma (*H. defensa* donc, mais aussi *S. symbiotica* et *R. insecticola*) protègent le puceron *A. pisum* contre la guêpe Braconidae *Aphidius ervi* (Oliver *et al.*, 2003). *R. insecticola* protège également son hôte contre les pathogènes fongiques, et jugule la contamination en réduisant la prolifération des champignons sur les cadavres de pucerons (Scarborough *et al.*, 2005). Certaines souches de *Rickettsia* sp. protégeraient elles-aussi *A. pisum* contre les champignons (Lukasik *et al.*, 2013). *Spiroplasma* pourrait coupler résistance et manipulation du comportement de l'hôte, écartant les pucerons mourants de leurs congénères dans une stratégie là encore de jugulation de la contamination (Lukasik *et al.*, 2013). Un cas intéressant de protection concerne des pucerons de l'espèce *A. pisum* normalement teintés d'un rouge criard que les coccinelles détectent facilement. Chez ces insectes, une Protéobactérie Gamma *Rickettsiella viridis* peut leur conférer une couleur verte, leur permettant ainsi de se fondre dans le feuillage (Tsuchida *et al.*, 2010).

Les symbiotes peuvent également permettre à un hôte d'infecter des plantes inaccessibles aux individus non infectés, tel le trèfle, que la présence de *R. insecticola* rend comestible pour le puceron (Tsuchida *et al.*, 2004). Les endosymbiotes peuvent donc améliorer la qualité nutritive globale du végétal consommé, en en atténuant les systèmes de défense, par exemple ceux sous la régulation des acides jasmoniques et salicyliques (Casteel, 2012). Ils peuvent enfin induire une réponse agressive de la plante : si cette réponse désavantage les compétiteurs de l'hôte mais pas ce dernier, cela lui confère un avantage. On trouve ce genre de relation mutualiste non pas chez les Hémiptères, mais entre le Coléoptère *Dendroctonus ponderosae* et un champignon (Colgan et Erbilgin, 2011).

### c) Mutualisme et conditions environnementales

Lorsque l'on parcourt la bibliographie, on se trouve confronté à un paradoxe intéressant : de nombreux endosymbiotes catégorisés en tant que « mutualistes » se sont révélés néfastes pour l'hôte dans de nombreuses expériences, affectant notamment sa durée de vie ou sa fécondité. C'est le cas de *H. defensa* (Vorburger et Gouskov, 2011), *R. insecticola* (Chen *et al.*, 2000 ; Russel et Moran, 2006), *S. symbiotica*, *Rickettsia* sp. (Chen *et al.*, 2000) etc. On a grâce à ces expériences un aperçu très clair des coûts physiologiques générés par les symbiotes. Cela signifie surtout que dans ces expériences en laboratoire, les conditions rencontrées par l'insecte ne permettaient pas au symbiote d'apporter un bénéfice suffisant, qui aurait permis de compenser ces coûts. Il en découle de manière intuitive que la relation sera « mutualiste » uniquement si les bénéfices retirés par l'hôte sont supérieurs aux coûts de l'infection.

Cela est parfaitement illustré par les bactéries conférant une protection, dont les bénéfices apportés dépendent des pressions exercées par les ennemis naturels de l'hôte. Par exemple, la psylle *Glycaspis brimblecombei* présente une corrélation entre le taux d'infection des populations naturelles par des symbiotes secondaires présumés protecteurs d'une part, et la densité en *Psyllaphaegus bliteus*, une guêpe parasitoïde solitaire, d'autre part (Hansen *et al.*, 2007). De même, en utilisant des cages à populations, le taux d'infection d'*H. defensa* décline très rapidement en absence de parasitoïdes et finit par disparaître, trahissant des coûts importants pour l'hôte (Oliver *et al.*, 2008 ; Dykstra *et al.*, 2014). La pression peut provenir de plusieurs ennemis naturels. Ainsi que nous l'avons déjà vu, *R. viridis* protégerait son hôte contre les coccinelles en participant à son camouflage sur le feuillage. Toutefois, les pucerons verts attirent davantage certains parasitoïdes (Tsuchida *et al.*, 2010). En conséquence, selon que la pression en l'un ou l'autre de ces deux ennemis naturels s'intensifiera, la teinte fournie par *R. viridis* pourra entraîner des coûts autres que physiologiques, attirant directement le parasitoïde jusqu'à l'hôte.

Enfin, les colonies d'insectes en laboratoire sont maintenues sur certains types de plantes et à certaines conditions de température. Il en va de même pour les expériences réalisées. Cela n'est pas sans conséquence sur le phénotype de leur symbiote, et donc sur les résultats des analyses, si l'on considère le rôle que jouent les endosymbiotes dans

l'utilisation des végétaux et dans la protection contre des stress thermiques. Par exemple, la plante hôte peut moduler le bénéfice retiré par la présence d'un endosymbiote : *R. insecticola* n'a pas les mêmes conséquences sur les traits d'histoire de vie de l'insecte selon la plante consommée, et influencerait le spectre de plantes hôtes de celui-ci (Tsuchida *et al.*, 2004). De même, les facteurs abiotiques, et parmi eux ce paramètre primordial qu'est la température, influencent énormément la nature de l'association. Par exemple, la résistance aux parasitoïdes conférée par *H. defensa* ne sera pas la même selon les températures utilisées lors des différentes expériences (Dykstra *et al.*, 2014). Il est donc extrêmement difficile de comprendre la relation entre les insectes et les endosymbiotes sur la base de quelques expériences, et il y a fort à parier que de nombreux bénéfices apportés par ces bactéries restent à découvrir, par exemple parce qu'ils s'exercent dans des conditions qui n'ont pas encore été testées.

#### **d) Interactions génotypiques**

Une seconde explication quant à cette hétérogénéité dans les résultats réside dans la variabilité génotypique qui existe. Ainsi, certaines études récentes essaient de distinguer la part relative des fonds génétiques des symbiotes et des eucaryotes dans les phénotypes observés, ainsi que de leur interaction (lire la revue de Feldhaar, 2011, qui leur fait la part belle).

Tout d'abord, différentes souches de bactéries apparentées peuvent conférer un bénéfice plus ou moins important, voire de nature différente à un même hôte. Par exemple, pour une même souche d'insecte hôte, différentes bactéries apparentées de l'espèce *H. defensa* ne présentent pas la même efficacité dans leur protection contre les parasitoïdes (Oliver *et al.*, 2005). Ensuite, différents hôtes ne répondent pas de la même façon à la présence de symbiotes. Là encore, une même souche d'*H. defensa* n'apporte pas les mêmes bénéfices à des hôtes possédant différents génotypes. De la même façon, *R. insecticola* n'a pas les mêmes conséquences sur la valeur adaptative des différentes souches d'*A. pisum*, ainsi que sur leur spécialisation à certaines plantes (Chen *et al.*, 2000 ; Tsuchida *et al.*, 2004 ; Leonardo, 2004). Un dernier exemple particulièrement intéressant concerne une fois encore le verdissement du puceron par *R. viridis* : toutes les souches vertes de pucerons n'hébergent pas la bactérie, tandis que certaines souches sont infectées mais demeurent

rouges (Tsuchida *et al.*, 2010). En outre, la présence de la bactérie peut influencer ou non la taille et le taux de développement de l'hôte en fonction du bagage génétique de ce dernier (Tsuchida *et al.*, 2010).

Enfin, il existe une interaction entre les génotypes des hôtes et des endosymbiotes, l'environnement pouvant influencer celle-ci, autant qu'elle peut influencer les génotypes individuels des partenaires. Cette interaction a été moins étudiée en ce qui concerne les associations insectes phloémophages/endosymbiotes. Simon et ses collaborateurs (2011) ont étudié chez le puceron l'impact du fond génétique de l'hôte et de la présence de symbiotes facultatifs sur différents traits d'histoire de vie. Des différences significatives ont ainsi été mises en évidence en fonction du génotype de l'insecte, de la présence de symbiotes, et de l'interaction entre les deux, pour l'**ensemble** des traits étudiés (à savoir, l'âge lors de la première reproduction, la fécondité, la durée de vie et la production des différents morphes).

L'« équation » permettant de désigner le phénotype d'un symbiote pour son hôte (**Figure 7**) peut se complexifier davantage encore si l'on prend la variabilité génotypique existant chez les autres partenaires (plantes, parasitoïdes etc.) mais nous ne nous appesantirons pas sur la question.

$$\text{Phénotype} = (\text{Génotype}_{(h)} + \text{Génotype}_{(es)} + \text{Génotype}_{(h)} \times \text{Génotype}_{(es)}) \times \text{Environnement}$$

**Figure 7.** Le phénotype d'un endosymbiote (es), et donc la balance des coûts et bénéfices associés à sa présence pour l'hôte (h), est sous la dépendance du génotype de chacun des deux partenaires, de leur interaction, mais également de l'environnement, qui va agir sur chacune des composantes.

## 5. Interactions entre différentes espèces de symbiotes au sein d'un même hôte

Cela peut également se complexifier si l'on considère le fait qu'un insecte n'est pas nécessairement infecté par une seule espèce d'endosymbiote. Ces multi-infections sont fréquentes chez nos insectes phloémophages, et nous les avons évoquées sans nous en rendre compte : puisque la quasi-totalité des insectes que nous présentons possède un symbiote obligatoire, toute évocation d'un endosymbiote secondaire implique une bi-infection. En outre, certains insectes sont infectés par plus d'un symbiote facultatif : chez *Diaphorina citri*, 10 individus sur les 42 testés par Subandiyah et ses collaborateurs (2000) étaient infectés par cinq espèces endosymbiotiques distinctes, dont quatre *a priori* secondaires. Les multi-infections sont surtout renseignées chez l'aleurode *Bemisia tabaci*, qui présente des taux de plus de 65% de multi-infections par des endosymbiotes secondaires (Gueguen *et al.*, 2010). Quelles conséquences vont-elles avoir pour l'hôte et son association avec les différents partenaires ?

### a) Maintien des multi-infections

Les pathogènes à transmission horizontale ont cristallisé la quasi-totalité des études théoriques et empiriques consacrées aux multi-infections, la prédiction la plus courante étant une virulence accrue de ces espèces qui va leur permettre de sortir « vainqueur » de la compétition interspécifique et intra-hôte (Ebert et Bull, 2003). L'on sait également qu'un endosymbiote à transmission maternelle a des intérêts divergents d'avec les pathogènes, comme c'est le cas entre wMel et les virus contre lesquels elle protège l'hôte. Nous allons pour notre part nous intéresser à la multi-infection entre endosymbiotes à transmission verticale, et notamment aux conditions permettant leur maintien (revue dans Vautrin et Vavre, 2009).

En suivant le même raisonnement que précédemment, si ces multi-infections se maintiennent dans la population malgré la transmission incomplète de ses membres, cela implique que la production de femelles multi-infectées soit supérieure à la production de femelles mono ou non-infectées (Vautrin et Vavre, 2009). Là encore, les stratégies de manipulation de la reproduction et de mutualisme vont être au cœur de ce maintien. La différence par rapport aux mono-infections va résider dans la complémentarité des phénotypes. Si deux symbiotes présents dans un insecte apportent un bénéfice de même

nature (synthèse d'un élément quelconque), ou ont les mêmes conséquences pour l'hôte (féminisation) la multi-infection ne devrait pas être maintenue : en effet, le coût infligé à l'hôte par les différents symbiotes ne sera compensé par aucun bénéfice supplémentaire (Vautrin et Vavre, 2009).

Une multi-infection stable implique donc la coexistence de symbiotes aux bénéfices distincts, voire synergiques. Conséquemment, le maintien des endosymbiotes secondaires passe par une stratégie obligatoirement différente de celle du symbiote primaire : résistance thermique, protection, etc. C'est ainsi qu'*H. defensa*, *R. insecticola* ou *S. symbiotica* peuvent cohabiter avec le symbiote primaire du puceron. Un manipulateur de la reproduction et un endosymbiote mutualiste sont fréquemment retrouvés en multi-infection, comme c'est le cas chez la mouche tsétsé où l'on retrouve des bactéries du genre *Wolbachia* et des mutualistes du genre *Sodalis* (Wang *et al.*, 2013). Un cas particulier mérite d'être cité : différentes bactéries manipulatrices induisant de l'incompatibilité cytoplasmique – et donc le même effet pour leur hôte – peuvent être maintenues en co-infection (Mouton *et al.*, 2003). En effet, dans ce cas, le compartiment maternel pluri-infecté est le seul qui puisse produire une descendance viable avec l'ensemble des mâles de la population (mono-, bi- et non-infectés).

#### **b) Tragédie des biens communs**

La multi-infection entraîne une interaction forte entre les endosymbiotes, qui peut altérer la dynamique des populations de chacune des espèces (Vautrin et Vavre, 2009). En effet, l'hôte représente une ressource limitée que vont se disputer les endosymbiotes. Cette compétition peut rendre impossible le maintien de la multi-infection, en particulier lorsque sont affectées (i) la transmission maternelle d'au moins une des espèces, ou (ii) la valeur adaptative de l'hôte. Il y a peu d'éléments pour discuter du premier point. L'on sait toutefois que certains symbiotes pourraient en exclure d'autres, et ainsi empêcher leur co-transmission : hébergées dans les mêmes cultures cellulaires que *R. insecticola*, *H. defensa* est systématiquement éliminée par celle-ci (Darby *et al.*, 2005). Concernant le second point en revanche, différents exemples suggèrent que la présence simultanée de plusieurs espèces d'endosymbiotes peut s'avérer dommageable pour l'hôte. *In vitro*, il a été démontré que la longévité et la fécondité des pucerons hébergeant *H. defensa* et *R. insecticola* étaient



tellement affectées qu'on ne s'attend pas à retrouver les deux symbiotes dans le même individu dans les populations naturelles – et ce malgré une protection accrue des hôtes bi-infectés (Oliver *et al.*, 2006). En outre, la valeur adaptative de l'hôte peut être affectée si l'interaction entre les deux symbiotes nuit à leur phénotype individuel : *S. symbiotica* ne protège plus le puceron contre d'éventuels stress thermiques en cas de présence de *Rickettsia* sp. (Montllor *et al.*, 2002). À l'inverse, la présence d'*H. defensa* n'empêche pas le phénotype protecteur de *Rickettsia* sp. et *Spiroplasma* sp. (Lukasik *et al.*, 2013).

Ainsi, les multi-infections requièrent des conditions très particulières pour être maintenues. Pourtant, ces associations peuvent être très stables dans les populations naturelles, ce qui pose la question de leur régulation et de leur évolution (Vautrin et Vavre, 2009).

### **c) Régulations et évolution des multi-infections**

La multi-infection va générer des interactions fortes entre les différents protagonistes, au détriment parfois de l'ensemble des partenaires. Comment peut-elle évoluer au cours des générations d'holobiontes ? Si une multi-infection se maintient, et qu'il y a donc un avantage à être multi-infecté, la sélection va favoriser les holobiontes chez qui la transmission conjointe de l'ensemble des partenaires va être la plus efficace. L'on peut s'attendre alors à ce que des stratégies de régulation de la co-transmission émergent, tant de la part des symbiotes que de l'hôte.

Par exemple, pour éviter la compétition, une partition des niches à l'intérieur de l'insecte peut sembler efficace. Toutefois, elle n'empêche ni le surcoût lié à la présence de plusieurs espèces dans l'hôte, ni la compétition au moment de la transmission verticale dans les oocytes (Vautrin et Vavre, 2009), qui est maximale si l'on considère le goulot d'étranglement que celui-ci représente.

Une régulation de la densité des différents endosymbiotes pourrait permettre leur coexistence. *W. pipientis*, dont les prévalences sont élevées chez *Coelostomidia wairoensis*, voit sa densité à l'intérieur de l'hôte décroître lorsqu'un second symbiote du genre *Erwinia* est également hébergé par l'insecte (Dhami *et al.*, 2012). Lors d'une infection par *S.*

*symbiotica*, *B. aphidicola* voit ses effectifs chuter drastiquement (Koga *et al.*, 2003). La régulation de ces densités, autant sous la direction des symbiotes que de l'hôte, est cependant mal connue (Vautrin et Vavre, 2009). Rien n'indique que ces processus soient adaptatifs, et ils pourraient être simplement le reflet des compétitions qui s'exercent entre les symbiotes, qui sont présumées avoir des conséquences énormes sur la dynamique populationnelle des symbiotes (Vautrin et Vavre, 2009).

Enfin, certaines symbioses très particulières pourraient illustrer cette sélection des holobiontes maximisant la co-transmission des partenaires. Les cas les plus extrêmes concernent les endosymbiotes hébergés par d'autres endosymbiotes, tels *Moranella endobia* ou les bactéries du genre *Midichloria*, hébergés respectivement par *Tremblaya princeps* chez les cochenilles farineuses, et par les mitochondries chez les tiques *Ixodes ricinus* (Beninati *et al.*, 2004). Dans ces systèmes, les populations des symbiotes sont imbriquées, prévenant tout risque d'exclusion de l'un ou l'autre des deux partenaires.

## **6. Un système simplifié bien complexe...**

En conclusion, nous avons pris pour modèle un holobionte qui apparaît comme « simple », en tous les cas relativement au tube digestif des Vertébrés. Malgré le nombre limité de partenaires, le système n'est cependant pas sans complexité. La balance des coûts-bénéfices du symbiote pour son hôte, mouvante au gré des conditions environnementales et des autres partenaires symbiotiques, est également sous la dépendance des interactions avec le génotype de l'insecte qu'il va infecter. De plus, la présence d'autres espèces peut entraîner une compétition pour les ressources entre les symbiotes, mais également favoriser l'holobionte si les bénéfices prodigués sont complémentaires.

### III. Evolution génomique des endosymbiotes

Dans cette troisième partie, la transition entre bactéries libres et endosymbiotes nous intéressera tout particulièrement, ainsi que ses conséquences, la plus significative étant la signature caractéristique des génomes des endosymbiotes. Nous consacrerons les derniers paragraphes aux nuances qui existent entre ces endosymbiotes, notamment entre primaires et secondaires.

#### 1. Symbiogenèse

Sachs et ses collaborateurs (2011) ont étudié 42 transitions vers un mode de vie symbiotique. L'ensemble des clades associés à un hôte aurait émergé depuis des bactéries environnementales, ce qui n'est pas surprenant si l'on considère le milliard d'années qui a séparé l'apparition des bactéries et celle des eucaryotes (Sachs *et al.*, 2011). Ainsi, les différents symbiotes que nous avons évoqués dans ce manuscrit sont-ils – lointainement – apparentés à des bactéries non symbiotiques, ou parfois avec des bactéries ayant une écologie comportant des phases libres et des phases en association avec un hôte (Figure). *S. symbiotica* appartient à un genre de bactéries ubiquistes (Petersen et Tisa, 2013). *H. defensa* et *R. insecticola* ont pour cousines les bactéries du genre *Yersinia*, pathogènes des mammifères (Degnan *et al.*, 2009). Les genres *Wolbachia* et *Rickettsia*, ainsi que la mitochondrie, sont des lointains parents des Rhizobiales (Moran *et al.*, 2008), ce clade comportant des espèces symbiotiques mais également des espèces vivant uniquement dans le sol.

Il est possible que certaines de ces bactéries libres ancestrales aient, de par leurs propriétés écologiques et/ou génomiques, été prédisposées à établir des associations avec un hôte. Pour preuve, certains clades n'ont jamais transité vers des formes symbiotiques, tandis que d'autres ont connu de multiples transitions (Sachs *et al.*, 2011). Tel est le cas des Protéobactéries, et notamment les Sous-Divisions Alpha et Gamma, qui comprennent la quasi-totalité des bactéries que nous avons évoquées dans le manuscrit, et dont le dernier ancêtre commun était une bactérie libre. Cependant, si l'on pondère par la longueur des branches de l'arbre phylogénétique de chaque clade, la notion de clades prédisposés à la symbiose semble peu pertinente aux yeux de Sachs et de ses collaborateurs (2011). À noter

toutefois que leur étude ne prend pas en compte l'ensemble de la diversité connue : les Protéobactéries Gamma des genres *Hamiltonella*, *Regiella*, *Sodalis* ou *Serratia*, pour ne citer qu'elles, n'y figurent pas.

La transition vers un mode de vie symbiotique serait basée sur des transferts horizontaux de gènes particuliers, couplés à une écologie favorisant la probabilité de rencontre avec des hôtes potentiels (Carvalho *et al.*, 2010 ; Toft et Andersson, 2010 ; Sachs *et al.*, 2011). Le scénario, dans ce cas, serait un passage direct depuis une forme libre vers une symbiose mutualiste. Pour de nombreux symbiotes dont les symbiotes intracellulaires, le scénario pourrait être une forme d'abord parasitaire, qui deviendrait mutualiste du fait d'un passage vers une transmission verticale (Sachs *et al.*, 2011). À noter toutefois que les nombreux manipulateurs, nous l'avons vu, sont des parasites à transmission verticale.

## **2. Évolution des génomes endosymbiotes**

Par rapport à leurs ancêtres pré-endosymbiotiques, deux aspects fondamentaux caractérisent l'écologie des endosymbiotes : la stabilité de leur environnement et la part importante de la dérive dans leur évolution, notamment lors des transferts de la femelle insecte à sa descendance. Ces deux paramètres vont énormément influencer sur le devenir de leurs génomes.

### **a) Écologie des endosymbiotes**

Les bactéries libres affrontent sans cesse la stochasticité de leur environnement. Leur niche écologique est en réalité composée d'une multitude de micro-niches présentant chacune des conditions très particulières (Moran *et al.*, 2005b). À cette hétérogénéité s'ajoute une dynamique temporelle, leur environnement étant parfaitement instable et imprévisible. Pour répondre à ces variations, elles disposent de tout un arsenal de facteurs de virulence, de senseurs, de régulateurs, voire de système de détection du quorum pour induire une action coopératrice. De plus, les populations de bactéries libres présentent des ordres de grandeur importants et possèdent différentes formes de sexualité. Du fait de ces deux propriétés, on estime à un milliard d'individus la taille efficace des populations d'*Escherichia coli* (Selander *et al.*, 1987). Ce chiffre impressionnant permet la prise très forte de la sélection chez ces bactéries sans cesse en compétition.

Par opposition à ces bactéries libres, les endosymbiotes vivent dans le milieu **riche** et **stable** que constitue la cellule eucaryote (Shigenobu *et al.*, 2000). Ils sont donc préservés des conditions de vies mouvantes auxquelles font face les bactéries libres. Cette stabilité va permettre un relâchement des pressions de sélection sur les fonctions de perception des stimuli, ainsi que sur les systèmes de régulation (Shigenobu *et al.*, 2000 ; Moran *et al.*, 2008). La transmission verticale joue également un grand rôle dans leur évolution : facilitant l'invasion de l'hôte, elle dispense là-encore le symbiote de tous les systèmes de virulence rencontrés chez les organismes pathogènes. Ainsi, de nombreux gènes sont totalement dispensables chez les endosymbiotes (Shigenobu *et al.*, 2000 ; Moran *et al.*, 2008).

Relativement aux bactéries « libres », les endosymbiotes présentent en outre des populations de très petite taille. On chiffre la présence de *B. aphidicola* à quelques dizaines de milliers de cellules chez le puceron *A. pisum* (Mira et Moran, 2002 ; **Tableau 3B**). De plus, les phases de transmission verticale s'apparentent à de véritables « goulots d'étranglement » : une fraction seulement de la population bactérienne est transmise à la descendance de l'hôte, ce qui a pour conséquence d'augmenter encore les effets de la dérive génétique (Moran, 1996). Pour différentes espèces symbiotiques, on estime à seulement quelques milliers, voire centaines, le nombre de bactéries transmises à la descendance chez le puceron (Mira et Moran, 2002).

Espèce hôte	Localisation	Nombre de symbiotes
<b>(A) Mesure du goulot d'étranglement</b>		
<i>Nasonia sp.</i>	Ovocyte	886 ± 60 (n = 4)
<i>Acyrtosiphon pisum</i>	Ovocyte	1872 ± 524 (n = 5)
	Bacteriocyte (embryon)	844 ± 48 (n = 4)
	Bacteriocyte (larve 1 <sup>er</sup> stade)	3234 ± 410 (n = 8)
<i>Uroleucon ambrosiae</i>	Ovocyte	8223 ± 428 (n = 4)
	Bacteriocyte (embryon)	821 ± 71 (n = 7)
<b>(B) Mesure à l'échelle de l'hôte</b>		
<i>Acyrtosiphon pisum</i>	Embryon	36 700
<i>Acyrtosiphon pisum</i>	Nymphe de 1 <sup>er</sup> stade	119 490

**Tableau 3.** Nombre de symbiotes *Buchnera* dans (A) les ovocytes et les bactériocytes et (B) dans les hôtes entiers

## b) Les premiers stades de l'endosymbiose

Nous allons dans un premier temps nous focaliser sur les premiers stades de l'endosymbiose. En effet, les conditions extrêmes rencontrées par les bactéries nouvellement acquises à la vie intracellulaire, que nous appellerons par souci de simplification des **néo-endosymbiotes**, vont avoir des répercussions drastiques sur leurs génomes.

Du fait d'une évolution régie par la dérive plutôt que par la sélection, les mutations faiblement délétères ne seront pas nécessairement contre-sélectionnées et s'accumulent de manière irréversible, selon le processus du « cliquet de Muller » (Moran, 1996). Cela est d'autant plus vrai que certaines mutations vont inactiver des gènes permettant la recombinaison, et qui auraient pu compenser l'accumulation des mutations délétères par l'acquisition de nouvelles fonctions (Silva *et al.*, 2003). De plus, de nombreux gènes sont dispensables dans le milieu intracellulaire et peuvent être perdus sans conséquences néfastes pour les symbiotes. Il y a donc une érosion rapide du génome chez les endosymbiotes, dont des régions entières seraient parfois perdues en un seul événement (Delmotte *et al.*, 2006). En outre, il existe dans ces mutations des biais dits « universels » (c'est-à-dire qu'on les retrouve aussi bien chez des bactéries libres que chez des bactéries symbiotiques), et notamment un remplacement des bases G et C par des bases A et T. Cela conduit chez les endosymbiotes à un enrichissement progressif du contenu en adénine et en thymine (Baumann *et al.*, 2005 ; Moran *et al.*, 2008). Cela n'est pas sans conséquence sur le processus de pseudogénéisation : les codons d'arrêt étant composés d'au moins deux bases A et T (**TAA**, **TAG**, **TGA**), l'accumulation de ces bases va accroître la probabilité d'une mutation non-sens. De même, la présence de pseudogènes et de régions non codantes couplée à une sélection purifiante peu efficace permettent une prolifération des éléments mobiles à l'intérieur du génome des endosymbiotes (Baumann *et al.*, 2005 ; Moran *et al.*, 2008).

Ainsi, les néo-endosymbiotes possèdent des génomes de taille comparable à celles des bactéries libres, mais contenant une grande proportion de gènes inactivés (pseudogènes) et d'éléments mobiles (Oakeson *et al.*, 2014; **Tableau 4**). Par exemple, le néo-endosymbiote *Sodalis pierantonius* souche SOPE – seulement 2% de divergence synonyme

avec les bactéries libres apparentées (Clayton *et al.*, 2012) – a rapidement perdu 50% de sa fonctionnalité initiale, et les séquences d’insertion représentent 18% des quatre mégabases que constitue son génome (Clayton *et al.*, 2012 ; **Tableau 4**). L’activité de ces éléments mobiles provoque généralement un remaniement total de l’organisation des gènes dès les premiers stades de la symbiose (Oakeson *et al.*, 2014). La présence de pseudogènes en serait également pour partie responsable : plus la proportion d’ADN non codant est importante dans un organisme, plus la probabilité qu’un réarrangement ait des conséquences létales diminue (Moran et Plague, 2004 ; Plague *et al.*, 2008). Chez *Sodalis pierantonius* SOPE cependant, les éléments mobiles se cantonneraient dans le non-codant et ne participeraient pas au processus d’inactivations de gènes (Oakeson *et al.*, 2014). Autre particularité de ce symbiote, son système de recombinaison est intact, ce qui suggère que ces systèmes sont soumis à sélection chez les néo-endosymbiotes.

Caractéristiques	<i>Buchnera aphidicola</i> BCc	<i>Serratia symbiotica</i> SCc	<i>Hamiltonella defensa</i> 5AT	<i>Sodalis pierantonius</i> SOPE	<i>Serratia proteamaculans</i>
Phénotype	Symbiote primaire	Symbiote néo-primaire*	Symbiote secondaire	Symbiote néo-primaire**	Bactérie libre
Taille du génome (Mb)	0,4	2	2,2	4,5	5,5
Nombre de gènes	401	711	2420	2309	5064
Nombre de pseudogènes	3	58	187	1771	12
Longueur moyenne des gènes (bases)	994	1020	810	Inconnu	972
Densité régions codantes (%)	90	38,7	88,8	Inconnu	87,1
Bases G et C (%)	20	29	40	56	55
Présence d’éléments mobiles	Non	Non	Oui	Oui	Oui

**Tableau 4. Caractéristiques des endosymbiotes**

Nous avons fait figurer les différentes caractéristiques génomiques qui existent depuis les bactéries libres jusqu’aux symbiotes primaires. Des symbiotes nouvellement obligatoires y figurent également, phénotype qu’ils ont acquis depuis un endosymbiote secondaire ancestral (\*) ou directement depuis une bactérie libre (\*\*).

Les néo-endosymbiotes possèdent également une composition nucléotidique peu biaisée, témoin de leur récente transition évolutive (Oakeson *et al.*, 2014 ; Burke et Moran, 2011). De plus, de nombreux gènes inactifs étant généralement encore présents dans leur génome, il est possible de connaître le type de mutations les ayant affectés (Burke et Moran, 2011). D’ailleurs, il est probable que nombre de leurs gènes apparemment fonctionnels

soient en réalité des pseudogènes n'ayant pas subi de mutations délétères apparentes. Lamelas et ses collaborateurs (2011) ont essayé de déterminer le nombre de ces pseudogènes « cryptiques » chez *Serratia symbiotica*, et l'ont estimé à 200 parmi les 2100 gènes présumés actifs, portant le total de pseudogènes à 740.

### c) Symbioses anciennes

Le processus d'érosion et de mutations initié dès la mise en place de l'association endosymbiotique va se poursuivre pendant des dizaines de millions d'années. Les symbiotes plus anciens ont ainsi une composition en nucléotide clairement biaisée, avec des génomes riches en bases A et T (Moran *et al.*, 2008). Les gènes inactivés étant progressivement éliminés, leurs génomes contiennent des pseudogènes, mais en proportion moindre par rapport au néo-endosymbiotes. Le nombre de gènes conservés est par contre bien plus faible, ainsi que la taille de leurs génomes. Il y a donc un biais énorme vers la perte de fonction, dû au processus du cliquet de Muller qui régit l'évolution de leur génome.

#### *Les « génomes-nains » des endosymbiotes primaires*

Les endosymbiotes primaires possèdent ainsi des génomes parmi les plus petits qui soient dans le monde procaryote, fortement enrichis en bases AT, et qui conservent un nombre restreint de gènes parmi le répertoire ancestral (Moran *et al.*, 2008 ; **Tableau 4**). Les premiers génomes séquencés passaient sous la barre des 500 000 bases, stagnant à une valeur proche de ce qu'on pensait être la taille minimale d'un organisme bactérien (Nakabachi *et al.*, 2006). Des génomes présentant une réduction encore plus drastique ont alors été révélés. Ainsi en est-il du symbiote primaire des Psylloidea, *Carsonella ruddii*, qui, lorsqu'il a été séquencé, proposait avec ses 159 662 bases un génome qui réduisait de deux tiers le record précédemment établi (Nakabachi *et al.*, 2006). Auparavant avait été séquencé *B. aphidicola*, notre symbiote modèle, fort de ses 450 000 bases chez *A. pisum*, qui ferait presque figure de géant en comparaison (Shigenobu *et al.*, 2000). Le symbiote primaire de certains Coccoidae, la Proéobactérie Beta *Tremblaya princeps*, propose également des génomes tutoyant le répertoire minimal requis pour la vie : le symbiote de la cochenille farineuse des agrumes *Planococcus citri* possède un génome de 138 931 bases (López-Madrugal *et al.*, 2011). Outre les Protéobactéries, cette convergence vers des génomes réduits se retrouve également chez les Bacteroidetes obligatoires de



Coccoidae, *Walczuchella monophebidarum* et *Llaveia axin axin* (Rosas-Pérez *et al.*, 2014 ; Sabree *et al.*, 2013). Le record pour l'instant est cependant à rechercher chez les Homoptères : le génome de *Nasuia deltocephalinicola*, symbiote de *Macrosteles quadrilineatus* (Cicadellidae), dépasse ainsi tout juste les 110 000 bases et ne comporte que 137 gènes codant pour une protéine (Bennett et Moran, 2013).

De plus, les endosymbiotes primaires ne possèdent généralement plus de système permettant la recombinaison. De ce fait, les éléments mobiles sont progressivement éliminés de leurs génomes. Cette érosion conduit à une densité très forte en séquences codantes, laissant peu de place aux régions intergéniques. Sur les 182 gènes du génome de *C. ruddii*, 164 chevauchent au moins l'un des deux gènes qui leur sont adjacents, pour une densité (record ?) de 97,3% pour l'ensemble des séquences codantes (auxquels il convient d'ajouter les ARN ribosomiaux ainsi que les ARNt) (Nakabachi *et al.*, 2006). Enfin, la synténie dans les génomes des symbiotes primaires est très conservée, ce qui peut être mis en lien avec cette faible part occupée par les séquences intergéniques, ainsi qu'à la perte des systèmes de recombinaison dans ces génomes (Silva *et al.*, 2003). On estime par exemple que 50 millions d'années se sont écoulées sans que n'ait été perturbé l'ordre des gènes chez certaines souches de *B. aphidicola* (Tamas *et al.*, 2002). Chez le symbiote primaire de certains aleurodes et cochenilles cependant, des régions non codantes répétées permettraient tout de même un certain remaniement de la synténie sans qu'il y ait présence de systèmes de recombinaison (McCutcheon et von Dohlen, 2011 ; Sloan et Moran, 2013). Ces éléments répétés engendrent-ils « mécaniquement » des taux importants de recombinaison, ou bien est-ce que ce sont les longues portions non codantes qui favorisent la fixation de réarrangements non délétères (Sloan et Moran, 2013) ?

#### Encadré 4. Endosymbiotes ou organelles ?

Nous avons parlé d'un répertoire minimal de gènes nécessaires à la vie. Pourtant, les mitochondries et les chloroplastes peuvent respectivement présenter des génomes de l'ordre de 6000 (Conway *et al.*, 2000) et 35000 bases (Kohler *et al.*, 1997). L'enjeu est donc pour certains scientifiques de trancher quant à l'appartenance de ces bactéries obligatoires des insectes aux endosymbiotes primaires ou aux organelles (López-Madrigal *et al.*, 2011 ; Husnik *et al.*, 2013). Un organisme vivant posséderait trois propriétés : une homéostasie, une reproduction autonome et une capacité à évoluer (Luisi, 2002). Force est de constater que selon cette définition, nombre de symbiotes primaires – si ce n'est l'ensemble de ceux que nous avons abordé dans le manuscrit – n'appartiendraient pas aux organismes « vivants ». C'est en tous les cas l'avis de López-Madrigal et de ses collaborateurs (2011), qui admettent tout de même lors de la publication de « leur » génome de *T. princeps* que ce symbiote ne ressemble pas aux organelles décrites. Le cas extrême de *Tremblaya* a depuis particulièrement stimulé le besoin de catégorisation des scientifiques, puisque deux années plus tard, Husnik et ses collaborateurs (2013) concluent que la dégénérescence extrême de cette souche de *T. princeps* suggère une transition vers un organisme « plus proche d'une organelle » (« *more akin to an organelle* »). Selon eux, une définition claire de ce qu'est une organelle comparativement à un endosymbiote fait défaut. Ils éliminent finalement *T. princeps* des organelles sur la base de l'absence d'un transfert intense de gènes bactériens – fonctionnels – dans le génome de l'hôte, ce processus ayant systématiquement eu lieu chez les différentes organelles (Keeling et Archibald, 2008 ; Timmis *et al.*, 2004).

#### *Symbiotes facultatifs*

Les symbiotes facultatifs, malgré un cycle de vie proche des symbiotes primaires, ne présentent pas une réduction génomique aussi intense que ces derniers (Tableau 4). Là encore, les chiffres sont très variables, la longueur du génome de *S. symbiotica* atteignant trois mégabases chez le puceron (Burke et Moran, 2011), tandis que le manipulateur de la reproduction *Cardinium hertigii* présente un génome long de moins de 900 000 bases chez la guêpe *Encarsia pergandiella* (Penz *et al.*, 2012). Cela reste toutefois très supérieur aux valeurs que l'on retrouve chez les symbiotes primaires. Cette différence n'est pourtant pas dû au temps écoulé depuis leur transition vers un mode de vie endosymbiotique : les genres *Rickettsia* et *Wolbachia*, dont les génomes font plus d'un mégabase (Wu *et al.*, 2004 ; Ogata *et al.*, 2005), auraient divergé il y a plus de 400 millions d'années, leur transition vers l'endosymbiose précédant donc largement la symbiogenèse de l'ensemble des partenaires obligatoires des Hémiptères (Moran *et al.*, 1993) ! Il faut donc rechercher dans leur écologie – au sens très large – les bases de ces différences entre endosymbiotes primaires et secondaires.

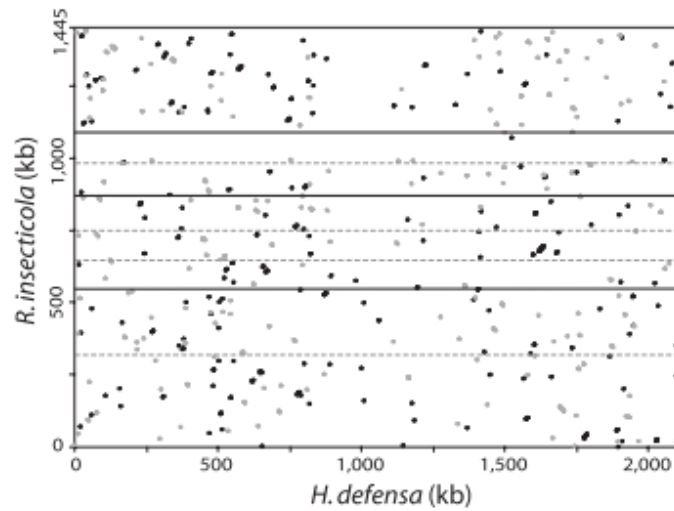
Les endosymbiotes primaires ont un spectre tissulaire très réduit, ne colonisant qu'un seul type cellulaire, généralement bactériocytaire. L'homéostasie dans cette cellule est contrôlée par l'hôte afin de protéger le symbiote des variations environnementales, mais également du système immunitaire de l'insecte (Braendle *et al.*, 2003). À l'inverse, l'habitat des symbiotes facultatifs ne se cantonne pas aux bactériocytes, voire, certains peuvent connaître des phases extracellulaires (voir en amont). Du fait de ce spectre plus élargi de tissus, la bactérie va être confrontée à des variations environnementales plus importantes que les symbiotes primaires. Par conséquent, les systèmes de perception, de régulation, de synthèse et de transports de métabolites divers seront davantage conservés chez les symbiotes secondaires (Moran *et al.*, 2008 ; Degnan et Moran, 2009). Cette colonisation active des tissus passe également par un contournement du système immunitaire de l'hôte, auquel pourraient contribuer de nombreux facteurs dits « de virulence » tels que la paroi cellulaire de la bactérie ou des systèmes de sécrétion (Dale *et al.*, 2001). Ces systèmes peuvent également contribuer à l'établissement de la bactérie dans les cellules de l'hôte lors de la transmission verticale.

Moins intense que chez les symbiotes primaires, le processus d'érosion n'en est pas moins continu chez les endosymbiotes facultatifs. En conséquence, de nombreux pseudogènes sont présents dans leurs génomes, contrairement aux endosymbiotes primaires (**Tableau 4**). Près de 200 ont été répertoriés dans les génomes de *R. insecticola* souche LSR1 et d'*H. defensa* souche 5AT (Degnan *et al.*, 2009 ; Degnan *et al.*, 2010), ce qui est supérieur au nombre que l'on peut trouver chez *Escherichia coli* ou *Yersinia pestis* (environ 150 pseudogènes, Degnan *et al.*, 2010), et d'autant plus si l'on pondère par la taille de leurs génomes respectifs. Là encore, une très grande variabilité existe chez les endosymbiotes facultatifs. Chez *C. hertigii* symbiote d'*Encarsia pergandiella*, le nombre de pseudogènes est extrêmement faible puisqu'on n'en dénombre que trois (Penz *et al.*, 2012), mais cela pourrait être dû à des problèmes d'annotation (Santos-Garcia *et al.*, 2014).

Les symbiotes secondaires conservent entre autre un système de recombinaison fonctionnel, bien qu'ils aient perdu de nombreux gènes assumant cette fonction. Cette capacité à recombiner, combinée à la part importante de pseudogènes sont des conditions favorables au maintien des éléments mobiles (phages, transposases, plasmides etc.),

d'autant plus dans des populations soumises à la dérive. Dans le génome de la souche wMel du genre *Wolbachia*, 158 familles d'éléments mobiles ont par exemple été recensées (Wu *et al.*, 2004). Chez *H. defensa* et *R. insecticola*, les éléments mobiles représentent respectivement 15 et 20% du nombre total de séquences codantes que contiennent leurs génomes (Degnan *et al.*, 2009 ; Degnan *et al.*, 2010). Il est toutefois à noter que nombre de ces gènes pourraient être inactifs (Degnan *et al.*, 2009). La présence d'éléments répétés n'est pas sans conséquence sur la synténie dans ces génomes, largement malmenée chez les endosymbiotes secondaires, même entre organismes apparentés (**Figure 8**).

En conclusion, malgré une évolution génomique présentant de nombreuses similarités (érosion du contenu ancestral, enrichissement en bases A et T, réduction de la taille etc.), il existe des différences intéressantes entre les génomes des symbiotes secondaires et obligatoires. Celles-ci concernent autant la dimension de ces génomes que leurs propriétés, les uns de petite taille et à l'architecture relativement stable, les autres de dimension plus importante et à l'architecture dynamique. Une telle variabilité existe également entre différents endosymbiotes secondaires, pour lesquels il est finalement difficile de dresser un portrait-type. C'est justement cette variabilité qui va nous intéresser dans le cadre de cette thèse. Nous allons notamment essayer de comprendre comment les caractéristiques propres à chaque symbiote en termes d'écologie et d'association avec son hôte vont influencer sur sa trajectoire évolutive.



**Figure 8. Comparaison de l'organisation des gènes chez *R. insecticola* et *H. defensa***

Ce graphique représente la localisation et l'orientation relatives des différents orthologues présents dans les génomes des deux organismes. Les cercles noirs représentent des gènes présents dans la même orientation, les cercles gris des gènes dans des orientations inversées. Les quatre lignes horizontales noires séparent les différents fragments de l'assemblage de *R. insecticola*, et sont eux-mêmes séparés en plusieurs *contigs* (lignes pointillées).

# PROBLEMATIQUE

---

Les endosymbiotes primaires et secondaires suivent des trajectoires distinctes dans l'évolution de leurs génomes, corrélées à des différences importantes en termes d'écologie. Les endosymbiotes secondaires se distinguent par une écologie plus complexe, plus riche, plus diversifiée que les endosymbiotes primaires. Leur capacité à envahir les différents tissus de l'hôte, voire à survivre hors du corps de celui-ci, ainsi que leur propension à transférer horizontalement, en font des bactéries intracellulaires particulières, qui vont faire face à des variations environnementales inconnues de leurs partenaires obligatoires.

Qu'en est-il entre des endosymbiotes secondaires qui ne partagent pas les mêmes niches écologiques, n'apportent pas le même bénéfice à l'hôte, ne possèdent pas les mêmes capacités de transmission ? Il est certain que l'ensemble de ces paramètres aura des conséquences importantes sur les pressions de sélection auxquelles le symbiote sera soumis, et affecteront la trajectoire évolutive empruntée par son génome. Ainsi, ces différences en termes d'écologie chez les endosymbiotes secondaires devraient expliquer pour partie leur variabilité génomique. Cette question de l'influence de l'écologie sur l'évolution de leur génome sera au cœur de ce travail de thèse.

Les niveaux de sélection auxquels sont confrontés les génomes des endosymbiotes secondaires seront particulièrement à prendre en compte pour comprendre cette évolution. Nous tâcherons ainsi de confronter nos résultats à la théorie holobiontique en fonction de ces niveaux de sélection, en essayant de déterminer si les intérêts de l'hôte coïncident toujours avec ceux de ces endosymbiotes secondaires qu'il héberge, et ce quelle que soit l'échelle évolutive prise en considération. Nous verrons également que les endosymbiotes vont interagir avec différents partenaires, dont les intérêts convergent ou non avec ceux de l'hôte, et dont la présence va avoir une influence énorme sur l'évolution du génome des symbiotes.



# MODELE BIOLOGIQUE

---

Les bactéries appartenant à l'espèce *Hamiltonella defensa* sont revenues tout le long de mon introduction, leur phénotype chez les pucerons ayant fait l'objet de nombreuses publications. Nous allons revenir plus en détail sur cette bactérie, qui a constitué le centre d'intérêt majeur de mon travail de thèse.

## I. *Hamiltonella-Regiella* : un clade d'endosymbiotes protecteurs

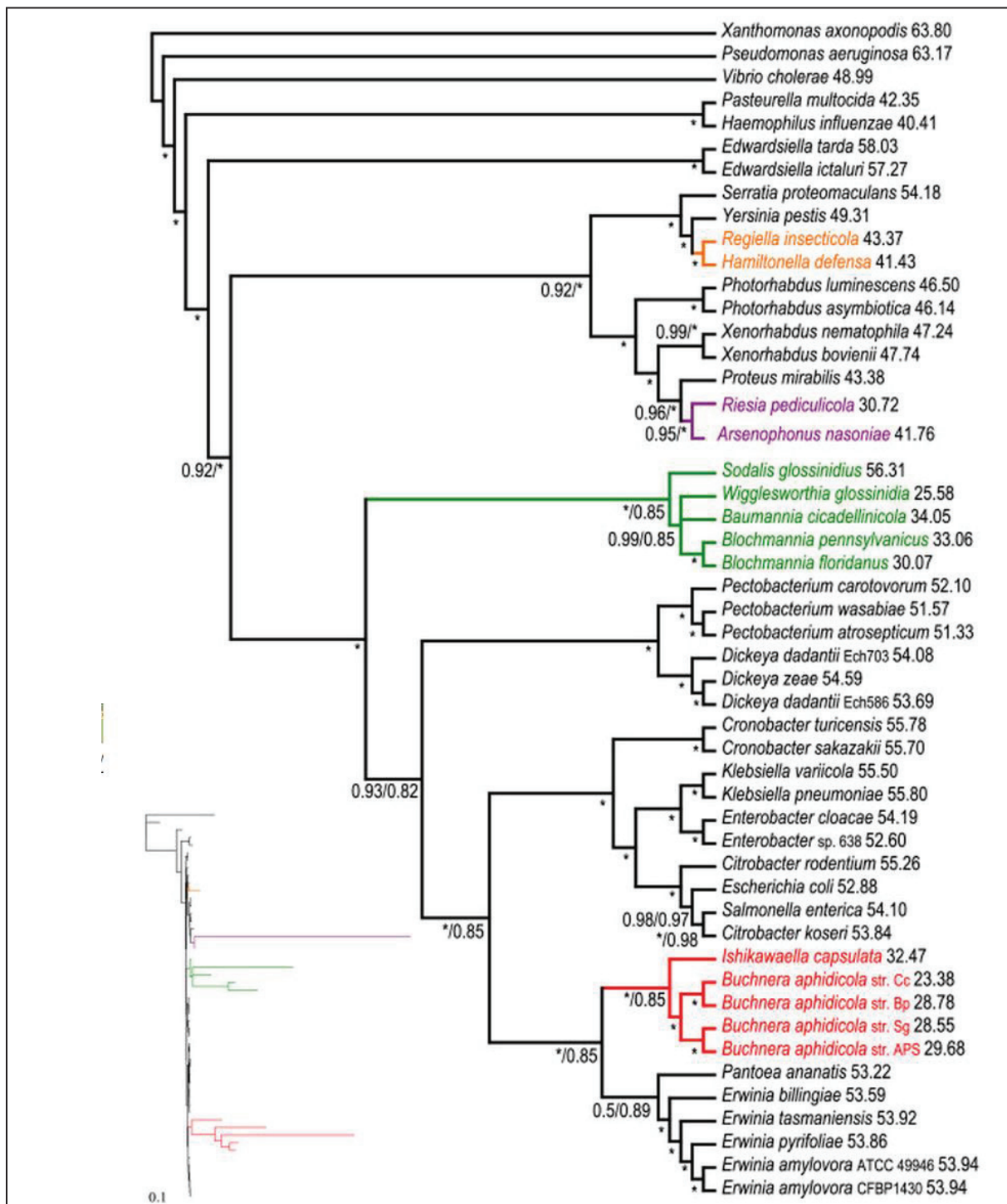
*H. defensa* est une Enterobacteriaceae, appartenant par conséquent à la Classe des Protéobactéries Gamma (**Figure 9**). Sa plus proche parente, la bactérie endosymbiotique *Regiella insecticola*, a elle aussi été évoquée plusieurs fois dans mon Introduction. Comme nous l'avons vu, leur phénotype principal à toutes les deux est la protection du puceron contre certains ennemis naturels tels que les guêpes parasitoïdes ou les champignons pathogènes. Cette caractéristique serait une déviation d'un probable phénotype pathogène chez leurs ancêtres, si l'on se réfère aux nombreux facteurs de virulence dont disposent ces deux symbiotes (Degnan *et al.*, 2009 ; Degnan *et al.*, 2010), ainsi qu'à leur parenté avec les bactéries du genre *Yersinia*.

## II. La symbiose puceron-*Hamiltonella defensa*

### 1. Un symbiote versatile

*H. defensa* a été répertoriée pour la première fois chez le puceron vert du pois en 2001 par Darby et ses collaborateurs, une décennie après sa première identification chez les aleurodes du genre *Bemisia* (Clark *et al.*, 1992). Ceci explique pourquoi elle fut tout d'abord baptisée « PABS », pour « Pea Aphid *Bemisia*-like Symbiont ». Ce terme a coexisté pendant plusieurs années avec celui beaucoup plus abstrait de bactérie du « type T » (Sändstrom *et al.*, 2001), avant que le nom de « *Candidatus Hamiltonella defensa* » ne soit proposé par Moran et ses collaborateurs (Moran *et al.*, 2005c).

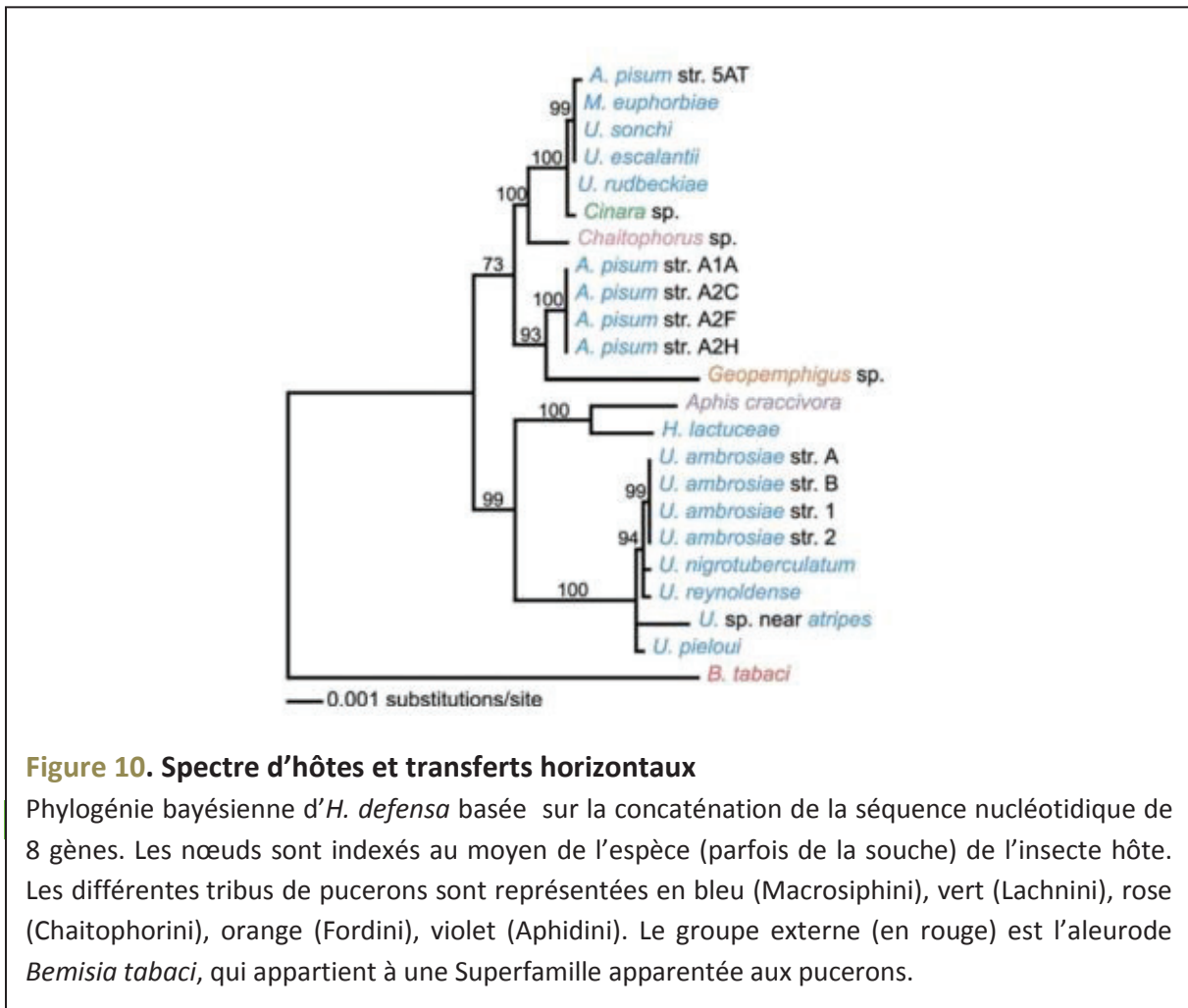




**Figure 9. Positionnement phylogénétique d'*Hamiltonella defensa*** (d'après Husnik *et al.*, 2011)

Ce cladogramme a été réalisé avec PhyloBayes en utilisant la concaténation de 69 séquences protéiques. Les astérisques signifient que le clade est très soutenu. Les valeurs à droite représentent le contenu en G et C des 69 gènes correspondants. Cette analyse propose que le clade monophylétique *Hamiltonella defensa* et *Regiella insecticola* a pour plus proches parentes les bactéries pathogènes du genre *Yersinia*.

Depuis cette découverte, son spectre d'hôtes chez les pucerons s'est révélé important, et on estime que 14% des espèces pourraient être infectées (Oliver *et al.*, 2010). La phylogénie de l'insecte et celle du symbiote ne sont pas congruentes (Degnan et Moran, 2008 ; Telesnicki *et al.*, 2012), ce qui révèle sa faculté à transférer horizontalement typique d'un symbiote secondaire (**Figure 10**).



**Figure 10. Spectre d'hôtes et transferts horizontaux**

Phylogénie bayésienne d'*H. defensa* basée sur la concaténation de la séquence nucléotidique de 8 gènes. Les nœuds sont indexés au moyen de l'espèce (parfois de la souche) de l'insecte hôte. Les différentes tribus de pucerons sont représentées en bleu (Macrosiphini), vert (Lachnini), rose (Chaitophorini), orange (Fordini), violet (Aphidini). Le groupe externe (en rouge) est l'aleurode *Bemisia tabaci*, qui appartient à une Superfamille apparentée aux pucerons.

## 2. Un phénotype protecteur répandu

La présence du symbiote dans l'hémolymph est tout particulièrement intéressante (Moran *et al.*, 2005 ; Dykstra *et al.*, 2014), puisque c'est dans ce fluide que va se faire le dépôt des œufs de parasitoïdes, contre qui elle va protéger l'hôte. Par conséquent, sa liberté de mouvement dans le corps de son hôte faciliterait l'interaction avec les ennemis de celui-ci, ce qui suggère que la bactérie peut contourner les barrières défensives du puceron. De

manière intéressante, il a été démontré qu'*H. defensa* pouvait survivre dans les cellules phagocytaires du puceron (Schmitz *et al.*, 2012).

Ce phénotype protecteur semble étendu chez les pucerons, puisqu'on a retrouvé des souches d'*H. defensa* protectrices chez *A. pisum* (Oliver *et al.*, 2009), *A. craccivora* (Asplen *et al.*, sous presse, cité par Dykstra *et al.*, 2014), et *A. fabae* (Schmid *et al.*, 2012). Chez *Sitobion avenae* toutefois, nulle protection n'a pu être mise en évidence (Łukasik *et al.*, 2013). Peut-être les souches infectant cette espèce sont-elles réellement non protectrices, ou bien les conditions expérimentales étaient-elles défavorables à l'action du symbiote. Par exemple, peut-être protègent-elles contre des ennemis autres que ceux utilisés dans le cadre de cette expérience.

### **3. Des coûts constitutifs**

Son phénotype protecteur ne conduit jamais à la fixation d'*H. defensa* dans les populations naturelles de pucerons. Sa prévalence variait ainsi de 30 à 60% dans des populations nord-américaines d'*A. pisum* (Russel *et al.*, 2013), et de 0 à 85% dans des populations d'*A. craccivora* (Brady *et al.*, 2014 ; Brady et White, 2013). La transmission maternelle du symbiote, si elle est imparfaite, avoisine les 100% (Darby et Douglas, 2003 ; Dykstra *et al.*, 2014). En laboratoire, des lignées infectées ont par exemple été maintenues pendant près de 1500 générations (Weldon *et al.*, 2013). Le bénéfice apporté couplé à cette transmission efficace devrait mener à des prévalences importantes.

Cette proportion importante d'individus non-infectés témoigne du coût que la présence du symbiote inflige à son puceron d'hôte. Ces coûts ne sont pas toujours faciles à mettre en évidence en laboratoire. En revanche, les expériences dites de « cages à population » sont sans appel : en l'absence de pression parasitaire, les individus infectés connaissent un déclin rapide, soulignant leur coût important pour l'hôte (Oliver *et al.*, 2008).

### **4. Bases génétiques et génomiques de cette protection**

Le génome d'une bactérie de l'espèce *H. defensa*, symbiote de la souche 5AT d'*A. pisum* a été séquencé et assemblé par Degan et ses collaborateurs (2009). Il présentait des caractéristiques typiques d'un symbiote secondaire : un génome de dimension réduite par rapport aux bactéries libres, mais supérieure à celles que l'on peut trouver chez un symbiote

primaire, un biais en A et T important (60%), de nombreux pseudogènes (188) ainsi qu'une invasion du génome par divers éléments mobiles – introns de type II, séquences d'insertions, prophages et plasmides intégrés au génomes, etc. – qui couvrent au moins 20% de celui-ci. Il dépendrait du symbiote primaire pour l'approvisionnement en acides aminés essentiels, de l'hôte pour l'approvisionnement en sucres et divers autres métabolites. Il est cependant capable de synthétiser divers cofacteurs (pyridoxal-5-phosphate, niacine, folate, biotine etc.), dont il pourrait faire bénéficier le consortium. Le séquençage a aussi révélé l'existence d'un plasmide de recombinaison, pHD5AT, qui serait impliqué dans l'échange de gènes.

L'on peut corréliser la présence de certains gènes/systèmes avec le phénotype et l'écologie de la bactérie. Elle possède ainsi de nombreux gènes impliqués dans la perception des stimuli environnementaux (détection du quorum, systèmes à deux composantes), qui lui permettraient de faire face aux variations environnementales que le symbiote subit durant ses phases d'invasion des différents tissus. Surtout, l'analyse de son génome a révélé l'arsenal dont elle aurait hérité depuis ses ancêtres pathogènes, et qui pourrait là encore être en lien avec son pouvoir invasif, mais aussi directement avec la protection conférée à son hôte. Au moins trois systèmes de sécrétion ont ainsi été mis en évidence, accompagnés par tout un assortiment d'effecteurs et de toxines (effecteurs de type III, toxines RTX). Enfin une protéine CdtB (*cytolethal distending toxin B*), particulièrement toxique pour les eucaryotes, serait en outre encodée par le génome d'un phage intégré à son génome, et sur lequel nous allons revenir.

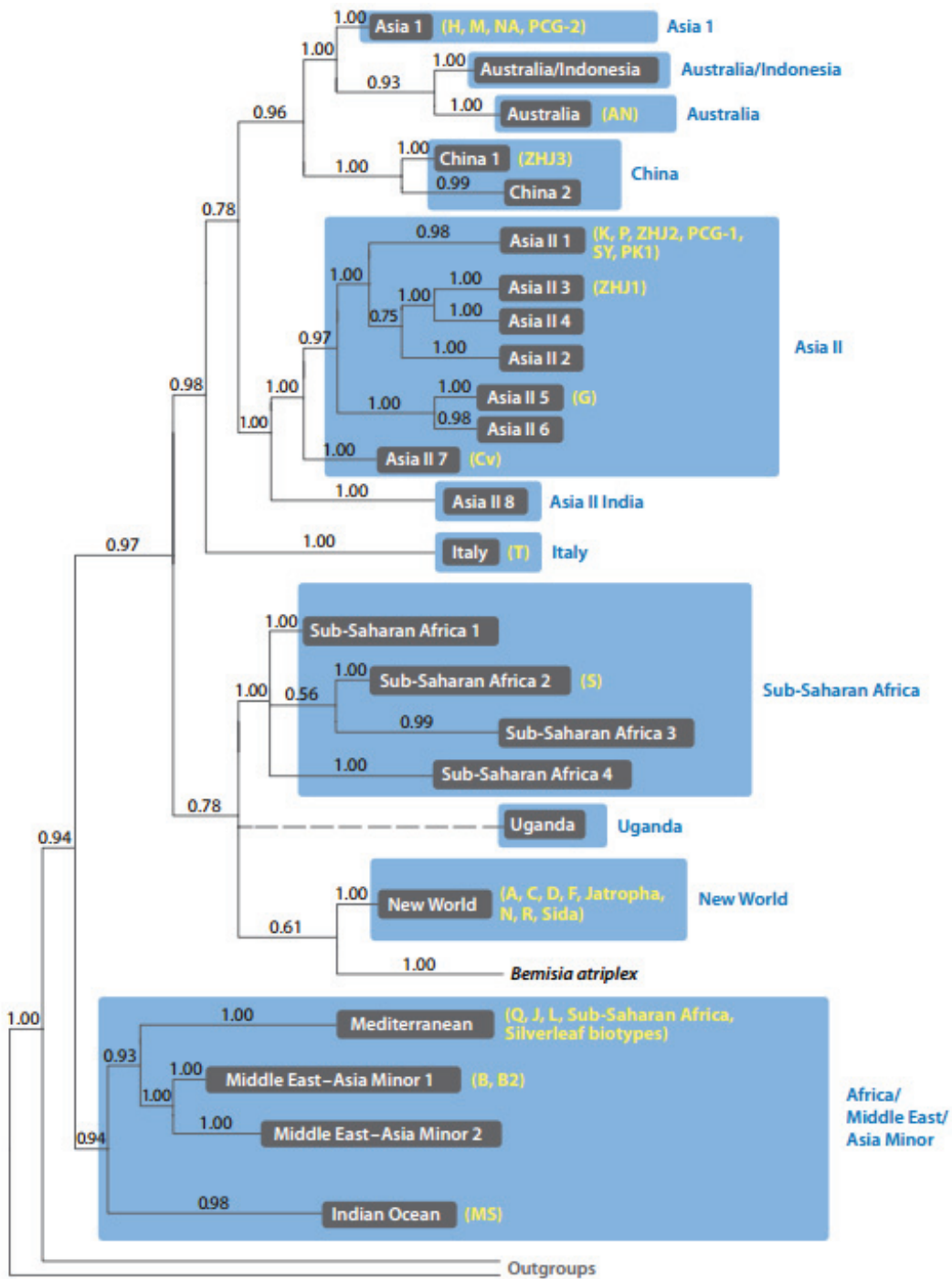
Ainsi, derrière le phénotype protecteur de la bactérie se cache en réalité un second protagoniste, le bactériophage APSE (« *Acyrtosiphon pisum* Secondary Endosymbiont »). Ce virus est apparenté au phage lambdaïde P22 (Podoviridae), et est un phage tempéré, que l'on retrouve intégré dans le génome de son hôte – bactérien cette fois – *H. defensa*. Différents variants ont été mis en évidence chez les pucerons, numérotés de APSE1 à APSE6 (APSE7 faisant référence au phage identifié chez le symbiote de l'aleurode *Bemisia tabaci*) : les phages APSE1, 2 et 3, sont des partenaires d'*H. defensa* infectant *A. pisum* ; APSE4, APSE5 et APSE6 sont des partenaires de bactéries *H. defensa* ayant été retrouvés respectivement chez *A. craccivora*, *Uroleucon rudbeckiae* et *Chaitophorus* sp. (Degnan et Moran, 2008). La protection a été démontrée chez APSE2 et APSE3, le second étant davantage plus virulent

pour le parasitoïde. L'enzyme impliquée serait une toxine potentielle nommée YDp, alors qu'ASPE2 encode la susmentionnée toxine CdtB (cytolethal distending toxin). Les autres variants encodent chacun au moins une toxine, bien qu'elle puisse différer, ce qui suggère que tous sont impliqués dans la protection des pucerons. Cette diversité dans les armes que possèdent ces phages indique que des transferts de gènes ont lieu entre eux. Cette acquisition de nouvelle fonction est à mettre en lien direct avec le phénotype protecteur du duo APSE/*H. defensa*, et avec la course aux armements engagés contre le parasitoïde.

### III. *Bemisia tabaci* et l'endosymbiose

*H. defensa* est également associé à l'aleurode du tabac *Bemisia tabaci* (Gennadius). Toutefois, la relation existant entre ces deux partenaires est bien moins établie. L'endosymbiose chez cet insecte a constitué le point de départ de mon travail de thèse, et a par conséquent particulièrement monopolisé mon attention. C'est pourquoi nous allons consacrer plusieurs chapitres à ce modèle particulier.

Précisons tout de suite que l'entité linnéenne *B. tabaci* ne fait pas référence à une espèce à proprement parler. Elle désigne un complexe d'espèces dites « **cryptiques** » du fait de l'incapacité à les discriminer sur la base de leur morphologie (**Figure 10**). Cette caractéristique est à l'origine de la confusion historique régnant dans la taxonomie de cet insecte, et dont les stigmates sont encore visibles de nos jours (**Encadré 5**).



**Figure 10. Phylogénie du complexe *Bemisia tabaci* (d'après de Baro *et al.*, 2011)**

Les espèces, désignées par un texte blanc dans un cadre gris, sont regroupées en onze grands «groupes» taxonomiques. Quant aux anciens biotypes, ils sont référés en jaune à droite du nom de l'espèce à laquelle ils appartiennent.



## Encadré 5. Un historique... complexe

L'historique de la taxonomie de cet aleurode mérite qu'on s'y arrête quelque peu (pour une revue, lire par exemple Perring, 2001). En effet, la première description officielle de l'insecte en 1889 par Gennadius, sous le nom d'*Aleyrodes tabaci*, n'a pas empêché la coexistence d'une vingtaine de termes qualifiant ce que nous appelons couramment de nos jours « *Bemisia tabaci* » (Tableau 5).

Synonymes de *B. tabaci* (Gennadius) avec leur localité et leur plante hôte (adapté de Perring, 2001)

Synonyme	Localité	Plante hôte	Famille
<i>Aleyrodes tabaci</i> (Gennadius, 1889)	Grèce	Tabac <i>Nicotiana</i> sp.,	Solanaceae
<i>Aleyrodes inconspicua</i> (Quaintance, 1900)	États-Unis	<i>Physalis</i> sp.	Solanaceae
<i>Bemisia inconspicua</i> (Quaintance)(Quaintance and Baker, 1914)	États-Unis	<i>Physalis</i> sp.	Solanaceae
<i>Bemisia emilliae</i> (Corbett, 1926)	Sri Lanka	<i>Emilia sonchifolia</i> (L.) DC. Ex Wight	Asteraceae
<i>Bemisia signata</i> (Bondar, 1928)	Brésil	<i>Nicotiana glauca</i> Graham	Amaranthaceae
<i>Bemisia bahiana</i> (Bondar, 1928)	Brésil	<i>Nicotiana tabacum</i> , L.	Solanaceae
<i>Bemisia costa-limai</i> (Bondar, 1928)	Brésil	<i>Euphorbia hirtella</i> Boiss.	Euphorbiaceae
<i>Bemisia gossypiperda</i> (Misra and Lamba, 1929)	Inde, Pakistan	Coton, <i>Gossypium</i> sp.	Malvaceae
<i>Bemisia achyranthes</i> (Singh, 1931)	Inde	<i>Achyranthes aspera</i> L.	Amaranthaceae
<i>Bemisia hibisci</i> (Takahashi, 1933)	Taiwan	<i>Hibiscus rosa-sinensis</i> L.	Malvaceae
<i>Bemisia longispina</i> (Priesner and Hosny, 1934)	Égypte	<i>Psidium guajava</i> L.	Myrtaceae
<i>Bemisia gossypiperda</i> var. <i>mosaicivectura</i> (Ghesquiere, 1934)	Zaire	<i>Jatropha multifida</i> L.	Euphorbiaceae
<i>Bemisia goldingi</i> (Corbett, 1935a)	Nigeria	Coton, <i>Gossypium</i> sp.	Malvaceae
<i>Bemisia nigeriensis</i> (Corbett, 1935a)	Nigeria	Cassava, <i>Manihot</i> sp.	Euphorbiaceae
<i>Bemisia rhodesiaensis</i> (Corbett, 1936)	Zimbabwe	Tabac, <i>Nicotiana</i> sp.	Solanaceae
<i>Bemisia tabaci</i> (Gennadius) (Takahashi, 1936)	Îles Mariannes	<i>Brassica oleracea</i> , L.	Cruciferaeae
<i>Bemisia manihotis</i> (Frappa, 1937)	Madagascar	Cassava, <i>Manihot</i> sp.	Euphorbiaceae
<i>Bemisia vayssierei</i> (Frappa, 1938)	Madagascar	Tabac, <i>Nicotiana</i> sp.	Solanaceae
<i>Bemisia</i> ( <i>Neobemisia</i> ) <i>hibisci</i> (Visnya, 1941)	Taiwan	<i>Hibiscus rosa-sinensis</i> L.	Malvaceae
<i>Bemisia</i> ( <i>Neobemisia</i> ) <i>rhodesiaensis</i> (Visnya, 1941)	Zimbabwe	Tabac, <i>Nicotiana</i> sp.	Solanaceae
<i>Bemisia loniceriae</i> (Takahashi, 1957)	Japon	<i>Lonicera japonica</i> Thunb. Ex Murray	Caprifoliaceae
<i>Bemisia minima</i> (Danzig, 1964)	Géorgie	<i>Elsholtzia patrini</i> (Lepech.) Garcke	Lamiaceae
<i>Bemisia miniscula</i> (Danzig, 1964)	Géorgie	<i>Lamium purpureum</i> L.	Lamiaceae

## Tableau 5. Synonymes de *Bemisia tabaci* (tous étant désormais désuets)

Pour résoudre cet imbroglio, des campagnes de synonymisation menées entre 1952 et 1978 ont permis de regrouper sous la même désignation ces insectes aux morphes indistincts, tout en rendant à Gennadius ce qui lui revenait de droit : la paternité de la description de *Bemisia tabaci*. Toutefois, en dépit de leur ressemblance, les populations de *B. tabaci* forment un ensemble écologiquement hétérogène. Afin de capturer cette diversité, le terme de « **biotypes** » a été introduit, ceux-ci étant établis sur la base d'informations enzymatiques (estérases) puis moléculaires. Ce terme avait ceci de pratique qu'il est taxonomiquement vague, cette imprécision permettant de ne point spéculer sur le rang à attribuer aux variants de *B. tabaci* faute de données déterminantes. De Baro et ses collaborateurs (2011) proscrirent ensuite l'emploi du terme « biotypes » dans ce qui constitue à l'heure actuelle le dernier rebondissement de cette histoire, et découperont *B. tabaci* en un complexe non plus de races mais bien d'espèces cryptiques, celles-ci étant alors au nombre de 24 et se répartissant en 11 groupes taxonomiques bien distincts (Figure). D'autres espèces ont depuis régulièrement été décrites (Hu *et al.*, 2011 ; Liu *et al.*, 2012 ; Firdaus *et al.*, 2013), pour arriver aux 35 ayant été recensées jusqu'à ce jour. Aujourd'hui encore, la question des espèces chez *B. tabaci* fait débat. Lors d'un symposium en Crête portant sur les aleurodes, le manque d'homogénéité dans les communications a d'ailleurs conduit Boykin à lancer un appel à la normalisation (2014). Pour notre part, nous nous plierons aux recommandations de de Baro et parlerons bien d'espèces et non de biotypes.

## 1. Ennemi d'états

La première caractéristique de cet insecte, qui ne nous intéressera que de loin mais qui justifie la stimulation s'exerçant autour de ces aleurodes, tient dans l'impressionnante capacité invasive de certaines espèces ou sous-espèces de ce complexe. Elles font partie de la liste des 100 espèces les plus invasives élaborée par l'Union Internationale pour la Conservation de la Nature, et les dégâts qu'elles occasionnent se chiffrent en milliards de dollars (Brown *et al.*, 1995). Plusieurs paramètres expliquent le fléau que représente *B. tabaci* pour l'Homme. Tout d'abord, son régime alimentaire polyphage comprendrait un millier de plantes hôtes, et elle s'attaque aussi bien aux cultures qu'aux plantes ornementales (Byrne et Bellows, 1991). Ensuite, en plus des dégâts directs imputés à la consommation de phloème, cet Hémiptère typique excrète du miellat qui va diminuer la valeur marchande des végétaux et attirer fonges et bactéries. Enfin, et surtout, l'insecte est en association avec des centaines de virus phytopathogènes, notamment de la famille des Geminiviridae (Czosnek et Ghanim, 2011)

Toutes les espèces de *B. tabaci* ne figurent pas à la même position sur le tableau de chasse de l'Homme, et se distinguent par une grande hétérogénéité écologique (Delatte *et al.*, 2009 ; de Baro *et al.*, 2011). Elles n'ont pas le même spectre d'hôtes, la même résistance aux insecticides, et présentent des taux de fécondité et d'espérance de vie différents. Ces paramètres vont influencer énormément sur leur capacité d'invasion. Deux espèces sont ainsi parmi les plus nuisibles qui soient : *B. tabaci* MEAM1 (pour « **M**iddle-**E**ast/**A**sia **M**inor 1 », anciennement appelée biotype B) et MED (pour « **MED**iterranée », qui correspond aux anciens biotypes Q, J, K, L et ASL). Depuis leurs foyers d'origines, elles ont colonisé le monde entier (de Baro *et al.*, 2011), et sont capables d'exclure les espèces autochtones, tels qu'aux États-Unis ou sur l'île de la Réunion (Delatte *et al.*, 2009). Ces deux espèces vont être au cœur des deux premiers travaux figurant dans ce manuscrit.

## 2. Le compartiment endosymbiotique de *Bemisia tabaci*

Une seconde caractéristique de ce complexe va nous intéresser cette fois tout particulièrement dans le cadre de ce manuscrit. *B. tabaci* est en effet réputée pour la diversité de bactéries symbiotiques qu'elle héberge dans ses cellules, et pour ses taux importants de multi-infections (Gueguen *et al.*, 2010). Ainsi, l'on retrouve des souches



proches d'endosymbiotes parmi différentes espèces d'aleurodes, mais également différents endosymbiotes à l'intérieur d'une même espèce. Autre fait intéressantissime, les symbiotes primaires comme les symbiotes secondaires infectent les mêmes bactériocytes primaires, un cas unique chez les Hémiptères (Goettlieb *et al.*, 2008).

#### **a) Un endosymbiote obligatoire : *Portiera aleyrodidarum***

Tout d'abord, comme – pratiquement – tout Hémiptère qui se respecte, *B. tabaci* dépendrait d'un endosymbiote primaire, *Portiera aleyrodidarum*, pour l'acquisition de certains acides aminés essentiels (Baumann, 2005). Cette hypothèse n'a pas été démontrée expérimentalement comme pour de nombreux endosymbiotes primaires présumés, mais est étayée par l'ubiquité de l'endosymbiote chez les aleurodes, fixé dans toutes les populations et pour toutes les espèces étudiées, ainsi que par sa localisation stricte dans les bactériocytes. Au début de ma thèse, aucun autre élément n'était connu dans la relation entre ce symbiote et son hôte. Nous reviendrons cependant sur les avancées qui ont été réalisées durant ces dernières années, et desquelles nous sommes pour partie responsables.

#### **b) L'endosymbiose secondaire chez *Bemisia tabaci***

Outre *P. aleyrodidarum*, sept espèces de symbiotes facultatifs ont été répertoriées jusqu'à aujourd'hui dans ce complexe d'espèces qu'est *Bemisia tabaci*. Leur prévalence dans les populations est importante, notamment dans certaines populations qui comprennent 95% d'individus arborant au moins un de ces endosymbiotes secondaires (Gueguen *et al.*, 2010). Deux d'entre eux nous intéresseront davantage, dont notre bactérie *Hamiltonella defensa*. Nous expliquerons plus en aval le pourquoi de cet intérêt. Faisons dans un premier temps un rapide tour d'horizon de cette diversité endosymbiotique chez notre aleurode.

Nous trouvons tout d'abord de nombreuses Protéobactéries. Trois espèces d'endosymbiotes appartiennent ainsi à la famille des Rickettsiales (Sous-Division Alpha des Protéobactéries). *B. tabaci* héberge différentes souches de *Wolbachia*, dont le Supergroupe O, endémique de *B. tabaci* (Bing *et al.*, 2014), qui ne semblent pas manipulatrices. De la même façon, *B. tabaci* abrite des souches de *Rickettsia* proches de *Rickettsia bellii*, qui pourraient combiner des phénotypes mutualistes et manipulateurs (Parrella *et al.*, 2013 ; Himler *et al.*, 2011), et être à l'origine de l'invasion de *B. tabaci* en Italie et aux États-Unis.

Enfin, notre dernière protéobactérie alpha est un endosymbiote récemment décrit et endémique aux Hémiptères, *Hemipteriphilus asiaticus* (Bing *et al.*, 2013), dont les effets pour l'hôte demeurent totalement inconnus.

Deux autres symbiotes infectent les aleurodes et n'appartiennent pas au clade des Protéobactéries. Le premier, *Fritschea bemisiae* appartient à l'Ordre des Chlamydiales et à la famille peu connue des Simkaniaceae, qui ne comprend que deux genres (Corsaro et Greub, 2009). Des bactéries du genre *Fritschea* infectent également la cochenille *Eriococcus spurius*, mais chez cet hôte comme chez *B. tabaci*, le phénotype demeure totalement inconnu (Everett *et al.*, 2005). Le second, *Cardinium hertigii*, du phylum des Bacteroidetes, appartient à un genre qui pourrait infester 7% des espèces arthropodes (Zchori-Fein et Perlman, 2004 ; Gruwell *et al.*, 2009; Nakamura *et al.*, 2009). Certaines souches utilisent les mêmes stratégies de manipulation de la reproduction que les bactéries du genre *Wolbachia*, d'autres pourraient être mutualistes. La guêpe parasitoïde des aleurodes *Encarsia pergandiella* bénéficierait par exemple de la biotine synthétisée par ce symbiote (Penz *et al.*, 2012).

Viennent ensuite deux espèces d'Enterobacteriaceae, *Arsenophonus* sp. et donc *Hamiltonella defensa*. Les bactéries du genre *Arsenophonus* sont connues pour leurs phénotypes manipulateurs (par exemple, l'organisme modèle *A. nasoniae*), phytopathogènes (par exemple *A. phytopathogenicus* ; revue dans Bressan, 2014), et mutualistes (par exemple *Riesia pediculicola*, symbiote du pou, en réalité membre du clade *Arsenophonus* ; Perotti *et al.*, 2007). Ils pourraient également être mutualistes chez les Hémiptères, bien que peu d'éléments étayent cette hypothèse (Novakova *et al.*, 2009). Quant à *H. defensa*, dont nous avons déjà détaillé l'action protectrice chez le puceron, nous avons expliqué que son association chez l'aleurode demeurerait peu connue.

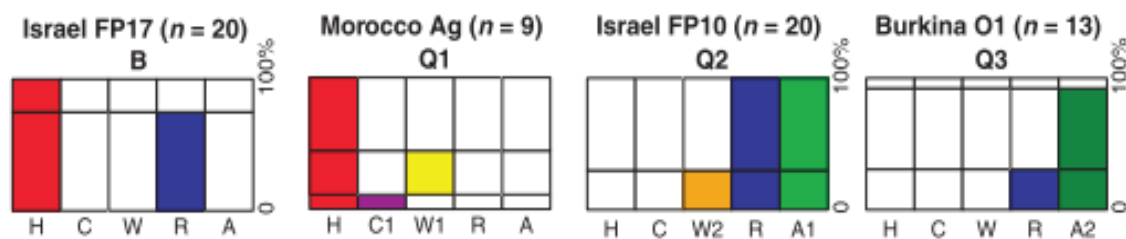
Il est à noter que comme certaines espèces de *B. tabaci* ont été clairement privilégiées dans les études qui ont été réalisées (notamment les espèces invasives MED et MEAM1), on peut imaginer que de nombreuses souches – voire espèces – d'endosymbiotes restent à découvrir dans ce complexe. De plus, il existe une diversité importante à l'intérieur de chaque espèce, puisqu'ont déjà été identifiées des souches distinctes de *Wolbachia* sp.,

*C. hertigii*, et *Arsenophonus* sp. (Gueguen *et al.*, 2010 ; Tsagkarakou *et al.*, 2012 ; Bing *et al.*, 2014).

### c) Endosymbiotes secondaires à forte prévalence chez l'aleurode

Cependant, et de manière intéressante, *Arsenophonus* sp. et *H. defensa* présentent des prévalences importantes chez *B. tabaci*, puisqu'on les retrouve près de la fixation dans les populations. Les multi-infections entre symbiotes primaires et secondaires sont par conséquent fréquentes. Par exemple, dans une étude portant sur les espèces MEAM1 et MED, 95% des 653 individus analysés étaient infectés par un symbiote secondaire en plus du symbiote primaire (Gueguen *et al.*, 2010). Du fait des prévalences moyennes des autres espèces endosymbiotiques, la présence simultanée de deux symbiotes secondaires est courante dans *B. tabaci* MED et MEAM1, et on la retrouve chez plus de 50% des individus.

De plus, les bactéries infectent uniquement les deux espèces les plus invasives, MED et MEAM1. Le patron de répartition de ces symbiotes chez *B. tabaci* est très loin d'être aléatoire. On retrouve des associations préférentielles en fonctions des espèces de *B. tabaci* (Chiel *et al.*, 2007). Par exemple, on ne retrouve pas *Arsenophonus* sp. chez MEAM1 alors qu'*H. defensa* y est pratiquement fixée (Gueguen *et al.*, 2010). En outre, des variants mitochondriaux ont été caractérisés chez *B. tabaci* MED, appelés « cytotypes », et dont la particularité majeure est d'être associés à des cortèges spécifiques de symbiotes. Si l'on regarde à l'échelle intra-spécifique, *H. defensa*, n'est jamais retrouvée dans les cytotypes Q2 et Q3 de l'espèce MED, et présente des prévalences très fortes chez le cytotype MED-Q1, quand *Arsenophonus* sp. présente le profil inverse, sa présence étant importante chez les individus Q2 et Q3, et inexistante chez les individus Q1 (**Figure 11** ; Gueguen *et al.*, 2010).



**Figure 11. Exemples de cortèges endosymbiotiques chez *Bemisia tabaci* MEAM1 et MED**

La prévalence de chaque symbiote est indiquée par différentes bandes verticales. Sont représentés *H. defensa* (H), *C. hertigii* (C), *Wolbachia* sp. (W), *Rickettsia* sp. (R) et *Arsenophonus* sp. (A). Les différents cortèges sont séparés par des lignes horizontales, et la taille de chaque bande indique la fréquence des différentes associations : dans la population de *B. tabaci* MEAM1 (biotype « B », à gauche), tous les individus sont infectés par *H. defensa*, certains étant en outre infectés par *Rickettsia* sp. ; dans la population de *B. tabaci* MED-Q1, ce sont *C. hertigii* (souche désignée par le terme « C1 ») et *Wolbachia* sp. (souche « W1 ») qui accompagnent *H. defensa* ; chez MED-Q2 et Q3, deux souches d'*Arsenophonus* sp. sont respectivement fixées, appelées ici A1 et A2.

Pour ce qui est des autres endosymbiotes, il existe une forte variabilité entre les populations à l'intérieur d'un même cytotype. Dans les populations tunisiennes du cytotype Q1, le plus invasif parmi l'espèce MED, Gueguen et ses collaborateurs (2010) n'ont retrouvé que le symbiote secondaire à forte prévalence, *H. defensa*. Dans les autres pays, *H. defensa* est accompagnée par différents autres endosymbiotes, *C. hertigii* (Uruguay, France, Maroc, Soudan, Cameroun), *Rickettsia* sp. (Soudan), et *Wolbachia* sp. (Cameroun, plus certaines populations de France et du Maroc). En Chine, ces quatre symbiotes sont retrouvés dans les populations, 1% des individus étant d'ailleurs infectés à la fois par *C. hertigii*, *Rickettsia* sp. et *H. defensa* (Pan *et al.*, 2012).

À l'inverse, *H. defensa* et *Arsenophonus* sp. ne sont jamais retrouvés dans le même hôte, pas plus que ne le sont *C. hertigii* et *Rickettsia* sp. (Gueguen *et al.*, 2010). Dans le cytotype ASL de l'espèce MED, trois souches d'*Arsenophonus* sp. ont été découvertes, sans que jamais n'aient été observés de cas de cohabitation au sein d'un même insecte (Gueguen *et al.*, 2010). Différentes hypothèses permettent d'expliquer ce phénomène. Cette répartition des symbiotes pourrait « simplement » refléter des infections indépendantes dans les différentes espèces ou sous-espèces (Gueguen *et al.*, 2010). Une autre hypothèse est que des interactions antagonistes à l'intérieur de l'hôte conduisent à des exclusions d'un symbiote par un autre. Enfin, la sélection inter-holobiontique peut contre-sélectionner certaines associations. Par exemple, nous avons vu que les hôtes infectés par des symbiotes

apportant la même fonction à l'holobionte seront contre-sélectionnés, subissant des coûts liés à la multi-infection sans compensation par un nouveau bénéfice.

Si l'on revient à la fixation d'*Arsenophonus* sp. et *H. defensa* chez leurs hôtes respectifs, cet évitement des deux bactéries est très intéressant. Ainsi, le fait qu'on ne les retrouve pas en association pourrait impliquer que leur rôle soit le même, en plus d'être potentiellement primordial. Devant la répartition particulièrement intéressante de ces symbiotes chez les espèces MED et MEAM1, ainsi que l'absence de données concernant leur phénotype, un projet collaboratif a vu le jour, impliquant différents partenaires en France (Équipes GEIHP et BAOBAB, Laboratoire de Biométrie et de Biologie Évolutive, Lyon), en Espagne (Équipe du Pr. Amparo Latorre, Institut Cavanilles, Valence,) ainsi qu'en Israël (Dr. Einat Zchori-Fein et Dr. Murad Ghanim, Agricultural Research Organization, Bet Dagan). Le but de cette collaboration était d'obtenir le métagénome intracellulaire des espèces MEAM1 et MED (cytotype Q2). Il consistait en le séquençage du métagénome par le Génoscope (Dr. Valérie Barbe Évry, France) et l'assemblage des génomes par nos soins. Les populations – israéliennes – utilisées étaient infectées respectivement par *H. defensa* et *Rickettsia* sp., et par *Arsenophonus* sp. et *Wolbachia* sp, sans oublier bien évidemment le symbiote primaire.

Au sein de ce projet, mon travail s'est focalisé sur ce dernier symbiote, *H. defensa*. Plusieurs faits expliquent ce choix. Tout d'abord, le séquençage d'*Arsenophonus* sp. n'a pas été mené avec succès, certainement du fait d'une fragilité de la bactérie, ou tout du moins de son génome (Diego Santos-Garcia, communication personnelle). Nous nous sommes donc intéressés davantage à l'espèce MEAM1, pour laquelle nous disposons de l'ensemble des partenaires. Parmi eux, *H. defensa*, de par sa fixation dans les populations, nous semblait la plus intéressante. Nous avons en outre établi une seconde collaboration avec l'équipe chinoise du Pr. Liu, qui avait assemblé le génome d'*H. defensa* infectant l'espèce MED-Q1 de *B. tabaci*, mais requérait nos conseils pour l'analyser. Nous avons ainsi la possibilité de mener à bien l'analyse génomique d'*H. defensa* MED-Q1, afin de détecter dans le génome de la bactérie des indices quant à son phénotype. Nous avons ensuite pu replacer l'évolution de ce symbiote dans un contexte plus général, en comparant les bactéries infectant le puceron et celles infectant les espèces MED et MEAM1 de *B. tabaci*, avant de nous intéresser de manière plus générale à l'évolution des symbiotes du clade *Hamiltonella-Regiella*.

Nous allons désormais reprendre le fil conducteur de mon travail de thèse, et les différentes questions suscitées au fur et à mesure des résultats.

### **Problématique Première. Interactions métaboliques entre endosymbiotes chez *Bemisia tabaci***

Dans les populations chinoises de *Bemisia tabaci* MED-Q1 étudiées, *P. aleyrodidarum* et *H. defensa* résident dans les mêmes bactériocytes. Le premier est un endosymbiote obligatoire, certainement impliqué dans la synthèse d'acides aminés. Le second est un endosymbiote *a priori* facultatif, mais présentant des prévalences importantes au sein des populations, puisqu'on le retrouve pratiquement fixé. Quel peut être le phénotype de cette bactérie qui justifie cette quasi-fixation ?

L'hypothèse que nous avons retenue est que cet endosymbiote pourrait lui aussi fournir des métabolites importants pour l'holobionte. En effet, les symbiotes primaires ont généralement des voies métaboliques incomplètes, qui suggèrent que l'hôte, ou d'autres partenaires, puissent jouer un rôle dans la synthèse des nutriments concernés (Hansen et Moran, 2014). De plus, la colocalisation de ces différents endosymbiotes de *B. tabaci* dans les bactériocytes primaires pourrait être le « catalyseur » de ces échanges métaboliques, puisqu'elle favoriserait les interactions entre les différents partenaires. *P. aleyrodidarum* possède-t-elle l'ensemble des gènes impliqués dans la synthèse d'acides aminés essentiels ? Ou bien, est-ce qu'une forme de complémentarité existe entre *H. defensa* et le symbiote primaire dans la synthèse de ces métabolites, qui justifierait la fixation de notre symbiote présumé secondaire ? À partir de l'assemblage des génomes des deux partenaires, nous avons opté pour une reconstruction de leurs capacités métaboliques, et ce afin de déterminer les capacités supposées de chaque endosymbiote, puis de dégager de possibles phénomènes de complémentarité.

### **Problématique Seconde. *Hamiltonella defensa* : deux clades d'hôtes, deux phénotypes ?**

D'après notre première analyse, nous avons conclu à un rôle primordial, sinon obligatoire, d'*H. defensa* chez cet aleurode, qui serait en passe de devenir un endosymbiote primaire pour celui-ci. Comment ce phénotype, cette transition écologique, se répercute-t-elle dans le génome de la bactérie ?

Chez le puceron, *H. defensa* est un symbiote protecteur contre les parasitoïdes. Loin d'être fixé dans les populations, il inflige un coût important à son hôte, et présente des prévalences moyennes. De plus, ce symbiote est un électron libre dans le corps du puceron, tandis qu'*H. defensa* chez l'aleurode demeure cloisonné dans les bactériocytes de l'aleurode.

Aussi aurions-nous à faire à un clade ayant établi deux associations n'ayant pas grand-chose en commun, dans deux clades d'insectes distincts. Nous nous sommes intéressés à leur contenu génique en relation avec les particularités de leur écologie. Nous avons déjà évoqué le génome de la souche 5AT infectant le puceron *Acyrtosiphon pisum*, séquencé par Degnan et ses collaborateurs (2009). Ce symbiote dispose de nombreux systèmes qui pourraient lui permettre de faire face aux variations environnementales, et de protéger son hôte contre ses ennemis (systèmes de sécrétion, toxines, systèmes à deux composantes, régulateurs etc.).

Comment a évolué le contenu du génome d'*H. defensa* chez l'aleurode sous l'hypothèse d'un phénotype nutritif ? La bactérie présente-t-elle le même arsenal que sa cousine ? Nous nous attendons à ce que le tropisme tissulaire réduit d'*H. defensa* chez *B. tabaci*, ainsi que son phénotype non-protecteur, rendent dispensables ces différents systèmes et s'en ressentent dans l'évolution de son génome.

### **Problématique Troisième. *Hamiltonella defensa* et *Regiella insecticola*, des taux de mutations différents.**

Notre troisième problématique va nous permettre de replacer notre espèce d'étude dans un contexte évolutif plus large. Ainsi, *H. defensa* et *R. insecticola* sont deux endosymbiotes formant un clade monophylétique, tous deux étant des symbiotes connus pour leur phénotype protecteur apparentés aux pathogènes du genre *Yersinia*. Pourtant, lors d'une analyse préliminaire, nous nous sommes aperçus qu'*H. defensa* était plus distante – à l'échelle protéique – de *Yersinia* que *R. insecticola*.

Nous avons formulé différentes hypothèses afin de comprendre ces observations, qui ne sont pas forcément mutuellement exclusives. Tout d'abord, les deux lignées de *H. defensa* et *R. insecticola* pourraient avoir connu une transition indépendante vers l'endosymbiose. En ce cas, *H. defensa* représenterait un clade plus ancien d'endosymbiotes.

Si l'ancêtre commun aux deux bactéries était bien un endosymbiote, il faudrait chercher ailleurs une explication à ces taux de substitutions apparemment importants chez *H. defensa*. Les taux de mutations sont-ils supérieurs dans les génomes d'*H. defensa* ? Cela pourrait-il être le fait de gènes de la réparation de l'ADN qui aurait été perdus ? L'écologie des endosymbiotes peut également être en cause. Si *H. defensa* présente – ou a présenté pendant des millions d'années – des populations à la taille efficace plus faible que celles de *R. insecticola*, l'on s'attendrait à ce qu'elles aient subi une plus forte action de la dérive, qui aurait conduit à un taux de **fixation** des mutations plus important.

Il y a donc de nombreuses explications possibles à ces différentiels en termes de taux d'évolution des endosymbiotes. Nous avons pensé que la génomique permettrait, sinon de trancher entre ces différentes hypothèses, au moins de discriminer les plus plausibles. Nous avons donc mené une analyse de génomique comparative, qui devrait nous permettre de mieux comprendre comment l'écologie des endosymbiotes influe sur l'évolution de leur génome, mais également de résoudre la question de l'ancêtre du clade *Hamiltonella-Regiella*.





# ARTICLE PREMIER

---

## Réduction génomique et complémentations métaboliques potentielles entre deux endosymbiotes de l'aleurode *Bemisia tabaci*

---

### Contexte

Durant mes quatre années de Doctorat, différentes données génomiques concernant le symbiote obligatoire *Portiera aleyrodidarum* ont été obtenues. Dans le cadre de notre collaboration avec l'Équipe du Pr. Amparo Latorre, nous avons assemblé le génome d'un symbiote primaire infectant des populations espagnoles de l'espèce *B. tabaci* MED-Q1 (**Annexe 1**). Dans le même temps, différents génomes de *P. aleyrodidarum* ont été séquencés chez les espèces MED et MEAM1 de *B. tabaci* (Sloan et Moran, 2012b ; Jiang *et al.*, 2013b).

Ces génomes possèdent des caractéristiques identiques, les quelques différences relevées provenant essentiellement de seuils de détection de gènes et d'annotation des pseudogènes propres aux différentes équipes ayant procédé à leur analyse. Ainsi, *P. aleyrodidarum* chez *B. tabaci* MED-MEAM1 présente-il un génome typique de symbiote primaire à bien des égards, que ce soit en terme de taille, de contenu génique et d'enrichissement en bases A et T. Il se distingue toutefois des autres symbiotes obligatoires par une fraction importante d'ADN non codant, qui représente tout de même 20% du génome et génère de nombreux événements de recombinaison (Sloan et Moran, 2013).

S'il semble capable de synthétiser différents acides aminés essentiels (thréonine, tryptophane), d'autres voies de synthèse sont incomplètes (histidine, lysine, leucine, isoleucine, valine), ce qui suggère des complémentations métaboliques avec l'hôte, voire avec d'autres partenaires. Or, dans ces espèces MED et MEAM1, certains endosymbiotes a priori facultatifs tels que *H. defensa*, présentent des prévalences anormales, qui laissent entendre que la contribution de ces bactéries à leur hôte pourrait être de grande importance. Ce bénéfice pourrait-il être d'ordre nutritif ? *H. defensa* et *Arsenophonus* sp.

pourraient-ils être impliqués dans la synthèse de nutriments importants dans leur hôte respectif ?

Cela fait sens au regard de la localisation de ces endosymbiotes secondaires chez *B.tabaci*, puisqu'on ne les retrouve que dans les bactériocytes primaires, qui constituent le lieu où s'effectuent les échanges métaboliques entre le symbiote primaire et l'hôte. La proximité entre ces endosymbiotes à fortes prévalences d'une part, et l'endosymbiote primaire d'autre part, devrait favoriser les interactions, notamment d'ordre métabolique.

### **Attentes et objectifs**

Dans ce premier article, nous nous sommes intéressés aux complémentations métaboliques qui peuvent s'exercer entre deux partenaires nichant dans les bactériocytes des populations chinoises de *Bemisia tabaci* MED-Q1 : le symbiote primaire *Portiera aleyrodidarum* et le symbiote secondaire *Hamiltonella defensa*.

Nos objectifs étaient multiples. Tout d'abord, il s'agissait de déterminer les capacités métaboliques caractérisant chaque partenaire pris indépendamment. Il est en effet possible que le symbiote secondaire synthétise des nutriments originaux, qui expliquent son maintien dans les populations. Cela est d'autant plus probable que les symbiotes primaires possèdent des capacités synthétiques réduites, généralement dédiées à la production d'acides aminés essentiels. Notamment, les premiers génomes séquencés de *Portiera aleyrodidarum* dans les espèces MED et MEAM1 ont révélé des « lacunes » dans la synthèse des différents cofacteurs et vitamines.

Notre second objectif était de déterminer qui de l'hôte ou d'*H. defensa* pouvait compléter les voies de synthèse correspondantes. Cette question est cruciale quant au rôle d'*H. defensa* à l'intérieur de l'holobionte. En effet, dans l'éventualité où cette bactérie serait concernée par la synthèse de métabolites essentiels, son statut de symbiote « secondaire » pourrait être remis en question. Sa fixation dans les populations d'insecte justifie pleinement que cette question soit posée.

---

# Genome reduction and potential metabolic complementation of the dual endosymbionts in the whitefly *Bemisia tabaci*

---

Qiong Rao, Pierre-Antoine Rollat-Farnier, Dan-Tong Zhu, Diego Santos-Garcia, Francisco J. Silva, Andrés Moya, Amparo Latorre, Cecilia C. Klein, Fabrice Vavre, Marie-France Sagot, Shu-Sheng Liu, Laurence Mouton et Xiao-Wei Wang

Cet article a été soumis à *BMC Genomics*

## ABSTRACT

**Background** The whitefly *Bemisia tabaci* is an important agricultural pest with global distribution. This phloem-sap feeder harbors a primary symbiont, “*Candidatus Portiera aleyrodidarum*”, that compensate for the deficient nutritional composition of its food sources, and a variety of secondary symbionts. Interestingly, all these secondary symbionts are found in co-localization with the primary symbiont within the same bacteriocytes, which should favor the evolution of strong interactions between symbionts.

**Results** In this paper, we analyze the genome sequences of the primary symbiont *Portiera* and of the secondary symbiont *Hamiltonella defensa* in the *B. tabaci* Mediterranean (MED) species in order to gain insight into the metabolic role of each symbiont in the biology of their host. The genome sequences of the uncultured symbionts *Portiera* and *Hamiltonella* were obtained from one single bacteriocyte of MED *B. tabaci*. As already reported, the genome of *Portiera* is highly reduced (357 kb), but has kept a number of genes encoding most essential amino-acids and carotenoids. On the other hand, *Portiera* lacks almost all the genes involved in the synthesis of vitamins and cofactors. Moreover, some pathways are incomplete, notably those involved in the synthesis of some essential amino-acids. Interestingly, the genome of *Hamiltonella* revealed that this secondary symbiont can not only provide vitamins and cofactors, but also complete the missing steps of some of the pathways of *Portiera*. In addition, some critical amino-acid biosynthetic genes are missing in the two symbiotic genomes, but analysis of whitefly transcriptome suggests that the missing steps may be performed by the whitefly itself or its microbiota.

**Conclusions** These data suggest that *Portiera* and *Hamiltonella* are not only complementary but could also be mutually dependent to provide a full complement of nutrients to their host. Altogether, these results illustrate how functional redundancies can lead to gene losses in the genomes of the different symbiotic partners, reinforcing their inter-dependency.

**Keywords** Endosymbiont; genome; *Hamiltonella*; *Portiera*; whitefly; metabolic complementation

## I. Background

Interactions between endosymbionts and insects are widespread in nature (Moran *et al.*, 2008; Moya *et al.*, 2008), and are particularly important in members of the suborder Sternorrhyncha such as whiteflies, aphids, psyllids and mealybugs. The ability of these insects to use an unbalanced food source such as plant sap is allowed by the interaction with primary bacterial endosymbionts that complement the diet by providing their host with essential amino acids (Douglas, 1998). These primary endosymbionts are confined in specialized host cells, the bacteriocytes, and are strictly vertically transmitted (Thao & Baumann, 2004), leading to a parallel evolution with their insect hosts for millions of years (Wilson *et al.*, 2010; Husnik *et al.*, 2013; Sloan *et al.*, 2014). One evolutionary consequence of this lifestyle is an extreme genome reduction and degradation compared to free living relatives. This convergent evolution has been observed in primary symbionts of many insects, particularly in members of the suborder Sternorrhyncha, such as *Buchnera aphidicola* in aphids (from 422 to 655 kb) (Shigenobu *et al.*, 2000; Perez-Brocal *et al.*, 2006), *Carsonella ruddii* in psyllids (158-166 kb) (Sloan & Moran, 2012), *Portiera aleyrodidarum* in whiteflies (281-358 kb) (Sloan & Moran, 2012; Sloan & Moran, 2013; Santos-Garcia *et al.*, 2012), as well as *Moranella endobia* (538 kb) and *Tremblaya princeps* (139-171 kb) in mealybugs (McCutcheon & von Dohlen, 2011; Husnik *et al.*, 2013; Lopez-Madrigal *et al.*, 2012).

Besides these primary symbionts, insects often harbor secondary symbionts that are not required for their host's survival or reproduction. Nonetheless, they can have broad and important effects on the host biology and ecology, ranging from mutualism to reproductive manipulation, which allow them to spread and be maintained in host populations (Zchori-Fein & Bourtzis, 2011). These secondary symbionts are both vertically and horizontally transmitted (Sintupachee *et al.*, 2006; Chiel *et al.*, 2009) and inhabit a variety of tissues, including bacteriocytes. As they are intracellular, secondary symbionts also show a pattern of genome reduction but to a lesser extent than primary symbionts. Part of their genome is devoted to the biosynthesis of vitamins and cofactors. For example, *Hamiltonella defensa* in aphids is able to supply all the essential vitamins except for thiamine (B1) and pantothenate (B5) (Degnan *et al.*, 2009).

Interestingly, recent advances have shown that these secondary or co-resident symbionts can complement the metabolic network of the primary symbionts. For example, the pair of endosymbionts *Carsonella eucalypti* (primary symbiont) and *Heteropsylla cubana* (secondary symbiont) in psyllids exhibits strict complementarity in the biosynthesis of tryptophan (Sloan & Moran, 2008), as is also the case in the aphid *Cinara cedri* where *Serratia* complements lineage specific gene losses of the primary endosymbiont *Buchnera* (Lamelas *et al.*, 2011). Similarly, the primary endosymbiont *Sulcia muelleri* supplies most of the essential amino acids to its hosts while the remaining ones are provided by different co-resident symbionts: *Baumannia cicadellinicola* in sharpshooters, *Hodgkinia cicadicola* in cicadas, *Zinderia insecticola* in spittlebugs and *Nasuia deltocephalinicola* in leafhoppers (Wu *et al.*, 2006; McCutcheon *et al.*, 2009; McCutcheon & Moran, 2010; Bennett & Moran, 2013). Furthermore, *Candidatus Moranella endobia* and *Candidatus Tremblaya princeps* contribute to intermediate steps of the pathways for the synthesis of amino acids (McCutcheon & von Dohlen, 2011). These insects must thus be seen as holobionts where the full community of organisms inhabiting a host must be taken into account to understand its phenotype, including its metabolic capabilities (Ferrari & Vavre, 2011). Interestingly, comparative genomics has revealed that these losses are generally specific to some lineages of primary endosymbionts, suggesting that the ancestor was performing the lost function. The most probable evolutionary scenario is as follows. Co-infection between the primary and the secondary symbiont preceded the function loss, generating a redundancy of some metabolic functions. As selection acts at the level of the holobiont, any loss of a metabolic function in only one symbiotic partner is neutral (Vautrin & Vavre, 2009; Vautrin *et al.*, 2008). Such losses in essential functions might thus lead to the observed pattern of complementation and to an inter-dependency between the symbiotic partners.

One interesting case for investigating metabolic complementation within symbiotic communities is the whitefly *Bemisia tabaci* Gennadius (Homoptera: Aleyrodidae). This whitefly is a complex of at least 35 cryptic species that differ in many ecological respects, including their potential for causing damage. Some of them are serious destructive pests of agricultural, horticultural and ornamental crops with worldwide distribution (Dinsdale *et al.*, 2010; De Barro *et al.*, 2011; Hu *et al.*, 2011). This phloem-feeding insect harbors the primary

bacterial symbiont "*Candidatus Portiera aleyrodidarum*" (Oceanospirillales, referred as "*Portiera*" in this study) that is located within bacteriocytes (Thao & Baumann, 2004), as well as a variety of secondary symbionts (Gottlieb *et al.*, 2008). *Portiera* is an ancient symbiont of whiteflies, acquired 100-200 million years ago (Baumann, 2005). The recent sequencing of two genomes of *Portiera* for the Mediterranean (MED, formerly referred to as the 'Q' biotype) species and two for the Middle East Asia Minor I (MEAM1, formerly referred to as the 'B' biotype) species of *B. tabaci* indicate that this symbiont not only synthesizes essential amino acids but could also protect its host against oxidative stress by supplying carotenoids (Sloan & Moran, 2012; Sloan & Moran, 2013; Santos-Garcia *et al.*, 2012). Another remarkable feature of these genomes is their extreme reduction with a size ranging from 281 to 358 kb, which is intermediate between those observed for classical primary symbionts and the most reduced ones. Such reduction opens widely the possibility that secondary symbionts complement some metabolic functions of *Portiera*. In whiteflies, secondary symbionts have been reported to affect the life parameters of their host including virus transmission (McCutcheon & von Dohlen, 2011), resistance to natural enemies like parasitic wasps (Mahadav, 2008), heat stress (Brumin *et al.*, 2011) and insecticide (Ghanim & Kontsedalov, 2009). Until now, at least seven secondary symbionts of whitefly have been reported in *B. tabaci*, including *Hamiltonella defensa* (referred as "*Hamiltonella*" in this study), *Arsenophonus*, *Wolbachia*, *Rickettsia*, *Cardinium*, *Fritschea* and *Hemipteriphilus* (Bing *et al.*, 2013a; Bing *et al.*, 2013b). They have different patterns of localization but all of them share bacteriocytes with *Portiera*. These frequent infections with secondary symbionts in addition to *Portiera* make *B. tabaci* an interesting model to investigate metabolic complementation within symbiotic communities. Moreover, two of them, *Hamiltonella* and *Arsenophonus*, reach almost fixation when present in the insect populations, but they have never yet been found together within the same host individual (Bing *et al.*, 2013). In addition, *Hamiltonella* was recently found to provide fitness advantage under nutritional stress conditions, making it a good candidate for analyzing complementation in *B. tabaci* (Su *et al.*, 2014).

Among the *B. tabaci* species complex, the MEAM1 and the MED species are the most widespread and economically important. Over the last twenty years, they have spread rapidly to more than eighty countries over six continents, highlighting their ability to adapt



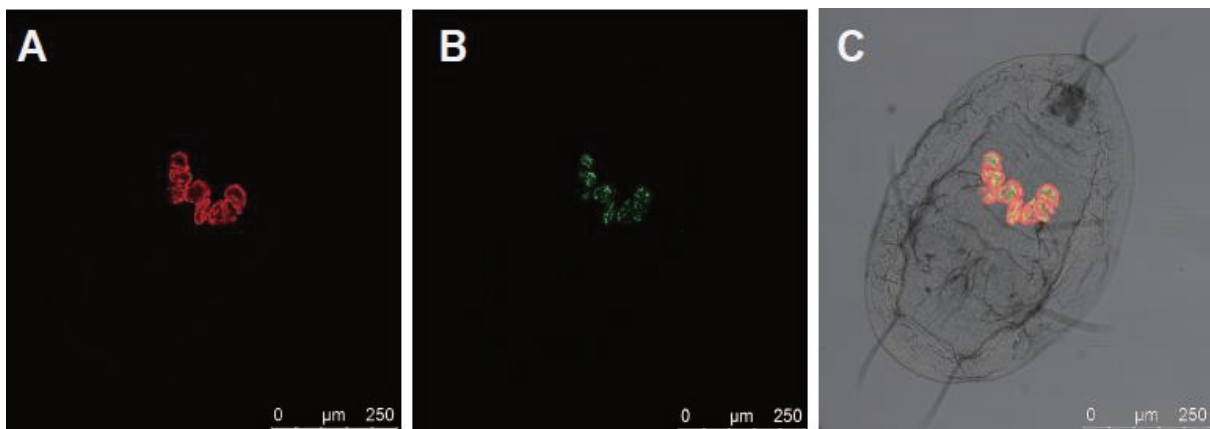
to various environments, and cause serious economic damages worldwide (Hu *et al.*, 2011; Rao *et al.*, 2011). While MEAM1 is notorious for its survivability under extreme conditions, MED is highly resistant to some classes of pesticides, especially neonicotinoids (Rao *et al.*, 2012). Interestingly, *Hamiltonella* has so far only been detected in these two species of *B. tabaci*.

The aim of the present study was to acquire, analyze and study the genome sequences of *Hamiltonella* and *Portiera*. Most often *B. tabaci* individuals harbor more than one secondary symbiont species, which can lead to complex interactions (Gueguen *et al.*, 2010). Here, we used a natural line belonging to the MED species that only harbors *Hamiltonella* in addition to *Portiera*. The data obtained provide insights into the dynamics and evolution of symbiont genomes in intracellular ecosystems, and the mechanisms involved in the interactions among the symbiotic partners, especially in terms of metabolic complementation. They confirm that complementation in whiteflies is an on-going process, and that the co-evolution of different symbiotic partners can lead to inter-dependency.

## II. Results and Discussion

### 1. Isolation of endosymbionts and DNA amplification

PCRs performed using specific primers of *Portiera*, *Hamiltonella*, *Cardinium*, *Wolbachia*, *Rickettsia*, *Arsenophonus* and *Hemipteriphilus* confirmed that the MED line used in this study only harbors *Portiera* and *Hamiltonella* (data not shown). In addition, fluorescence *in situ* hybridization (FISH) revealed that *Hamiltonella* shares bacteriocytes with *Portiera* in this line, and disperses in the cytoplasm of bacteriocytes (**Figure 1**), as already observed (Gottlieb *et al.*, 2008). This contrasts with the situation observed in the aphid *Acyrtosiphon pisum* where *Hamiltonella*, is localized within sheath cells and do not share the same bacteriocytes with *Buchnera* (Sandstrom *et al.*, 2001; Moran *et al.*, 2005). This suggests that the two endosymbionts in MED whiteflies have a more intimate relationship. The endosymbionts were isolated by micromanipulation from a single bacteriocyte cell. Bacterial DNA was amplified by multiple displacement amplification. Diagnostic PCRs were then performed using specific primers to test for the presence of the endosymbiont DNA and the absence of host nuclear DNA contamination.



**Figure 1: Fluorescent *in situ* hybridization (FISH) of *B. tabaci* nymphs using *Portiera* (red) and *Hamiltonella* (green) specific probes. A) *Portiera* channel; B) *Hamiltonella* channel; C) *Portiera* and *Hamiltonella* channels on bright field channel, combined optical sections.**

### 2. Sequencing and general features of the *Portiera* and *Hamiltonella* genomes

Amplified symbiotic DNA was sequenced using the Illumina HiSeq 2000 sequencer and the genomes of both *Portiera* and *Hamiltonella* were independently assembled. The main features for these assemblies are summarized in **Table 1**. The genome of *Portiera* was assembled into a single contig. It has the classical characteristics of primary endosymbionts: a biased G+C content of 26.1% with a reduced genome (357,461 base pairs), but which does

not reach the smaller sizes reported to date, such as for the genomes of *Nasuia deltocephalinicola* (112 kb), *Tremblaya princeps* (139 kb), *Hodgkinia cicadicola* (144 kb), *Carsonella ruddii* (160 kb), *Zinderia insecticola* (209 kb) and *Sulcia muelleri* (246 kb), *Uzinura diaspidicola* (263 kb) (**Table 2**) (McCutcheon & von Dohlen, 2011; Zchori-Fein & Bourtzis, 2011; Wu *et al.*, 2006; McCutcheon & Moran, 2010; Nakabachi *et al.*, 2006; McCutcheon *et al.*, 2009; Sabree *et al.*, 2013). Despite its reduced size, this genome has a relatively low percentage of coding DNA (67.4%), which is much less than the coding density of *Buchnera* in aphids (83%). Large intergenic regions have been shown to represent an important substrate for genome rearrangements in *Portiera* (Sloan & Moran, 2013). A total of 272 genes were predicted in the *Portiera* genome, and 84.5% of them (230 genes) have homologs present in GenBank. The genome encodes three rRNA genes (16S, 23S and 5S), two non-coding RNAs (rnpB, tmRNA) and 33 tRNA genes including at least one for each of the 20 amino acids (**Table 3**), as observed in the two other assembled *Portiera* genomes of the MED whitefly species and the two genomes of *Portiera* in the MEAM1 whitefly species (Sloan & Moran, 2012; Santos-Garcia *et al.*, 2012; Jiang *et al.*, 2012). The gene content is the same in all deposited genomes, the differences being only due to variations in the annotations and pseudogene detection parameters.

	<i>Portiera</i>	<i>Hamiltonella</i>
Total number of scaffold/contig	1	92
Current genome assembly (bp)	357,461	1,800,792
Average length (bp)	-	19,574
N50 length (bp)	-	102,662
N90 length (bp)	-	15,36
Maximum contig length (bp)	-	214,721
Predicted genes	272	1,884
Assigned function genes	230	1,672
Gene average length	889	806
Coding density, %	67.4	84.4
tRNA genes	33	40
rRNA genes	3 (5S, 16S, 23S)	3 (5S, 16S, 23S)
GC content, %	26.12	40.49

**Table 1: General statistics and features of *Portiera* and *Hamiltonella* genomes from the MED whitefly**

The draft genome of *Hamiltonella* in this MED line has an approximate size of 1,800,792 bp with an average G+C content of 40.49% and comprises 92 large scaffolds (**Table 1**) with a N50 size of approximately 102.66 kb. The genome encodes a total of 1,884 putative CDSs for which 88.75% (1,672) can be assigned to a putative biological function. A total of 40 tRNA genes including at least one for all of the 20 amino acids and five small RNAs were predicted (**Table 3**). The average coding gene identity with *Hamiltonella defensa* Hd-Ap is very high (96.57%). Globally, the general features of this genome are very similar to those of other genomes of secondary symbionts.

	Chromosome size (bp)	G+C (%)	No. of CDS	Coding density (%)	Avg. CDS length
<b><math>\alpha</math>-proteobacteria</b>					
<i>Rickettsia belli</i>	1,522,076	31.6	1,429	85.2	908
<i>Wolbachia pipientis</i>	1,267,782	35.2	1,195	80.2	851
<i>Hodgkinia cicadicola</i>	143,795	58.4	169	91.3	777
<b><math>\beta</math>-proteobacteria</b>					
<i>Profftella armatur</i>	459,399	24.2	366	88.0	1,104
<i>Zinderia insecticola</i>	208,564	13.5	202	89.7	926
<i>Tremblaya princeps</i>	138,927	58.8	121	66.1	759
<i>Nasuia deltocephalinicola</i>	112,091	17.1	137	91.9	752
<b><math>\gamma</math>-proteobacteria</b>					
<i>Escherichia coli</i>	4,641,652	50.8	4,14	85.1	954
<i>Sodalis glossinidius</i>	4,171,146	54.7	2,432	50.9	873
<i>Arsenophonus nasoniae</i>	3,567,128	37.4	3,332	-	-
<i>Hamiltonella defensa</i>	2,110,331	40.1	2,094	80.4	810
	1,843,969	40.3	1,806	76.8	784
<i>Serratia symbiotica</i>	1,762,765	29.2	672	38.8	1,019
<i>Blochmannia pennsylvanicus</i>	791,654	29.6	610	76.7	995
<i>Blochmannia floridanus</i>	705,557	27.4	583	83.2	1,007
<i>Baumannia cicadellincola</i>	686,194	33.2	595	85.5	986
<i>Buchnera aphidicola</i>	640,681	26.3	564	87.0	988
	416,38	20.2	357	85.1	992
<i>Portiera aleyrodidarum</i>	357,461	26.1	272	67.7	890
	357,472	26.1	246	67.7	984
	358,242	26.2	256	67.6	945
	280,663	24.7	269	94.3	984
<i>Carsonella ruddii</i>	159,662	16.6	182	94.1	826
<b>Bacteroidetes</b>					
<i>Uzinura diaspidicola</i>	263,431	30.2	227	86.5	1,004
<i>Sulcia muelleri</i>	245,53	22.4	227	92.1	996

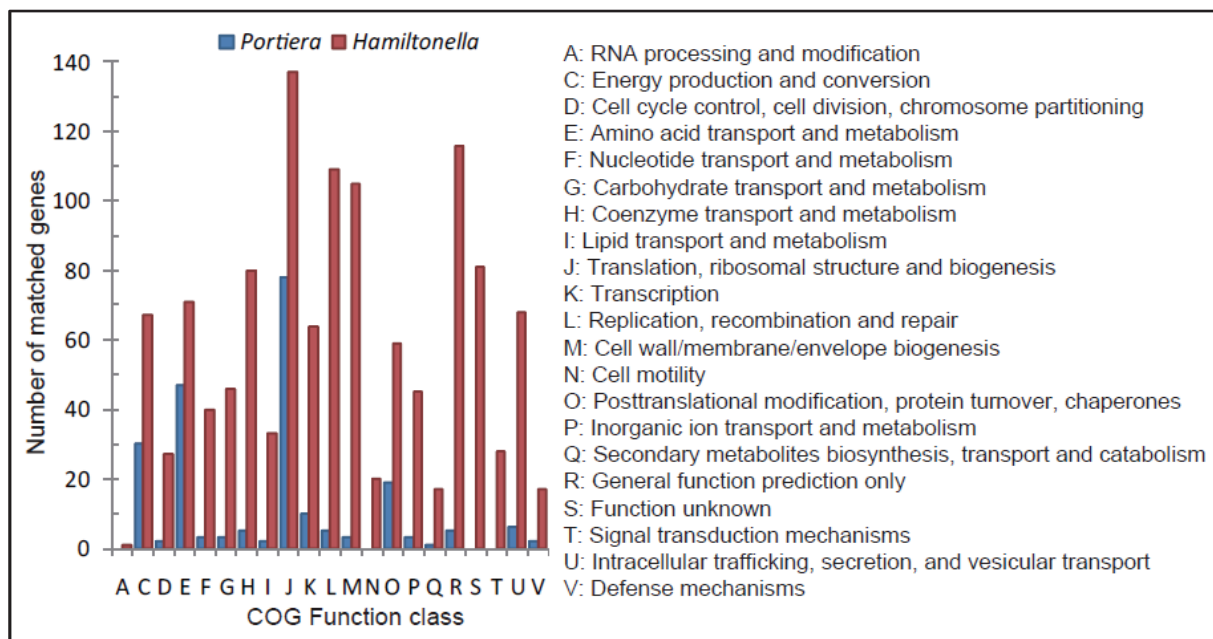
**Table 2: General genomic properties of representative free-living bacteria and insect symbionts**

<b>tRNA Type</b>	<b><i>Portiera</i> Anti-codon</b>	<b><i>Hamiltonella</i> Anti-codon</b>
Ala	TGC	TGC/GGC
Arg	TCT/ACG/CCG	TCT/CCT/CCG/ACG/TCT
Asn	GTT	GTT
Asp	GTC	GTC
Cys	GCA	GCA
Gln	TTG	CTG/TTG
Glu	TTC	TTC
Gly	GCC/TCC	TCC/GCC
His	GTG	GTG
Ile	GAT	GAT
kIle	CAT	CAT
Leu	TAG/GAG/TAA	CAG/GAG/TAG/TAA/CAA
Lys	TTT	TTT/CTT
iMet	CAT	CAT
Met	CAT	CAT
Phe	GAA	GAA
Pro	TGG	TGG/GGG
Ser	TGA/GCT/GGA/ CGA	GCT/GGA/TGA
Thr	CGT/TGT/AGT	TGT/GGT
Trp	CCA	CCA
Tyr	GTA	GTA
Val	GAC/TAC	TAC/GAC

**Table 3: tRNA type of *Portiera* and *Hamiltonella* from the MED whitefly**

### **3. Functional annotation of the genomes of *Portiera* and *Hamiltonella* and metabolic reconstruction**

Protein-coding genes of *Portiera* and *Hamiltonella* were classified into Clusters of Orthologous Groups (COGs) (Tatusov *et al.*, 1997). A total of 209 *Portiera* genes and 1,119 *Hamiltonella* genes were assigned to COGs (**Figure 2**). The three most prominently represented COG categories in the *Portiera* genome are “Translation, ribosomal structure and biogenesis”, “Amino acid transport and metabolism”, and “Energy production and conversion” (COGs J, E, and C, respectively). Conversely, genes related to “RNA processing and modification”, “Chromatin structure and dynamics”, “Cell motility and Signal transduction mechanisms” (COGs A, B, N, and T, respectively) are absent from *Portiera*. In *Hamiltonella*, several COG categories are prominently represented, namely “Translation ribosomal structure and biogenesis”, “Replication, recombination and repair”, and “Cell wall/membrane/envelope biogenesis” (COGs J, L, and M, respectively) (**Figure 2**).



**Figure 2: COG-based characterization of all proteins with annotated functions in *Portiera* and *Hamiltonella*.**

The metabolic networks of the two symbionts were reconstructed and used to identify the sources needed by each symbiont (Tables 4 and 5). *Portiera* needs 11 different sources to produce less than 180 metabolites, while *Hamiltonella* needs 15 sources for less than 600 metabolites (Pearson test,  $p$ -value < 0.05), which conforms with the expectation of a more reduced and more host-dependent metabolism in this primary endosymbiont than in *Hamiltonella*. While most of these sources might be provided by the host, the metabolism of *Hamiltonella* needs some sources which would be only synthesized by *Portiera*. AroA is a 3-phosphoshikimate-1-carboxyvinyltransferase involved in the sixth step of the chorismate pathway, leading to the production of 5-enolpyruvyl-shikimate-3-phosphate. A mutation in the *aroA* gene, determined by the *in silico* analysis, and later confirmed by Sanger sequencing on independent samples, makes 5-enolpyruvyl-shikimate-3-phosphate a source for the *Hamiltonella* metabolism. *Portiera* can produce this source and probably export it to the secondary endosymbiont. These results suggest that the two symbionts not only depend on their host for a complete metabolic functioning, but also on their bacterial symbiotic partner, and that this evolution of a mutual dependence is an ongoing process as *aroA* is intact in the *Hamiltonella* from *A. pisum* (Degnan *et al.*, 2009).

Inputs	Class	Putative source
(Seleno-)Homocysteine	Amino acid	<i>B. tabaci</i>
Geranylgeranyl diphosphate	Isoprenoid	<i>B. tabaci</i>
HCO <sub>3</sub>	Anion	<i>B. tabaci</i>
L-Ornithine	Amino acid	<i>B. tabaci</i>
L-Aspartate	Amino acid	<i>B. tabaci</i> and <i>Hamiltonella</i>
Erythrose-4-phosphate	Sugar Phosphate	<i>B. tabaci</i> and <i>Hamiltonella</i>
Phosphoenol-pyruvate	Carboxylic acid	<i>B. tabaci</i> and <i>Hamiltonella</i>
Farnesyl-Diphosphate	Isoprenoid	<i>B. tabaci</i> and <i>Hamiltonella</i>
Ribose-5-phosphate	Sugar Phosphate	<i>B. tabaci</i> and <i>Hamiltonella</i>
Protoheme	Heme	Mitochondria

**Table 4: Necessary sources for the metabolism of *Portiera*.** The sources potentially provided by *Hamiltonella* have been assessed from our analyses. Most of the sources not produced by the bacterial partner are classic metabolites of eukaryotes, and the corresponding biosynthetic pathways in MetaCyc are referred. Some sources do not seem to be produced by eukaryotes, but previous works on phloemophagous insects proposed that they were acquired from the host or its diet. The same assumptions have been made in this study, and the corresponding references have been indicated in the table.

Among sources provided by the host, hydrogen sulfide is of particular interest. Indeed, while the two symbionts of the MED whitefly lack the whole gene sets for sulfur metabolism (**Figure 3**), *cysN/cysC* [bifunctional enzyme EC:2.7.7.4, 2.7.1.25], *cysQ* [EC:3.1.3.7] and *cysJ* [EC:1.8.1.2] of this pathway have been identified in the whitefly transcriptome (Wang *et al.*, 2010). It suggests that the host whitefly or its gut microbiota might help the symbionts for sulfate reduction. Notably, the sulfate reduction pathway can transform sulfate into hydrogen sulfide, a precursor for the cysteine synthesis pathway present in *Hamiltonella* (**Table 5**).

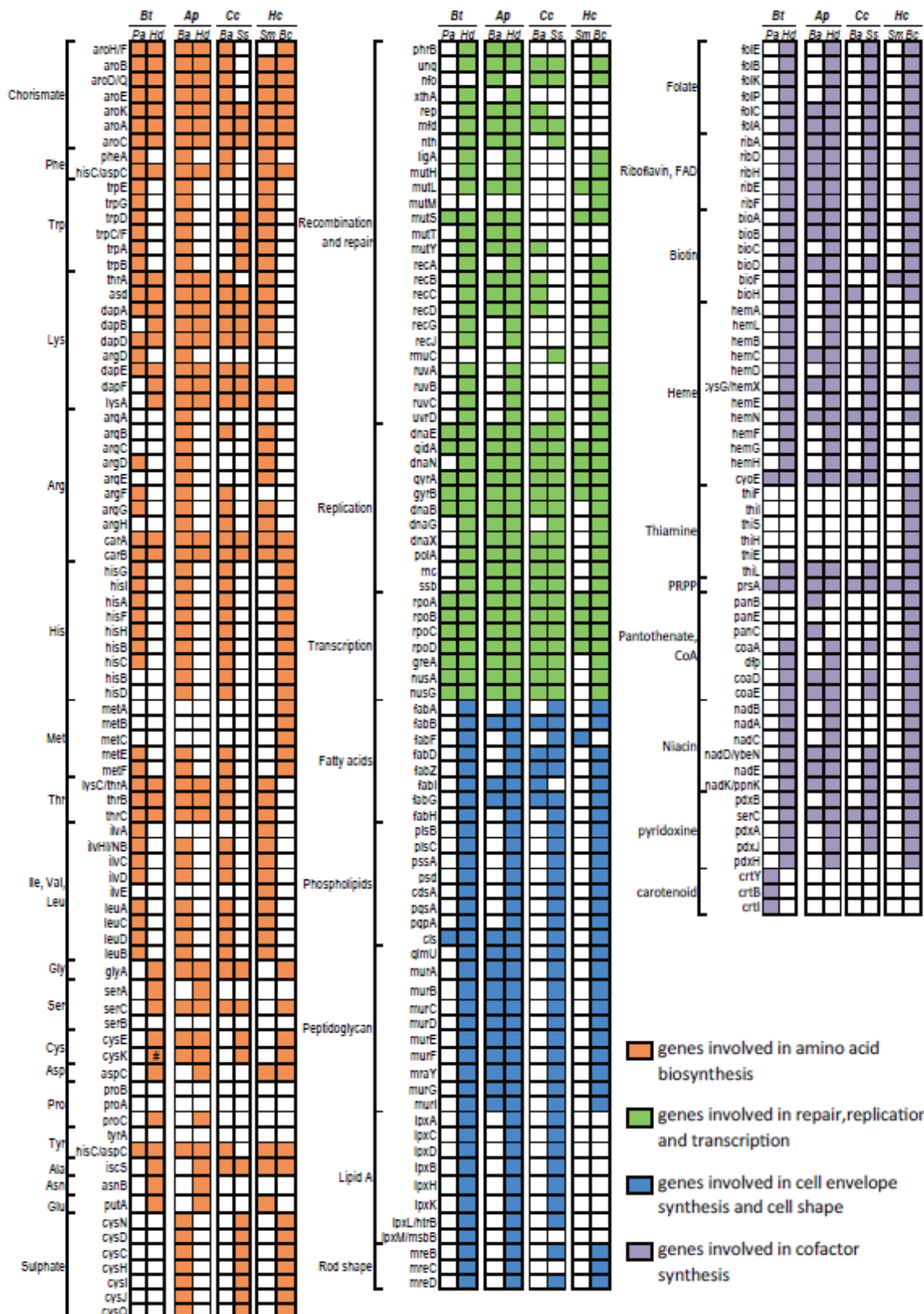
Metabolite	Product/Pathway	Putative source
H <sub>2</sub> S (S <sub>2</sub> O <sub>3</sub> )	(Seleno)Cysteine	<i>B. tabaci</i>
Pantothenate	Co-enzyme A	<i>B. tabaci</i>
HCO <sub>3</sub>	Fatty acids, nucleic acids etc.	<i>B. tabaci</i>
Dihydroneopterin	Folate	<i>B. tabaci</i>
P-Amino-Benzoate	Folate	<i>B. tabaci</i>
Glucose	Glucolysis	<i>B. tabaci</i>
Serine	Glycine, Cysteine etc.	<i>B. tabaci</i>
Fe <sup>2+</sup>	Heme o, general cofactor	<i>B. tabaci</i>
Proline	Glutamate	<i>B. tabaci</i>
E4P	Pyridoxine	<i>B. tabaci</i>
SAM	Methionine	<i>B. tabaci</i>
Protoporphyrin	Heme o	Mitochondria
5-ES-3P	Chorismate	<i>Portiera</i>
N-S-LL-2,6-D	Lysine	<i>Portiera</i>
Phenyl-pyruvate	Phenylalanine	<i>Portiera</i>

**Table 5: Necessary sources for the metabolism of *Hamiltonella*.** The sources potentially provided by *Portiera* have been assessed from our analyses. Most of the sources not produced by the bacterial partner are classic metabolites of eukaryotes, and the corresponding biosynthetic pathways in MetaCyc are referred. Some sources do not seem to be produced by eukaryotes, but previous works on phloemophagous insects proposed that they were acquired from the host or its diet. The same assumptions have been made in this study, and the corresponding references have been indicated in the table. Abbreviations: D-erythrose-4-phosphate (E4P); S-adenosyl methionine (SAM); 5-enolpyruvyl-shikimate-3-phosphate (5-ES-3P); N-succinyl-L,L-2,6-diaminopimelate (N-S-LL-2,6-D).

#### 4. Essential amino acid synthesis depends on within-pathway complementation between symbionts

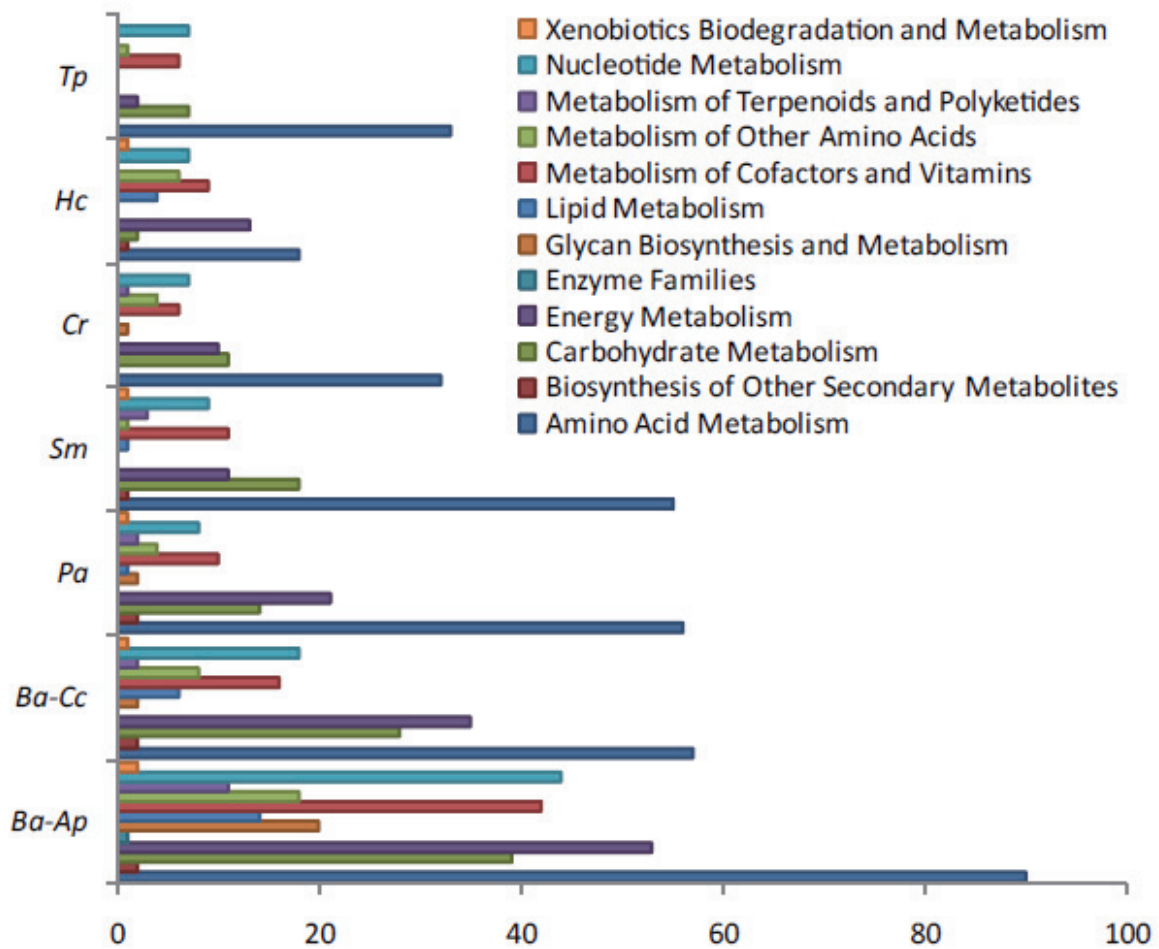
A major metabolic contribution of symbionts in phloemophagous insects concerns the synthesis of essential amino acids. As highlighted by the COG analysis, some metabolic functions related to the biosynthesis of amino-acids have been retained in the highly reduced genome of *Portiera*, which contains 56 genes dedicated to this function (**Figure 4**). *Portiera* is therefore able to synthesize or to participate in the synthesis of several amino acids, and especially essential ones, like other primary endosymbionts (**Figure 4**). These results further demonstrate the evolutionary convergence between symbionts from distinct phylogenetic lineages hosted by hosts with similar diets (Lopez-Sanchez *et al.*, 2009).





**Figure 3: Gene content of reduced genomes from symbionts.** Abbreviations : *Bt*- *Bemisia tabaci*, *Ap*- *Acyrtosiphon pisum*, *Cc*- *Cinara cedri*, *Hc*- *Homalodisca coagulata*; *Pa*- *Portiera aleyrodidarum*, *Hd*- *Hamiltonella defensa*, *Ba*- *Buchnera aphidicola*, *Ss*- *Serratia symbiotica*, *Sm*- *Sulcia muelleri*, *Bc*- *Baumannia cicadellinicola*.

However, *Portiera* has only two complete pathways for the synthesis of essential amino-acids (Sloan & Moran, 2012; Santos-Garcia *et al.*, 2012). Interestingly, some of the missing genes in these pathways are present in the *Hamiltonella* genome (**Figure 3**). For example, *aspC*, a gene encoding an enzyme required for the phenylalanine synthesis, is missing in the genome of *Portiera* but is encoded by *Hamiltonella*. Furthermore, the genome of *Portiera* lacks three enzymes involved in the biosynthesis of lysine, namely *dapF*, *lysA*, which are absent, and *dapB* which is pseudogenized, while *Hamiltonella* possesses all of these genes but lacks the *argD* gene in the same pathway (**Figure 3**). As for chorismate in *Hamiltonella*, the production of lysine has only been recently lost in *Portiera*, since the entire pathway is present and intact in the *Portiera* of *Trialeurodes vaporariorum*, another member of Aleyrodidae (Sloan & Moran, 2013). This is consistent with the hypothesis that acquisition of secondary endosymbionts generates metabolic redundancies with the primary endosymbiont, leading to specific gene losses in both genomes (Sloan & Moran, 2012; Lamelas *et al.*, 2011). These losses, if they affect essential functions ancestrally performed by the primary endosymbiont, can rapidly make secondary endosymbionts indispensable for the holobiont (Sloan & Moran, 2012; Lamelas *et al.*, 2011). Nevertheless, the functions of *aspC* in *Portiera* and *argD* in *Hamiltonella* could also be performed by other aminotransferases encoded by their respective genomes, an unknown enzyme, or by an already known enzyme which would have changed or extended its prior function, as has already been proposed for other symbiotic genes (Sloan & Moran, 2012; Charles *et al.*, 2011). Interestingly, the lysine (in *Carsonella*, *Sulcia* and some *Buchnera*) and phenylalanine (in *Carsonella*, *Buchnera*, *Tremblaya* and *Ishikawella*) pathways are often incomplete in the genomes of primary endosymbionts, and have been proposed to be in some cases complemented by other partners, hosts or co-symbionts (reviewed in Hansen *et al.*, 2014 and Hansen & Moran, 2014).



**Figure 4: Comparative analysis of genes by metabolism functional KEGG categories from the genomes of *Portiera* and other bacteria.** For abbreviations of symbionts, Tp: *Tremblaya princeps*, Hc: *Hodgkinia cicadicola*, Cr: *Carsonella ruddii*, Sm: *Sulcia muelleri*, Pa: *Portiera aleyrodidarum*, Ba-Cc: *Buchnera aphidicola* (*Cinara cedri*), Ba-Ap: *Buchnera aphidicola* (*Acyrtosiphon pisum*).

The host itself could also provide enzymes for completing some pathways, as has been recently shown in aphids where the host metabolism has been redirected to complement some missing reactions for amino acid synthesis in *Buchnera* (Russell *et al.*, 2013). For example, it has been hypothesized that phloem-feeder hosts encode a homolog of the *aspC* gene allowing the production of phenylalanine (Hansen & Moran, 2014). The same situation might apply to *B. tabaci* MED for the synthesis of valine, leucine and isoleucine, for which *ilvE*, the gene encoding the common last enzyme of these pathways is absent from the genomes of both *Portiera* and *Hamiltonella* and has been found in the transcriptome of the MED whitefly (Wang *et al.*, 2010).

The case of the histidine synthesis pathway is different because the genes which are not present in the genome of *Portiera* (*hisB* [EC 3.1.3.15] and *hisD*; Sloan & Moran, 2012; Santos-Garcia *et al.*, 2012) are neither present in the *Hamiltonella*'s genome (**Figure 3**), nor have been detected in the whitefly's transcriptome (Wang *et al.*, 2010). This suggests that histidine may be provided by the food. This hypothesis is relevant since previous studies indicated that this essential amino acid is present in high concentration in the phloem sap (Douglas, 2006).

Finally, genes involved in the synthesis of non-essential amino acids like serine and proline are missing in the genomes of both endosymbionts. All genes of these pathways (including *serA*, *serC*, *serB*, *proB* and *proA*) can be identified in the sequences obtained from the MED transcriptome (Wang *et al.*, 2010). This highlights the fact that gene losses in the symbiotic genomes can also occur when some functional redundancies exist with the host's genome. The phenomenon was also confirmed in pea aphid (Wilson *et al.*, 2010).

##### **5. *Portiera* and *Hamiltonella* synthesize different vitamins and cofactors**

Other types of metabolites frequently provided by the symbionts in sap-feeding insects are vitamins and cofactors. *Portiera* lacks almost all the genes involved in the synthesis of vitamins and cofactors (**Figure 3**). However, *Portiera* is the first symbiont reported to be able to produce carotenoids (Sloan & Moran, 2012; Santos-Garcia *et al.*, 2012). Interestingly, a thorough BLAST search of the whitefly transcriptome (Wang *et al.*, 2010) suggested that the insect cannot synthesize carotenoids, unlike aphids (Moran & Jarvik, 2010). Conversely, *Hamiltonella* possesses a number of gene sets (76 genes) dedicated to vitamin production, which include riboflavin (vitamin B2), NAD, pyridoxine (vitamin B6), biotin (vitamin B7), folic acid (vitamin B9) (**Figure 3-4**). However, it is unable to produce thiamine (vitamin B1) and pantothenic acid (vitamin B5) and, in the MED transcriptome, we did not find any genes for synthesizing the two missing vitamins, except *panC*, suggesting that whiteflies might capture thiamine and pantothenic acid from the phloem sap.

## 6. Transport capabilities in link with exchanged metabolites

As revealed by previous analyses, the metabolic networks of both endosymbionts could be highly intertwined requiring efficient transport systems of metabolites. Compared to secondary symbionts, primary symbionts lack many genes for environmental information processing, such as membrane transport and signal transduction. In addition, most primary endosymbionts have kept only a few transporter systems although metabolite fluxes are required within the symbiotic system (Ren & Paulsen, 2005). Indeed, the metabolism of *Portiera* needs amino acid, nucleic acid and isoprenoid uptake (**Table 4**). Moreover, the products of the metabolism of *Portiera* (carotenoids, essential amino acids) need also to be exported to the host.

However, the genome of *Portiera* encodes only eleven putative inner-membrane-located transporters. As the gene content of *Portiera* has quite not changed since the last ancestor between the obligate endosymbionts of *B. tabaci* and *T. vaporarium* (Sloan & Moran, 2013), this suggests an ancient loss of transporters. The low number of transporters is of the same order of magnitude as in *Buchnera* Cc (12 transporters), the symbiont of *Cinara cedri*, and less than half of the number of transporters identified in other *Buchnera* genomes (30 to 34) (Charles *et al.*, 2011). Different hypotheses have been proposed to explain the weak number of transporters in *Buchnera* Cc (Charles *et al.*, 2011). First, this could be linked to the strong specialization of the aphid host of this symbiont (only found in cedar plants), but in our case, the large plant spectrum of *B. tabaci* is not consistent with this hypothesis. Second, the presence of general transporters with low affinity with the substrates, or the recruitment of transporters from the host, could compensate the losses. For example, it has been proved that *Buchnera* perceived glutamine from aphid transporter (Price *et al.*, 2014). Finally, the majority of the metabolites in *Buchnera* Cc could be transported by passive diffusion, due to a highly simplified envelop (Charles *et al.*, 2011). This is also totally plausible in *Portiera*, this symbiont exhibiting an atypical cell wall that could be devoid of the external membrane of Gram-negative bacteria (Costa *et al.*, 1993). Consistently, *Portiera* in *B. tabaci* as *Buchnera* in *C. cedri* lost the LspA lipoprotein signal peptidase, which would indicate a probable deficiency in the outer membrane assembly (Charles *et al.*, 2011).

In addition, this reduction of the ancestral transporter stock can be explained by the small number of classes of chemical compounds transported by *Portiera*. Its genome encodes exporters/importers for all of these classes, excepted sugars, but as it is expected in *Buchnera Cc*, *Portiera* probably acquires them by passive diffusion (Perez-Brocal *et al.*, 2006). . For example, it encodes a GltP proton dependent transporter involved in *Escherichia coli* in the uptake of aspartate (Wallace *et al.*, 1990), a probable source of the *Portiera* metabolism. In addition this transporter could also allow the uptake of the other required amino acids (**Table 4**). Another transporter is ArgO, which is normally involved in arginine export (Nandieni *et al.*, 2004). Interestingly arginine is no more synthesized by *Portiera* from *B. tabaci*, and consequently, ArgO could have no more role, or could be involved in the export of other essential amino acids synthesized by the endosymbiont. . Finally, the best hits of BTQP\_154 (>50% of homology) are *ditE*-like genes of *Pseudomonas spp.* The DitE permease could be involved in diterpenoid transport (Martin & Mohn, 2000). Thus, this transporter could be a putative candidate for the uptake of isoprenoid sources (**Table 4**) and the export of carotenoids.

### III. Conclusions

The analysis of the metabolic pathways in *Portiera* and *Hamiltonella* revealed the interdependency between two partners engaged in a mutualistic relationship. First, while the two partners depend principally on sources coming from the host diet and metabolism, they could also rely on some sources they provide to each other. Second, although *Portiera* encodes a large number of genes in the biosynthesis of essential amino-acids, it lacks almost all genes required for the synthesis of cofactors, while the latter are likely to be produced by *Hamiltonella*. Third, and in addition to this complementarity in terms of provided-sources and metabolites, some strict within-pathway complementation cases have been inferred by our *in silico* analysis, which concern two essential amino acids (lysine and phenylalanine). These results suggest that *Hamiltonella* could be an primary partner of the consortium, and could functionally explain the recent results showing the benefit it provides under nutritional stress conditions (Su *et al.*, 2014), and allow its fixation within populations. Interestingly, *Hamiltonella* probably replaced the primary endosymbiont for some functions that are still

assumed by *Portiera* in other species of *B. tabaci*. As *Hamiltonella* is probably a recent symbiont of *Bemisia*, it appears that both symbionts have been caught in the act of a switch towards reciprocal complementation. Thus, concerning these functions, *Hamiltonella* did not allow an extension of the ancestral metabolic ability of the holobiont in the *B. tabaci* MED and MEAM1 species. Emergence of dependence between the symbiotic partners is rather due to metabolic redundancies which have been eliminated, a phenomenon that led to an apparent mutualism. In order to confirm this scenario and gain insights into the evolutionary dynamics and history of these interactions, this analysis should be extended to other populations of MED species associated with *Arsenophonus*, and to other species of *B. tabaci*. This could allow determining when and in which *B. tabaci* species *Portiera* lost some key elements for the holobiont functioning, and whether some secondary endosymbionts only replaced these functions, or widened the initial niche of the holobiont, making it less dependent on specific environmental sources for some metabolites.

## IV. Methods

### 1. Whitefly rearing

The *B. tabaci* line belonging to the MED species (mtCO1 GenBank accession no: DQ473394) was maintained on cotton *Gossypium hirsutum* (Malvaceae) (cv. Zhe-Mian 1793) in climate chambers at  $27 \pm 1$  °C, 14 h : 10 h (light: darkness) and 40-60% relative humidity. The purity of this whitefly line was monitored every 3-5 generations using the RAPD-PCR technique with the primer H<sub>16</sub> (5'-TCTCAGCTGG-3') (De Barro & Driver, 1997). Our previous study confirmed this line only harbors two symbionts, the primary symbiont *Portiera* and the secondary symbiont *Hamiltonella* (Borenstein *et al.*, 2008).

### 2. Fluorescence *in situ* hybridization (FISH)

The symbiont localization was determined by performing FISH on nymphs as previously described (Gottlieb *et al.*, 2008; Gottlieb *et al.*, 2006) with symbiont-specific 16S/23S rRNA DNA probes: BTP1-Cy3 (TGTCAGTGTCAGCCCAGAAG) for *Portiera* and BTH-Cy5 (CCAGATTCCCAGACTTTACTCA) for *Hamiltonella*. The *Portiera*-specific probe was used for all the individuals for control. Stained samples were whole mounted and photographed on a



confocal microscope (Leica). Specificity of detection was confirmed using no probe staining and RNase-digested specimen.

### **3. Purification and amplification of endosymbionts DNA from a single bacteriocyte**

MED whitefly nymphs with paired, roundish and orange color bacteriome were dissected in PBS (GIBCO®, Invitrogen, USA). A single bacteriocyte was collected with a micropipette on a glass slide and symbiont cells were then isolated by Eppendorf® microcapillary using a TransferMan® NK2 micromanipulator with a CellTram Vario (Eppendorf, Germany) under an inverted microscope. To prepare the total DNA of the symbionts, the symbiont cells in PBS were amplified by multiple displacement amplification (MDA) using the Repli-g UltraFast Mini Kit (Qiagen, Germany) according to the manufacturer's protocol, with some modifications (Woyke *et al.*, 2010). The yield of amplified total DNA was about 300 ng per microliter.

### **4. Diagnostic PCR**

The amplified DNA was checked by diagnostic PCR using primers specific for *Portiera* and *Hamiltonella* (Zchori-Fein & Brown, 2002). Contamination by DNA from *B. tabaci* was also checked by performing PCR on  $\beta$ -actin gene and EF1 gene. The  $\beta$ -actin gene primers were Actin-F (GCTGCCTCCACCTCATTAGA) and Actin-R (AGGGCGGTGATTCCTTCT). The PCR parameters for  $\beta$ -actin gene of *B. tabaci* were 4 min at 94 °C, followed by 30 cycles of: 45 s at 94 °C, 45s at 60 °C, 1 min at 72°C, and a final extension of 10 min at 72 °C. The EF1 gene primers were EF-F (CGTCCCCATTCTGGATGGCACGG) and EF-R (CATCTCGACGGACTTAACTTCAGT), and the PCR parameters for EF1 gene are described in Ghanim *et al.* (2007).

### **5. Genome sequencing, assembly and annotation**

Using the amplified endosymbiotic DNA, two paired-end libraries with an average insert size of 200 bp and 2 kb were generated and sequenced respectively, using the Illumina HiSeq 2000 sequencer according to the manufacturer's protocols in Beijing Genomics Institute, BGI (Shenzhen, China). The reads belonging to *Portiera* were retrieved based on previous published *Portiera* genomes and then assembled using SOAPdenovo v1.05. The *Portiera* genome was assembled in a single circular closed contig. The rest of the reads were



used for *Hamiltonella* assembly and the resulted contigs were connected according to the 200 bp and 2 kb mate-paired relationships according to the instruction of SOAPdenovo v1.05. This assembly generated a draft genome of *Hamiltonella*. It has an approximate size of 1,800,792 bp and comprises 92 large scaffolds. CDSs were then predicted by Glimmer v3.0, and a homologous comparison to a nonredundant public database was performed by BLAST for function annotation. Genes were assumed to be pseudogenes if they underwent more than 20% of reduction compared to orthologs following (Lerat & Ochman, 2005). The annotation of COGs, gene ontologies, and EC numbers was performed using SWISSPROT. We used RNAmmer and tRNAscan to identify rRNAs (Lagesen *et al.*, 2007) and tRNAs (Schattner *et al.*, 2005). tRNA genes with anticodon CAT were discriminated according to Silva *et al.* (2006).

## **6. KEGG (Kyoto Encyclopedia of Genes and Genome) analysis**

For functional categorization, the protein sequences of the symbionts were annotated by mapping to KEGG pathways. All genes of *Portiera* and *Hamiltonella* were analyzed based on BBH (bi-directional best hit) by using the KAAS tool (Moriya *et al.*, 2007), which compared the metabolic capacity of the two bacteria with the other primary and secondary symbionts.

## **7. Metabolic complementation analysis**

PathwayTools (Karp *et al.*, 2002) was used to build the metabolic networks of both *Portiera* and *Hamiltonella* on the basis of E.C. numbers and annotations. MetExplore (Cottret *et al.*, 2010) was used to check and curate each metabolic reconstruction individually. Cytoscape was used for network visualisation (Smoot *et al.*, 2011). Each reaction removed or corrected in MetExplore was followed by a strict manual control (see below). The refined networks were used to identify the metabolic sources in the two symbionts, *i.e.* the inputs which permit to produce all the different metabolites. Thus, it was important to avoid the detection of sources with no biological sense. To do this, we removed reactions which were unlikely to take place in these organisms, or which did not interest us in the case of this work. More precisely, the automatic reconstruction generated isolated reactions which do not use any inputs produced by the metabolism of either symbiont. These reactions can correspond to spontaneous reactions that have been placed in the networks because they

do not need enzymes and can thus theoretically occur in all organisms. They can also correspond to enzymes that can theoretically perform different reactions. In that case only the reactions that are connected with the rest of the network were kept. Some reactions can also involve generic metabolites (“a sugar”) and were replaced by the specific metabolites (“glucose”), or removed if these specific reactions were not connected to the rest of the network. Finally, reactions that are not assumed to occur in prokaryotes were also removed. For each removed reaction in the metabolic network of a given endosymbiont, we checked that the concerned metabolites were not a product or a source of the metabolism of the other endosymbiont. Finally, to detect sources, we applied the Borenstein's method (Borenstein *et al.*, 2008) on the two metabolic reconstructions independently, using the Igraph package (R software) (Csardi & Nepusz, 2012). Some sources were due to missing reactions in the pathways. It is the case for lysine or folate biosynthesis in *Hamiltonella*. In order to check for the validity of these sources, the missing genes were searched on the basis of both MetaCyc and KEGG maps. For each missing reaction, we used queries based on Blast, domain identification, and enzyme name/E.C number to identify putative candidates – the last two steps were performed using the MaGe platform.

### **8. *Portiera* transporter analysis**

Analysis of the transporters of *Portiera* was performed as described in Charles *et al.* (2011). At first, Blast was used against the Transport Classification DataBase (TCDB) in order to find a set of potential transporters and to assess their classification according to the TCDB. Then GO annotations of the homologous genes present in UniprotKB were used to identify genes involved in transport. The genes found from these two steps were manually curated. The enzymatic reactions generally need some cofactors to occur. These last are not produced by *Portiera* and have also to be transported inside the cell from the environment. Thus, we considered these cofactors as an additional source of the metabolism of *Portiera*. The Uniprot database allowed us to find the cofactors associated with the different enzymes encoded by *the Portiera* genome.

### **9. Nucleotide sequence accession numbers**

The complete genome sequence of *Portiera aleyrodidarum* from the MED whiteflies has been deposited at GenBank under the accession number CP007563. The draft genome of

*Hamiltonella defensa* from the MED whiteflies has been deposited under the accession number AJLH00000000 (Rao *et al.*, 2012). The version described in this paper is the second version, AJLH02000000.

## ACKNOWLEDGEMENTS

This work was supported by grants from the National Natural Science Foundation of China (Projects 31321063 and 31390421). This work was supported by grants BFU2012-39816-C02-01 (co-financed by FEDER funds and Ministerio de Economía y Competitividad, Spain) and Prometeo 92/2009 from Generalitat Valenciana (Spain) and by EU COST Action FA0701. D.S-G. is a recipient of a contract from Prometeo 92/2009. These results have been achieved within the framework of the 1st call on Mediterranean agriculture carried out by ARIMNet, with funding from MOARD (IL), ANR (FR), INIA (ES), NAGREF-DEMETER (GR), and GDAR (TR). P. A. Rollat-Farnier is a recipient of a grant from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° [247073]10 SISYPHE. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## REFERENCE

- Baumann P. **2005**. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annual Review of Microbiology*, **59**, 155-189.
- Bennett G.M. et Moran N.A. **2013**. Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a Phloem-feeding insect. *Genome Biology and Evolution*, **5**, 1675-1688.
- Bing X.L., Ruan Y.M., Rao Q., Wang X.W. et Liu S.S. **2013a**. Diversity of secondary endosymbionts among different putative species of the whitefly *Bemisia tabaci*. *Insect Science*, **20**, 194-206.
- Bing X.L., Yang J., Zchori-Fein E., Wang X.W. et Liu S.S. **2013b**. Characterization of a newly discovered symbiont of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Applied and Environmental Microbiology*, **79**, 569-575.
- Blattner F.R., Plunkett G., Bloch C.A., Perna N.T., Burland V., Riley M. *et al.* **1997**. The complete genome sequence of *Escherichia coli* K-12. *Science*, **277**, 1453-1462.
- Borenstein E., Kupiec M., Feldman M.W. et Ruppin E. **2008**. Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proceedings of the National Academy of Sciences*, **105**, 14482-14487.
- Brumin M., Kontsedalov S. et Ghanim M. **2011**. Rickettsia influences thermotolerance in the whitefly *Bemisia tabaci* B biotype. *Insect Science*, **18**, 57-66.
- Charles H., Balmand S., Lamelas A., Cottret L., Pérez-Brocal V., Burdin B. *et al.* **2011**. A genomic reappraisal of symbiotic function in the aphid/*Buchnera* symbiosis: reduced transporter sets and variable membrane organisations. *PLoS One*, **6**, e29096.

- Chiel E., Zchori-Fein E., Inbar M., Gottlieb Y., Adachi-Hagimori T., Kelly S.E. *et al.* **2009**. Almost there: transmission routes of bacterial symbionts between trophic levels. *PLoS One*, **4**, e4767.
- Costa H.S., Westcot D.M., Ullman D.E. et Johnson M.W. **1993**. Ultrastructure of the endosymbionts of the whitefly, *Bemisia tabaci* and *Trialeurodes vaporariorum*. *Protoplasma*, **176**, 106-115.
- Cottret L., Milreu P.V., Acuña V., Marchetti-Spaccamela A., Stougie L., Charles H. *et al.* **2010**. Graph-based analysis of the metabolic exchanges between two co-resident intracellular symbionts, *Baumannia cicadellinicola* and *Sulcia muelleri*, with their insect host, *Homalodisca coagulata*. *PLoS Computational Biology*, **6**, e1000904.
- Cottret L., Wildridge D., Vinson F., Barrett M.P., Charles H., Sagot M.F. *et al.* **2010**. MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Research*, **38**, W132-W137.
- Csardi G. et Nepusz T. **2012**. The igraph software package for complex network research. *International Journal of Complex Systems*.
- Darby A.C., Choi J.H., Wilkes T., Hughes M.A., Werren J.H., Hurst G.D.D. *et al.* **2010**. Characteristics of the genome of *Arsenophonus nasoniae*, son-killer bacterium of the wasp *Nasonia*. *Insect Molecular Biology*, **19**, 75-89.
- De Barro P.J. et Driver F. **1997**. Use of RAPD PCR to distinguish the B biotype from other biotypes of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae). *Australian Journal of Entomology*, **36**, 149-152.
- De Barro P.J., Liu S.-S., Boykin L.M. et Dinsdale A.B. **2011**. *Bemisia tabaci*: a statement of species status. *Annual Review of Entomology*, **56**, 1-19.
- Degnan P.H., Lazarus A.B. et Wernegreen J.J. **2005**. Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects. *Genome Research*, **15**, 1023-1033.
- Degnan P.H., Yu Y., Sisneros N., Wing R.A. et Moran N.A. **2009**. *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proceedings of the National Academy of Sciences*, **106**, 9063-9068.
- Dinsdale A., Cook L., Riginos C., Buckley Y.M. et Barro P.D. **2010**. Refined global analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. *Annals of the Entomological Society of America*, **103**, 196-208.
- Douglas A.E. **1998**. Nutritional interactions in insect-microbial symbioses: Aphids and their symbiotic bacteria *Buchnera*. *Annual Review of Entomology*, **43**, 17-37.
- Douglas A.E. **2006**. Phloem-sap feeding by animals: problems and solutions. *Journal of Experimental Botany*, **57**, 747-754.
- Ferrari J. et Vavre F. **2011**. Bacterial symbionts in insects or the story of communities affecting communities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 1389-1400.
- Ghanim M. et Kontsedalov S. **2009**. Susceptibility to insecticides in the Q biotype of *Bemisia tabaci* is correlated with bacterial symbiont densities. *Pest Management Science*, **65**, 939-942.
- Ghanim M., Sobol I., Ghanim M. et Czosnek H. **2007**. Horizontal transmission of begomoviruses between *Bemisia tabaci* biotypes. *Arthropod-Plant Interactions*, **1**, 195-204.
- Gil R., Silva F.J., Zientz E., Delmotte F., González-Candelas F., Latorre A. *et al.* **2003**. The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proceedings of the National Academy of Sciences*, **100**, 9388-9393.

- Gottlieb Y., Ghanim M., Chiel E., Gerling D., Portnoy V., Steinberg S. *et al.* **2006.** Identification and localization of a *Rickettsia* sp. in *Bemisia tabaci* (Homoptera: Aleyrodidae). *Applied and Environmental Microbiology*, **72**, 3646-3652.
- Gottlieb Y., Ghanim M., Gueguen G., Kontsedalov S., Vavre F., Fleury F. *et al.* **2008.** Inherited intracellular ecosystem: symbiotic bacteria share bacteriocytes in whiteflies. *The FASEB Journal*, **22**, 2591-2599.
- Gueguen G., Vavre F., Gnankine O., Peterschmitt M., Charif D., Chiel E. *et al.* **2010.** Endosymbiont metacommunities, mtDNA diversity and the evolution of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. *Molecular Ecology*, **19**, 4365-4378.
- Hansen A.K. et Moran N.A. **2014.** The impact of microbial symbionts on host plant utilization by herbivorous insects. *Molecular Ecology*, **23**, 1473-1496.
- Hu J., De Barro P., Zhao H., Wang J., Nardi F. et Liu S.S. **2011.** An extensive field survey combined with a phylogenetic analysis reveals rapid and widespread invasion of two alien whiteflies in China. *PLoS One*, **6**, e16061.
- Husnik F., Nikoh N., Koga R., Ross L., Duncan R.P., Fujie M. *et al.* **2013.** Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell*, **153**, 1567-1578.
- Jiang Z.F., Xia F., Johnson K.W., Bartom E., Tuteja J.H., Stevens R. *et al.* **2012.** Genome sequences of the primary endosymbiont "*Candidatus portiera aleyrodidarum*" in the whitefly *Bemisia tabaci* B and Q biotypes. *Journal of Bacteriology*, **194**, 6678-6679.
- Karp P.D., Paley S. et Romero P. **2002.** The pathway tools software. *Bioinformatics*, **18**, S225-S232.
- Lagesen K., Hallin P., Rodland E.A., Staerfeldt H.H., Rognes T., Ussery D.W. **2007.** Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, **35**, 3100-3108.
- Lamelas A., Gosalbes M.J., Manzano-Marín A., Peretó J., Moya A. et Latorre A. **2011.** *Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont. *PLoS Genetics*, **7**, e1002357.
- Lerat E. et Ochman H. **2005.** Recognizing the pseudogenes in bacterial genomes. *Nucleic Acids Research*, **33**, 3125-3132.
- Lin H., Lou B., Glynn J.M., Doddapaneni H., Civerolo E.L., Chen C. **2011.** The complete genome sequence of "*Candidatus Liberibacter solanacearum*", the bacterium associated with potato zebra chip disease. *PLoS One*, **6**, e19135.
- López-Madrugal S., Latorre A., Porcar M., Moya A. et Gil R. **2013.** Mealybugs nested endosymbiosis: going into the 'matryoshka' system in *Planococcus citri* in depth. *BMC Microbiology*, **13**, 74.
- López-Sánchez M.J., Neef A., Peretó J., Patiño-Navarrete R., Pignatelli M., Latorre A. *et al.* **2009.** Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. *PLoS Genetics*, **5**, e1000721.
- Mahadav A., Gerling D., Gottlieb Y., Czosnek H. et Ghanim M. **2008.** Parasitization by the wasp *Eretmocerus mundus* induces transcription of genes related to immune response and symbiotic bacteria proliferation in the whitefly *Bemisia tabaci*. *BMC Genomics*, **9**, 342.
- Martin V.J. et Mohn W.W. **2000.** Genetic investigation of the catabolic pathway for degradation of abietane diterpenoids by *Pseudomonas abietaniphila* BKME-9. *Journal of Bacteriology*, **182**, 3784-3793.

McCutcheon J.P. et Moran N.A. **2007**. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proceedings of the National Academy of Sciences*, **104**, 19392-19397.

McCutcheon J.P., McDonald B.R. et Moran N.A. **2009a**. Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genetics*, **5**, e1000565.

McCutcheon J.P., McDonald B.R. et Moran N.A. **2009b**. Convergent evolution of metabolic roles in bacterial symbionts of insects. *Proceedings of the National Academy of Sciences*, **106**, 15394-15399.

McCutcheon J.P. et Moran N.A. **2010**. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biology and Evolution*, **2**, 708-718.

McCutcheon J.P. et von Dohlen C.D. **2011**. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Current Biology*, **21**, 1366-1372.

Moran N.A., Russell J.A., Koga R. et Fukatsu T. **2005c**. Evolutionary relationships of three new species of *Enterobacteriaceae* living as symbionts of aphids and other insects. *Applied and Environmental Microbiology*, **71**, 3302-3310.

Moran N.A., McCutcheon J.P. et Nakabachi A. **2008**. Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics*, **42**, 165-190.

Moran N.A. et Jarvik T. **2010**. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science*, **328**, 624-627.

Moriya Y., Itoh M., Okuda S., Yoshizawa A.C. et Kanehisa M. **2007**. KAAZ: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, **35**, W182-W185.

Moya A., Pereto J., Gil R. et Latorre A. **2008**. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nature Reviews Genetics*, **9**, 218-229.

Nakabachi A., Yamashita A., Toh H., Ishikawa H., Dunbar H.E., Moran N.A. *et al.* **2006**. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science*, **314**, 267.

Nakabachi A., Ueoka R., Oshima K., Teta R., Mangoni A., Gurgui M. *et al.* **2013**. Defensive bacteriome symbiont with a drastically reduced genome. *Current Biology*, **23**, 1478-1484.

Nandineni M.R. et Gowrishankar J. **2004**. Evidence for an arginine exporter encoded by *yggA* (*argO*) that is regulated by the LysR-type transcriptional regulator *ArgP* in *Escherichia coli*. *Journal of Bacteriology*, **186**, 3539-3546.

Ogata H., La Scola B., Audic S., Renesto P., Blanc G., Robert C. *et al.* **2006**. Genome sequence of *Rickettsia bellii* illuminates the role of *amoebae* in gene exchanges between intracellular pathogens. *PLoS Genetics*, **2**, e76.

Pérez-Brocá V., Gil R., Ramos S., Lamelas A., Postigo M., Michelena J.M. *et al.* **2006**. A small microbial genome: the end of a long symbiotic relationship? *Science*, **314**, 312-313.

Price D.R.G., Feng H., Baker J.D., Bavan S., Luetje C.W. et Wilson A.C.C. **2014**. Aphid amino acid transporter regulates glutamine supply to intracellular bacterial symbionts. *Proceedings of the National Academy of Sciences*, **111**, 320-325.

Rao Q., Luo C., Zhang H., Guo X. et Devine G.J. **2011**. Distribution and dynamics of *Bemisia tabaci* invasive biotypes in central China. *Bulletin of Entomological Research*, **101**, 81-88.

Rao Q., Wang S., Su Y.L., Bing X.L., Liu S.S. et Wang X.W. **2012a**. Draft genome sequence of “*Candidatus Hamiltonella defensa*,” an endosymbiont of the whitefly *Bemisia tabaci*. *Journal of Bacteriology*, **194**, 3558-3558.



- Rao Q., Xu Y.H., Luo C., Zhang H.Y., Jones C.M., Devine G.J. *et al.* **2012b**. Characterisation of neonicotinoid and pymetrozine resistance in strains of *Bemisia tabaci* (Hemiptera: Aleyrodidae) from China. *Journal of Integrative Agriculture*, **11**, 321-326.
- Ren Q. et Paulsen I.T. **2005**. Comparative analyses of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. *PLoS Computational Biology*, **1**, e27.
- Russell C.W., Bouvaine S., Newell P.D. et Douglas A.E. **2013**. Shared metabolic pathways in a coevolved insect-bacterial symbiosis. *Applied and Environmental Microbiology*, **79**, 6117-6123.
- Sabree Z.L., Huang C.Y., Okusu A., Moran N.A., Normark B.B. **2013**. The nutrient supplying capabilities of *Uzinura*, an endosymbiont of armoured scale insects. *Environmental Microbiology*, **15**, 1988-1999.
- Sandström J.P., Russell J.A., White J.P. et Moran N.A. **2001**. Independent origins and horizontal transfer of bacterial symbionts of aphids. *Molecular Ecology*, **10**, 217-228.
- Santos-Garcia D., Farnier P.A., Beitia F., Zchori-Fein E., Vavre F., Mouton L. *et al.* **2012**. Complete genome sequence of “*Candidatus Portiera aleyrodidarum*” BT-QVLC, an obligate symbiont that supplies amino acids and carotenoids to *Bemisia tabaci*. *Journal of Bacteriology*, **194**, 6654-6655.
- Schattner P., Brooks A.N., Lowe T.M. **2005**. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research*, **33**, W686-689.
- Shigenobu S., Watanabe H., Hattori M., Sakaki Y. et Ishikawa H. **2000**. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature*, **407**, 81-86.
- Silva F.J., Belda E. et Talens S.E. **2006**. Differential annotation of tRNA genes with anticodon CAT in bacterial genomes. *Nucleic Acids Research*, **34**, 6015-6022.
- Sintupachee S., Milne J.R., Poonchaisri S., Baimai V. et Kittayapong P. **2006**. Closely related *Wolbachia* strains within the pumpkin arthropod community and the potential for horizontal transmission via the plant. *Microbial Ecology*, **51**, 294-301.
- Sloan D.B. et Moran N.A. **2012a**. Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. *Molecular Biology and Evolution*, **29**, 3781-3792.
- Sloan D.B. et Moran N.A. **2012b**. Endosymbiotic bacteria as a source of carotenoids in whiteflies. *Biology Letters*, **8**, 986-989.
- Sloan D.B. et Moran N.A. **2013**. The evolution of genomic instability in the obligate endosymbionts of whiteflies. *Genome Biology and Evolution*, **5**, 783-793.
- Sloan D.B., Nakabachi A., Richards S., Qu J., Murali S.C., Gibbs R.A. *et al.* **2014**. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Molecular Biology and Evolution*, **31**, 857-871.
- Smoot M.E., Ono K., Ruscheinski J., Wang P.L. et Ideker T. **2011**. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431-432.
- Su Q., Xie W., Wang S., Wu Q., Liu B., Fang Y. *et al.* **2014**. The endosymbiont *Hamiltonella* increases the growth rate of its host *Bemisia tabaci* during periods of nutritional stress. *PLoS One*, **9**, e89002.
- Tatusov R.L., Koonin E.V. et Lipman D.J. **1997**. A genomic perspective on protein families. *Science*, **278**, 631-637.
- Thao M.L. et Baumann P. **2004**. Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. *Applied and Environmental Microbiology*, **70**, 3401-3406.

- Toh H., Weiss B.L., Perkin S.A., Yamashita A., Oshima K., Hattori M. *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.
- Vautrin E., Genieys S., Charles S. et Vavre F. **2008**. Do vertically transmitted symbionts co-existing in a single host compete or cooperate? A modelling approach. *Journal of Evolutionary Biology*, **21**, 145-161.
- Vautrin E. et Vavre F. **2009**. Interactions between vertically transmitted symbionts: cooperation or conflict? *Trends in Microbiology*, **17**, 95-99.
- Wallace B., Yang Y.J., Hong J.S. et Lum D. **1990**. Cloning and sequencing of a gene encoding a glutamate and aspartate carrier of *Escherichia coli* K-12. *Journal of Bacteriology*, **172**, 3214-3220.
- Wang X.W., Luan J.B., Li J.M., Bao Y.Y., Zhang C.X. et Liu S.S. **2010**. De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics*, **11**, 400.
- Wilson A.C., Ashton P.D., Calevro F., Charles H., Colella S., Febvay G. **2010**. Genomic insight into the amino acid relations of the pea aphid, *Acyrtosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. *Insect Molecular Biology*, **19**, 249-258.
- Woyke T., Tighe D., Mavromatis K., Clum A., Copeland A., Schackwitz W. *et al.* **2010**. One bacterial cell, one complete genome. *PLoS One*, **5**, e10314.
- Wu M., Sun L.V, Vamathevan J., Riegler M., Deboy R., Brownlie J.C. *et al.* **2004**. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biology*, **2**, 0327-0341.
- Wu D., Daugherty S.C., Van Aken S.E., Pai G.H., Watkins K.L., Khouri H. *et al.* **2006**. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biology*, **4**, e188.
- Zchori-Fein E. et Brown J.K. **2002**. Diversity of prokaryotes associated with *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae). *Annals of the Entomological Society of America*, **95**, 711-718.
- Zchori-Fein E. et Bourtzis K. **2011**. *Manipulative tenants: bacteria associated with arthropods*. CRC Press.
- Zientz E., Dandekar T. et Gross R. **2004**. Metabolic interdependence of obligate intracellular bacteria and their insect hosts. *Microbiology and Molecular Biology Reviews*, **68**, 745-770.



## BILAN

Nous avons dans cet article analysé les capacités métaboliques potentielles de deux protagonistes partageant les mêmes cellules au sein de l'holobionte *Bemisia tabaci* MED-Q1.

*Portiera aleyrodidarum* présentait des caractéristiques communes aux autres génomes assemblés chez *Bemisia tabaci* : un génome réduit, ne possédant pas les gènes nécessaires à la synthèse des vitamines et cofacteurs, et révélant des « brèches » dans les voies de synthèse de différents acides aminés essentiels. Autre fait intéressant, ces symbiotes sont impliqués dans la production de caroténoïdes (Sloan et Moran, 2012b ; Jiang *et al.*, 2013 ; Santo-Garcia *et al.*, 2012). C'est la première fois qu'un endosymbiote d'Hémiptère joue un rôle dans leur production : chez le puceron *A. pisum*, c'est l'hôte qui synthétise ces pigments du fait d'un transfert ancestral de gènes fongiques (Moran et Jarvik, 2010). Chez la psylle *Pachypsylla venusta* (Nováková et Moran, 2012), des caroténoïdes ont été détectés et seraient synthétisés par l'eucaryote, le symbiote primaire *Carsonella ruddii* ne possédant pas les gènes nécessaires (Nakabachi *et al.*, 2006).

Le fait que la synthèse de ces pigments soit possible dans au moins trois des quatre grands clades de Sternorrhyncha suggère un rôle important dans ce Sous-Ordre des Hémiptères (Sloan et Moran, 2012). Ainsi, leur action anti-oxydante pourrait compenser la perte des gènes dédiés à la réparation de l'ADN chez le symbiote primaire, protégeant de ce fait le génome contre différents stress, notamment la lumière, les phloémophages y étant particulièrement exposés. Ils permettraient également de prévenir les dommages liés à l'activité métabolique intense qui a lieu au sein des bactériocytes (Sloan et Moran, 2012).

Ces métabolites pourraient ainsi faciliter les interactions avec ses autres partenaires. En effet, en utilisant la reconstruction des métabolismes bactériens ainsi que des données de transcriptomique issues d'une précédente étude ciblant l'eucaryote (Wang *et al.*, 2010), il a été démontré que ce dernier assurait certainement les étapes manquantes pour différents acides aminés : la leucine, l'isoleucine et la valine. Il est également possible que l'hôte participe à la synthèse de la phénylalanine. Il y aurait donc en premier lieu une complémentarité métabolique entre *B. tabaci* et son symbiote primaire. Ensuite, seul le symbiote secondaire *H. defensa* semble en mesure d'assurer les réactions manquantes en ce

qui concerne la synthèse de la lysine, voire, il la synthétiserait seule. La bactérie pourrait en outre participer à la synthèse de la phénylalanine en complémentation avec le symbiote primaire, en lieu et place de l'hôte. Il est en tous les cas certain que le consortium dépende de ce symbiote pour la synthèse de la lysine, un acide aminé essentiel, de sorte qu'*H. defensa* pourrait en être une composante obligatoire. Cette analyse métabolique est corroborée par une étude récente, qui a démontré le bénéfice retiré par *B. tabaci* en période de stress alimentaire, quant à la présence d'*H. defensa* (Su *et al.*, 2014). Le symbiote jouerait donc bien un rôle nutritif chez cet aleurode.

Ainsi, si peu de cas de complémentations ont été mis en évidence entre les deux symbiotes, ils revêtent de par leur nature essentielle une importance capitale pour comprendre la relation qui s'exerce entre les trois partenaires. Il pourrait y avoir dépendance de *B. tabaci* et de *P. aleyrodidarum* à un troisième partenaire, *H. defensa*. La réciproque pourrait également être vraie. Certes, tout symbiote est dépendant de l'environnement intracellulaire. Cependant, les symbiotes facultatifs tels que ceux du genre *Hamiltonella* sont capables, nous l'avons vu, de transiter depuis leur hôte vers un autre. Or, *H. defensa* chez *B. tabaci* MED-Q1 a perdu la capacité à synthétiser le chorismate, du fait d'une inactivation du gène *aroA*, métabolite que seul le symbiote primaire peut lui fournir. Le symbiote pourrait donc être « piégé » dans l'aleurode, dépendant d'un environnement riche en chorismate, et donc incapable de survivre hors des bactériocytes primaires.



# ARTICLE SECOND

---

## Deux clades d'hôtes, deux arsenaux bactériens : Évolution par perte de gènes chez les endosymbiotes facultatifs

---

### Contexte

*Hamiltonella defensa* pourrait être un partenaire obligatoire de l'aleurode *Bemisia tabaci*, et engagé de manière irréversible dans cette relation. Cela est tout particulièrement intéressant au regard du phénotype et de l'écologie des symbiotes infectant les pucerons. Chez ces insectes, *H. defensa* propose des caractéristiques plus « typiques » des symbiotes facultatifs, et notamment une plus grande diversité dans les tissus infectés. Ce pouvoir invasif lui permet notamment d'envahir l'hémolymphe, la mettant en relation directe avec l'(es) œuf(s) de parasitoïdes qui s'y trouve(nt). Elle peut alors exercer son rôle protecteur au détriment de la progéniture de la guêpe. De manière intéressante, les bactéries de l'espèce *Regiella insecticola*, qui constituent le clade-frère d'*H. defensa*, sont également des endosymbiotes protecteurs. Ainsi, ce phénotype aurait été acquis depuis leur ancêtre commun, et nous aurions une transition écologique radicale chez les symbiotes de l'aleurode. Les « causes » semblent avoir été identifiées, et consisteraient simplement en quelques gènes de synthèse. Qu'en est-il des « conséquences », c'est-à-dire des répercussions pour le génome de la bactérie ?

### Attente et objectif

Dans ce second article, nous nous sommes attelés à la comparaison génomique des deux « clades » d'*H. defensa*. Nous nous sommes principalement préoccupés des différents systèmes qui sont l'apanage des symbiotes secondaires et sont généralement absents chez les symbiotes obligatoires. De tels systèmes (facteurs de virulence, toxines, effecteurs, systèmes à deux composantes, régulateurs etc.) sont nombreux chez les symbiotes infectant le puceron (Moran *et al.*, 2005d ; Degnan *et al.*, 2009), et sont en lien direct avec leur pouvoir invasif ainsi que leur phénotype de tueurs de parasitoïdes. Leur intérêt semble en revanche peu évident pour *H. defensa* chez l'aleurode, séquestrée dans un seul type cellulaire et potentiellement impliquée dans un rôle essentiellement nutritif. Leur devenir

nous intéressera donc tout particulièrement dans ces organismes potentiellement en transition vers un mode de vie obligatoire.

---

Two host clades, two bacterial arsenals:  
Evolution through gene losses in facultative endosymbionts

---

Pierre-Antoine Rollat-Farnier, Diego Santos-Garcia, Qiong Rao, Marie-France Sagot, Francisco J. Silva, Hélène Henri, Einat Zchori-Fein, Amparo Latorre, Andrés Moya, Valérie Barbe, Shu-Sheng Liu, Xiao-Wei Wang, Fabrice Vavre et Laurence Mouton

Cet article a été soumis à *Genome Biology and Evolution* (en révision)

**ABSTRACT**

Bacterial endosymbiosis is an important evolutionary process in insects, which can harbor both obligate and facultative symbionts. The evolution of these symbionts is driven by evolutive convergence, and they exhibit among the tiniest genomes in prokaryotes. The large host spectrum of facultative symbionts and the high diversity of strategies they use to infect new hosts probably impacts the evolution of their genome and explains why they undergo less severe genomic erosion than obligate symbionts.

*Candidatus Hamiltonella defensa* is suitable for the investigation of genomic evolution of facultative symbionts because the bacteria are engaged in specific relationships in two clades of insects. In aphids, *H. defensa* is found in several species with an intermediate prevalence and confers protection against parasitoids. In whiteflies, *H. defensa* is almost fixed in some species of *Bemisia tabaci*, which suggests an important role of and a transition towards obligate symbiosis.

In the present study, comparisons of the genome of *H. defensa* present in two *B. tabaci* species (MEAM1 and MED) and in the aphid *Acyrtosiphon pisum* revealed that they belong to two distinct clades and underwent specific gene losses. In aphids, it contains highly virulent factors that could allow protection and horizontal transfers. In whiteflies, the genome lost these factors and seems to have a limited ability to acquire genes. However it contains genes that could be involved in the production of essential nutriment, which is consistent with a primordial role for this symbiont.

In conclusion, while both lineages of *H. defensa* have mutualistic interactions with their hosts, their genomes follow distinct evolutive trajectories that reflect their phenotype and could have important consequences on their evolvability.

**Key-words:** comparative genomics, *Hamiltonella defensa*, aphids, *Bemisia tabaci*

## I. Introduction

Endosymbiosis is an important evolutionary process in insects, as more than 20% of the insect species depend on endosymbiotic bacteria for their development and survival (Moya *et al.*, 2008; Moran *et al.*, 2008). Insect endosymbioses can be classified based on the host dependency on the bacterium: obligate symbionts are referred to as primary symbionts (P-symbionts), while facultative symbionts are called secondary symbionts (S-symbionts). The first are mutualists for the insect, which hosts them in specialized cells named bacteriocytes, which in most cases form an organ-like structure known as the bacteriome (Buchner, 1965). In most known cases, the classical contribution of such bacteria is nutritional, and involves supplying metabolites that are essential to insect hosts that rely on unbalanced diets like phloem or blood (Buchner, 1965; Nogge, 1982). These endosymbiotic associations have allowed the hosts to explore new niches; supporting that symbiosis is an important source of evolutionary innovation. Regarding S-symbionts, their prevalence in insect populations is highly variable (Cheng *et al.*, 2000), contrary to the fixed P-symbionts. Their persistence relies on various – not mutually exclusive – strategies including both context-dependent mutualism (Moran *et al.*, 2005a; Lamelas *et al.*, 2011) and reproductive manipulation (Stouthamer *et al.*, 1999). S-symbionts can be located in the bacteriocytes but may also be found in different tissues of the host, including the hemolymph (Fukatsu *et al.*, 2000). They are often maternally transmitted, but unlike the P-symbionts whose vertical transmission is strict, S-symbionts are often imperfectly transmitted and can also be horizontally transferred among insects of the same or different species (Chiel *et al.*, 2009). These different dynamics are reflected by the incongruence of phylogenetic trees found between hosts and S-symbionts (Russel *et al.*, 2003; Baumann, 2005; Lamelas *et al.*, 2008), as opposed to the nearly perfect co-cladogenesis between P-symbionts and their hosts (Baumann, 2005). As a consequence of horizontal transfers, insect species harboring P-symbionts are also frequently infected by S-symbionts, leading to multiple infections within host individuals (Fukatsu *et al.*, 2000; Sloan and Moran, 2012a).

The unique lifestyle of symbionts has shaped their genome features. On the one hand, in each generation, only a small part of the symbiotic population is transmitted to the offspring (Moran, 1996). These frequent bottlenecks reduce the effective size of the



symbiotic population and thus the efficiency of selection with different consequences. Mildly deleterious mutations can accumulate, notably through AT mutational bias (Moran, 1996). Similarly, new insertions of mobile elements are less counter-selected, which can later on facilitate intragenomic recombination and large deletions (Silva *et al.*, 2001). On the other hand, different gene categories become dispensable in the new and stable conditions (i.e. the host environment), such as genes related to mobility, sensing of the environment and regulation (Shigenobu *et al.*, 2000). Finally, stability of the association at the intergenerational level favors reduction in the virulence arsenal of symbionts. As a consequence of these different processes, endosymbiont genomes have the smallest size and the highest AT bias in eubacteria (Shigenobu *et al.*, 2000; Santos-Garcia *et al.*, 2012; Sloan and Moran, 2012b; Lopez-Madrugal *et al.*, 2011). These characteristics are more pronounced in P-symbionts that only retain gene content close to the minimal set required for life (Gil *et al.*, 2004; Pérez-Brocal *et al.*, 2006). In contrast, S-symbionts that can still infect different tissues and different hosts have retained genes allowing for more versatile behavior. These genes include the ones involved in regulation and environmental perception, and above all virulence factors to overcome the host immune system (Ochman and Moran, 2001), some of them being carried by phages and plasmidic islands (Wu *et al.*, 2004; Degnan *et al.*, 2009). While genome differences between P- and S-symbionts are now well documented, it is less clear how specificities in the relationship between the host and S-symbionts can affect the evolution of the bacterial genomes. For example, in a recent genomic study the supply to the host with riboflavin was postulated to be the key of the establishment of *Serratia symbiotica* as a co-obligate endosymbiont in aphids from the genus *Cinara* (Manzano-Marín and Latorre, 2014).

*Hamiltonella defensa* is a suitable candidate to approach this question. These Gammaproteobacteria can mainly be found in two important clades of insects – aphids and whiteflies (Moran *et al.*, 2005b; Telesnicki *et al.*, 2012) – with which they are engaged in two specific relationships. Several aphid species are infected by *H. defensa*, which highlights its ability to horizontally transfer (Russel *et al.*, 2003; Telesnicki *et al.*, 2012). In this clade, *H. defensa* behaves as a protective symbiont against parasitoids, through the synthesis of toxins encoded by the bacteriophage APSE (Moran *et al.*, 2005b). Interestingly, the presence of *H. defensa* is also associated with infection costs (Vorburger and Gouskov, 2011), which

could explain the intermediate prevalence observed for this symbiont and its correlation with the intensity of parasitoid attacks. Within aphid individuals, *H. defensa* is located in the bacteriome, together with their P-symbiont *Buchnera aphidicola* (ensuring its maternal transmission), but also in the insect hemolymph, which may favor the contact with the parasitoid (Darby *et al.*, 2001; Moran *et al.*, 2005b). In whiteflies, contrary to the large host spectrum in the superfamily of aphids, *H. defensa* is restricted to two closely related species of the *Bemisia tabaci* species complex, namely Mediterranean (MED, formerly referred to as the 'Q' biotype) and Middle East Asia Minor 1 (MEAM1, formerly referred to as the 'B' biotype) species (Gueguen *et al.*, 2010). In addition, *H. defensa* is almost fixed in MEAM1 and in the Q1 sub-species of MED, while it is totally absent from the two other MED sub-species Q2 and Q3 (Gueguen *et al.*, 2010; Pan *et al.*, 2012). Also in contrast with the situation in aphids where the symbionts are in separated bacteriocytes (Tsuchida *et al.*, 2005), *H. defensa* co-infects the same bacteriocytes as the P-symbiont *Portiera aleyrodidarum* in *B. tabaci* (Gottlieb *et al.*, 2008; Skaljac *et al.*, 2010). Because *P. aleyrodidarum* is not able to fully compensate for the insect diet and present some incomplete amino acid biosynthetic pathways, bacteriocyte sharing could favor metabolic exchanges between the two symbionts. In addition, *H. defensa* has been shown to increase growth rate under nutritional stress in *B. tabaci* MED species (Su *et al.*, 2014). All these elements suggest an important metabolic contribution of *H. defensa* in whiteflies.

Recently, the genomes of *H. defensa* have been sequenced from both aphids and whiteflies. The genome of *H. defensa* from the aphid *Acyrtosiphon pisum* (strain 5AT) exhibits characteristics of an S-symbiont with a moderate AT bias and genome reduction (Degnan *et al.*, 2009). Consistent with the phenotype of the bacterium, it contains different virulence factors that may mediate both protection against parasitoids and the capacity to invade the tissues of the insect host. In contrast, the analysis of the *H. defensa* genome from *B. tabaci* MED-Q1 suggests that it could supply its host with different metabolites, including the essential amino acid lysine (Rao *et al.*, in prep.). This is consistent with a primordial role for this symbiont and could explain its fixation in some whitefly populations. In order to analyze the genomic consequences of the specific lifestyle and relationships *H. defensa* has established with its aphid and whitefly hosts, we compared the genomes of *H. defensa* in *A. pisum* and in *B. tabaci* MED-Q1. The genome of *H. defensa* from the MEAM1 species of *B.*

*tabaci* was sequenced and compared to the two other genomes. Special attention was given to the history of infection, and to gene content evolution in relation to the metabolism, virulence and environmental sensing. Because of its putative obligate status, it was hypothesized that *H. defensa* in *B. tabaci* would consistently have genomic characteristics and evolution revealing a transition towards primary endosymbiosis.

## II. Results

### 1. Genome features of hBtab\_MEAM1

The genome of hBtab\_MEAM1 was assembled into 183 contigs. The total length of the assembled genome is 1.72 Mb (**Table 1**), but this is probably an underestimation of the real size of the genome as it includes some multicopy sequences that were partially collapsed during the process. However, all the unique sequences in the genome are probably represented in the assembly, as 100-fold coverage is the average, which should guarantee a representation of a large part of the genome. Furthermore, the assembly contains the sequences of the 205 genes known to be universally present as single-copy genes in almost all Gammaproteobacteria (Lerat *et al.*, 2003). The persistence of multiple contigs thus most probably reflects the abundance of repetitive sequences, which represented a problem for the assembly. This assembly thus allows for a confident interpretation of the gene functions present in *H. defensa*.

Features	<i>B. aphidicola</i> APS	hBtab_MEAM1	hBtab_MED-Q1	hApis_5AT	<i>R. insecticola</i> LSR1	<i>E. coli</i> K12	<i>Y. pestis</i> CO92
Chromosome size	640 681	1 726 317	1 840 000	2 110 331	2 035 106	4 639 221	4 653 728
Extrachromosomal element	2	0	0	1	1	0	3
Total G + C (%)	26.2	40.1	40.3	40.1	42.4	50.8	47.6
Predicted CDS	571	1400	1806	2100	1761	4284	4012
Pseudogenes	13	232	NA	188	214	150	149
Average CDS Size	984	965	NA	812	856	950	998
Coding density (%)	86.7	78.8	NA	80.8	71.4	87.9	83.8
rRNA operons	2	2 <sup>a</sup>	NA	3	4	7	6
tRNAs	32	40	38	42	36	86	70

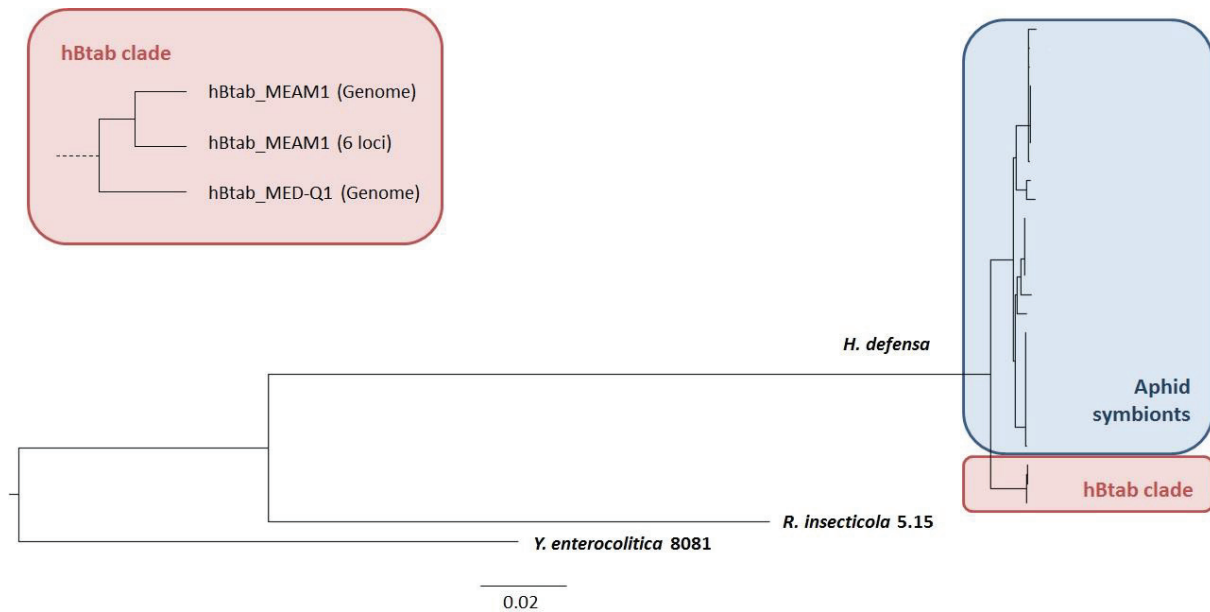
**Table 1: Properties of different draft/complete published Gammaproteobacteria genomes**

<sup>a</sup> In hBtab\_MEAM1, the real number of rRNA operons is not known: one operon contains the three rRNA genes (5S, 23S and 16S), and there is also one copy of the 16S rRNA and one copy of the 5S rRNA which are each near the end of a different contig. It is not possible to determine if these two copies are on the same operon, and if there is a second 23S copy or not.

Overall, the general characteristics of hBtab\_MEAM1 are very similar to those of both hBtab\_MED and hApis\_5AT in terms of size, gene and GC content (**Table 1**). All three genomes are typical of S-symbionts, showing moderately reduced size and AT bias, and a high number of repetitive sequences compared to the total absence in long-term P-symbionts. The most important difference between hApis\_5AT and the two *H. defensa* strains from *B. tabaci* seems to be the absence of a non-integrating plasmid in the two latter genomes. In hBtab\_MEAM1, 1,400 CDS and 232 pseudogenes were identified, which represent a slightly lower gene number compared to the other *H. defensa* genomes (**Table 1**). The differences between the two close genomes of hBtab\_MED-Q1 and hBtab\_MEAM1 could be due to differences in the respective assembly, i.e. in the parameters that have been applied for CDS detection and in pseudogene annotation, but they also reflect some real differences in gene content. The most telling example concerns an entire (inactivated) 15 kb long *tri* operon, present in both hBtab\_MED-Q1 and hApis\_5AT, but absent from the genome of hBtab\_MEAM1.

## 2. Phylogenetic reconstruction and infection history

In order to clarify the phylogenetic relationships between the strains associated with aphids and whiteflies, we concatenated six orthologous loci from 23 *H. defensa* strains, among which 20 were infecting aphids and three infecting *B. tabaci*. The phylogenetic analysis showed that the three bacteria of *B. tabaci* form a monophyletic clade (**Figure 1**), hereafter referred to as the hBtab clade. The phylogenetic proximity between the bacteria of *B. tabaci* is consistent with the 99.9% of similarity at the level of the 16S rDNA gene between the MEAM1 and MED-Q1 strains, which each have 98% similarity with the sequence of hApis\_5AT. Moreover, the comparative genomic analysis clearly revealed different events of gene inactivations and acquisitions specific to each clade (see below).



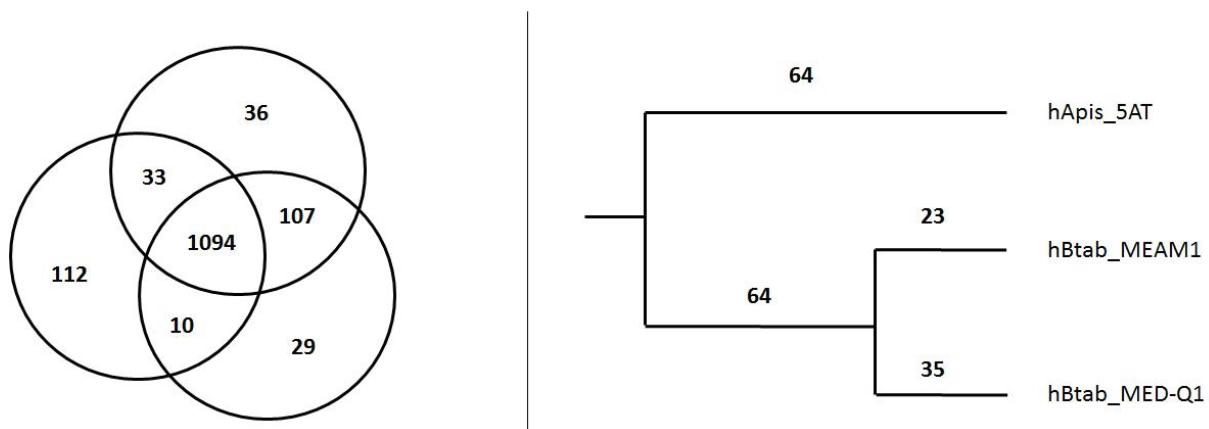
**Figure 1: Simplified phylogeny of the genus *Hamiltonella*.** The monophyly of the hBtab clade is supported by a concatenation of six gene sequences and a bootstrap value of 100. A zoom on the hBtab clade illustrates the separation between the symbionts of the MEAM1 and MED-Q1 species. The legend in brackets indicates if the used data came from the hBtab draft genomes or from the six loci sequenced by Telesnicki *et al.* (2012).

Due to their genomic divergence, the presence of *H. defensa* in aphids and whiteflies can only be explained by horizontal transfers (Darby *et al.*, 2001). Within the *B. tabaci* species complex, the two *H. defensa* from MEAM1 are monophyletic (**Figure 1**). However, it is still not possible to know whether *H. defensa* was present in the ancestor of MED and MEAM1, horizontally transferred between both species or independently acquired from unknown donors. First, the clustering of the two bacteria of MEAM1 is supported by only one mutation in one gene, the five other genes being identical in the three bacteria of the hBtab clade. Moreover, the divergence time between the two species is still under debate and thus cannot be used to determine the expected divergence between the symbionts: Boykin *et al.* (2013) estimated that the ancestor of the two species lived 13 mya ago, but a recent reevaluation placed this divergence below 1 my (Santos-Garcia *et al.*, in prep.). However the *B. tabaci* P-symbiont *P. aleyrodidarum*, whose ancestral presence is certain, can be a good reference. It exhibits a similarity of 99.8% for the 16S rDNA sequences between the MEAM1 and MED species (Santos-Garcia *et al.*, 2012), which is of the same order of magnitude as the similarity between the MED and MEAM1 *H. defensa*. Therefore an ancestral presence or two independent acquisitions are both possible even if differences

between the MED and MEAM1 *H. defensa* are sufficient to exclude a very recent horizontal transmission.

### 3. Gene content evolution in *H. defensa*

After the removal of pseudogenes, fragments due to the draft assembly, duplications and small ORFs without any significant match in the TrEMBL and Swiss-Prot databases, the orthology relationship between the three *H. defensa* strains revealed an important set of common genes (**Figure 2A**). The substantial difference in global gene number and content compared to previous studies (Degnan *et al.*, 2009; Rao *et al.*, 2012), stems from the strict cleaning step, which eliminated several hundreds of genes, especially those specific to the hBtab clade or to hApis\_5AT. For example, among the genes specific to hApis\_5AT, more than 300 were eliminated because they were composed of less than 100 amino acids and had no significant match in the database. More than 100 were pseudogenes, and more than 200 were eliminated because they were associated with duplication events.



**Figure 2:** (A) Euler diagram of the orthologs shared by the three bacterial strains after manual curation of the gene families and a cleaning step. (B) Gene losses (including pseudogenes) in the *H. defensa* clade are represented on the corresponding branches. Phagic and plasmidic genes have been removed.

To further analyze the gene contents of the *H. defensa* genomes, we reconstructed the putative ancestor of this genus based on the gene content of all *H. defensa* strains and one of their Gammaproteobacteria relatives. The reconstruction of the ancestor highlighted that the acquisition of new genes represents a small proportion of the functional differentiation between the bacteria. Indeed, the genes specific either to the hBtab bacteria

or to hApis\_5AT are mostly mobile elements and are located on inactive phage and plasmidic islands. They were thus removed from the analysis in order to focus on the functional aspects. It then appeared that new acquisitions of genes are restricted to operons which are dedicated to the synthesis of sugars and which are specific of each lineage, Type 3 effectors and RTX toxins (details below). Thus, the great majority of the differences observed come from different losses/inactivations from the 1,265 genes composing the ancestral gene content rather than from massive gene acquisitions (**Figure 2B**).

In contrast, gene loss has been extensive, and this trend is more pronounced in the hBtab clade than in hApis\_5AT (Pearson's test,  $p$ -value < 0.05). Losses in hBtab have affected many functions known to be important for the ecology of hApis\_5AT, such as virulence factors or two-component systems (Degnan *et al.*, 2009). However, despite its protective role, hApis\_5AT lost many virulence factors retained in the hBtab clade. Some of the most important losses in each clade are detailed below.

#### **4. Functional losses in the hBtab clade**

##### **a) Extra-chromosomal elements**

The hApis\_5AT strain is associated with a conjugative IncFII plasmid called pHD5AT, present both in non-integrated and integrated forms (Degnan *et al.*, 2009). While the non-integrated form remains intact in hApis\_5AT, the integrated form is inactive due to the absence of some genes (*traQ*, *traR*) and the inactivations of others (*traY*, *traT*, *traC*, *pilL*, *pilN*, plus the two *repA* genes).

The hBtab clade lacks entire plasmid regions (*traLMNOPQRT*) and only a fragment of the *repA* gene is found in these genomes, making the replication of any non-integrated plasmid impossible. As a consequence, the plasmid genes of this clade are probably inactive islands integrated in the bacterial chromosome. Consistently, a large plasmidic region (corresponding in hApis\_5AT to the region between HDEF\_p0016 and HDEF\_p0053) is surrounded in hBtab\_MEAM1 by two genes which are found in the bacterial genome of hApis\_5AT but absent from its plasmid. Similarly, the above-mentioned fragment of *repA* (corresponding to HDEF\_p0001) is found in the neighborhood of chromosomal genes in the two bacteria of the hBtab clade. The most probable scenario is thus of an ancestral

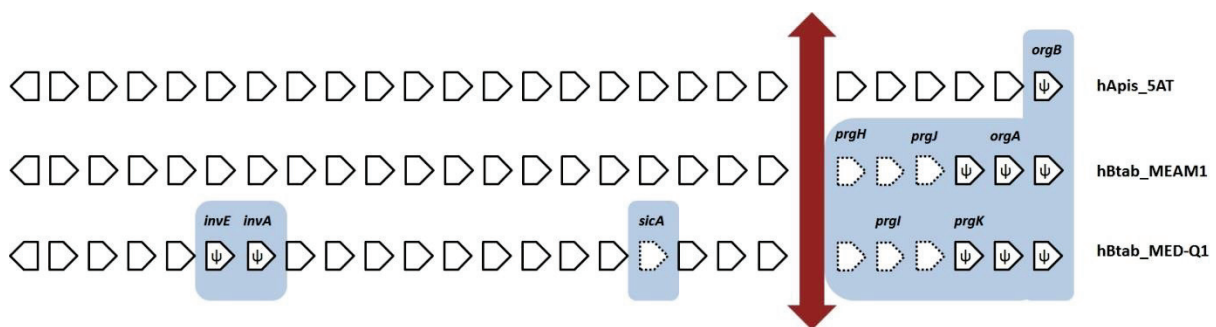
chromosomal integration and inactivation of the plasmid in the genus *Hamiltonella*, followed by the loss of the non-integrated plasmid in the hBtab clade.

The APSE phage is another well characterized partner of the *H. defensa* genus (Moran *et al.*, 2005a). Different inactivations occurred in the hBtab phage and concerned the major head protein P24, the *kilA*-domain protein G, the regulator protein I and the P38\_2 integrase, which allows both phage integration and excision. Thus, as for the conjugative plasmid, the APSE phage is probably inactive in hBtab. However, the CdtB toxin encoded by the phage, which is involved in the protection against parasitoids in aphids (Moran *et al.*, 2005a; Degnan and Moran, 2008), is also found in the hBtab genome and has conserved the 12 active residues described in Degnan and Moran (2008).

In addition to these two elements, many small phage and plasmidic islands exist in the hBtab genomes, like in hApis\_5AT (Degnan *et al.*, 2009) where they have been supposed to be inactive. As in hApis\_5AT, these islands contain a large proportion of pseudogenes, and are probably inactive and progressively degraded.

### b) Type 3 secretion systems (T3SS)

Two T3SS have been described in hApis\_5AT (Degnan *et al.*, 2009). They were called SPI-1 and 2 (for “*Salmonella* Pathogenicity Island”) because their organization was similar to the secretion systems of *Salmonella*. These two T3SS are largely degraded in hBtab, especially the SPI-1-like system (**Figure 3**). Some genes directly involved in the apparatus formation are missing or have become pseudogenized, making this system probably inactive.



**Figure 3: SPI-1 (T3SS) gene clusters in *H. defensa*.** The vertical red arrow represents an ancestral split of the operon. Pseudogenes are indicated by the letter  $\Psi$ , missing gene by dashed lines. The *orgB* ancestral inactivation could have already reduced the efficiency of this system in the last common



ancestor. In hBtab, *prgH*, *I*, *J* and *K* are absent or inactive, leading very certainly to a nonfunctional system.

The SPI-2 T3SS also contains pseudogenes in the hBtab clade. Indeed, the two-component system (TCS) SsrAB, which is involved in the activation of T3SS and is directly linked to virulence (Okan *et al.*, 2006; Okan *et al.*, 2010), is inactive due to a *ssrA* pseudogene. In addition, the *ssaK* gene, essential for the SPI-2 apparatus (Niemann *et al.*, 2011), has been independently truncated in the two bacteria. Even if there is not a massive gene loss like in SPI-1, the fact that two mutations independently occurred in a structural gene of the SPI-2 system suggests a relaxation of the selective pressure. Thus, this system may also be non- or sub-functional in the hBtab clade.

Consistent with its active T3SS, hApis\_5AT retains several Type 3 effectors that are absent or pseudogenized in the hBtab genomes (Degnan *et al.*, 2009; **Table 2**). All these effectors confer virulence and invasiveness to pathogenic bacteria and could be related either to protection against parasitoids or to evasion of the host immune system. This is the case of the *yopH* and *yopT* genes, which are important for resistance against phagocytosis (Andersson *et al.*, 1996; Grosdent *et al.*, 2002). Interestingly, *H. defensa* has recently been shown to persist in phagocytes in aphids (Schmitz *et al.*, 2012). In contrast, the hBtab genomes possess a single well characterized Type 3 effector that is absent in hApis\_5AT: the SPI-1 associated effector YopJ/P. *Yersinia* uses YopJ to induce cell death, inhibiting the host immune response by blocking the MAPK and NFκB pathways (Orth, 2002). It has also been presented as a candidate for protection against parasitoids in *Regiella insecticola* (Hansen *et al.*, 2012). However, the fact that the SPI-1 system is inactive in hBtab raises the question of its role in this clade, as is also the case for the three effectors that are shared between all *H. defensa*.

Gene	hBtab_MEAM1	hBtab_MED-Q1	hApis_5AT
<i>ssaE</i>			
<i>ssaB</i>			
<i>exoY_2</i>			
<i>yopJ</i>	Absent	Absent	Absent
<i>aexT_1</i>	Absent	Absent	
<i>aexT_2</i>	Absent	Absent	
<i>exoY_1</i>	Absent	Absent	
<i>nleD</i>	Absent	Absent	
<i>yopT_1</i>	Absent	Absent	
<i>yopT_2</i>	Absent	Absent	
<i>yopH</i>	Inactive	Inactive	

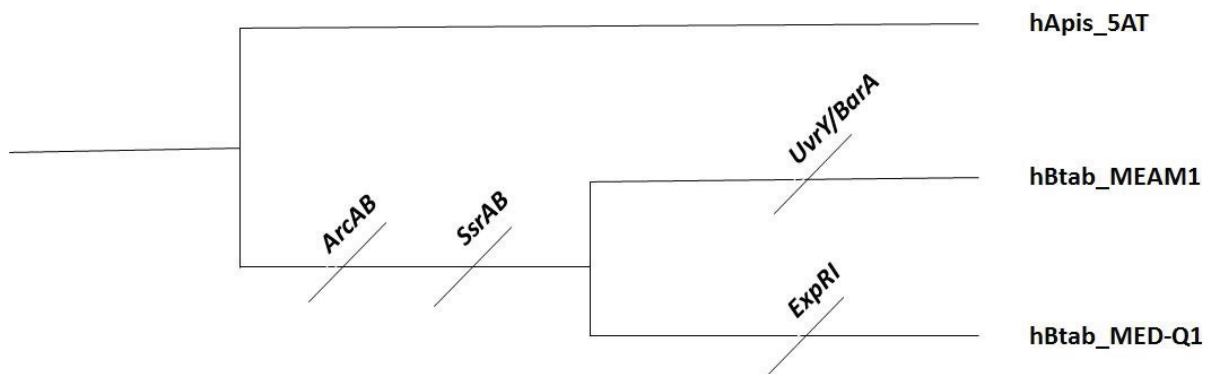
**Table 2: Type 3 effectors in the genomes of the *H. defensa* genus.** When it is not specified, the genes are present in the corresponding organism. If not, the gene can be absent or inactive.

In conclusion, the virulence arsenal within the hBtab clade is greatly reduced compared to hApis\_5AT, with both inactivation of the T3SS and loss of numerous effectors.

### c) Environmental perception

SsrAB is not the only TCS truncated in the hBtab clade (**Figure 4**). Indeed, two other systems involved in environmental perception have been inactivated, one in each hBtab strain (**Figure 4**). The UvrY/BarA TCS is an important regulator of the metabolism, also controlling quorum sensing, motility and the production of toxins (Lapouge *et al.*, 2008). It is well conserved among Gammaproteobacteria in which it coordinates for pathogenicity and social behavior. In hBtab\_MEAM1, BarA is fragmented in three, while in hBtab\_MED-Q1, the gene seems intact, but its reconstruction is not complete as it is cut by the end of a contig. The other system is the ExpRI-like quorum-sensing system, which was ancestrally present in *H. defensa*, but is inactive in hBtab\_MED-Q1 (*expI* is absent and *expR* inactive). Finally, the case of ArcAB is a bit more complicated. ArcAB is a global regulator modulating more than one hundred genes involved in central metabolism and respiratory pathways (Martínez-Antonio and Collado-Vives, 2003). It also allows for resistance against nitrogen species and reactive oxygen species in aerobic conditions (Lu *et al.*, 2002). The ancestor *R. insecticola/H.defensa* probably already underwent a mutation in ArcB: compared to its orthologs in Gammaproteobacteria, the sequences in these symbionts lack the PAS domain which is an oxygen/light sensor (Taylor and Zhulin, 1999). Nevertheless, the Ka/Ks ratio of 0.45 between the hBtab bacteria and hApis\_5AT reveals that it stayed under purifying

selection after this mutation. However, in hBtab, a second mutation introduced a premature stop codon, making the gene probably inactive.



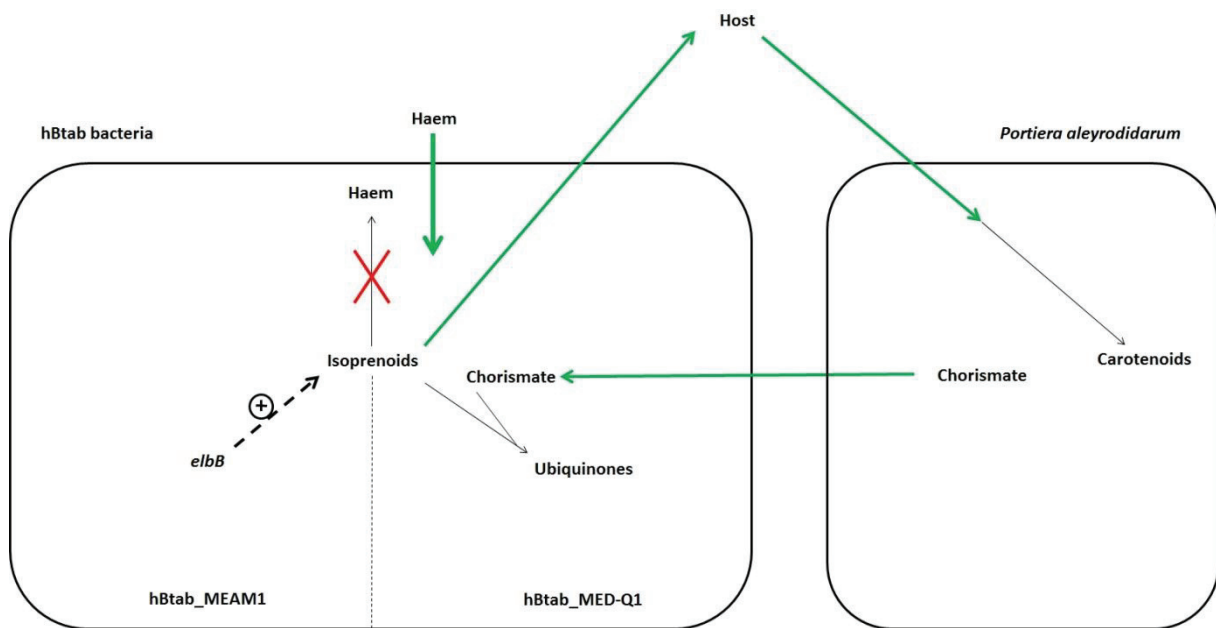
**Figure 4: Putative inactivation of the two-component and quorum-sensing systems along the evolution of *H. defensa*. The bars represent inactive systems.**

Altogether, these results suggest a lower response of the hBtab clade to environmental variations compared to the symbiont of aphids.

#### **d) Metabolism of cofactors**

hApis\_5AT can synthesize different cofactors such as biotin, pyridoxal-5-phosphate, coenzyme A, or heme, through the import of host metabolites like glucose or pantothenate (Degnan *et al.*, 2009). The hBtab clade has a metabolism close to the one of hApis\_5AT, but different inactivations prevent the production of some cofactors. An example is protoheme, an intermediary compound of the heme, which cannot be produced by the two bacteria of the hBtab clade due to many ancestral gene inactivations (*hemBCDEFG*L). However, the heme transport system (*hmuRTSUV* operon) still allows the bacterium to import host heme, which can be spontaneously converted into heme. Interestingly, heme contains an isoprenoid chain. Isoprenoids are also the precursors of carotenoids, an important metabolite furnished to the insect by the P-symbiont (Santos-Garcia *et al.*, 2012; Sloan and Moran, 2012b). It is possible to envision that this dependence to host heme/hemin could allow the latter to divert the isoprenoid fate in hBtab, using the hBtab isoprenoids to enhance the production of geranylgeranyl-phosphate, which is the probable source for carotenoids production by *P. aleyrodidarum* (Rao *et al.*, in prep.; **Figure 5**). This is also consistent with the specific inability of hBtab\_MED-Q1 to produce chorismate, which is used

with isoprenoids to produce ubiquinones (**Figure 5**). On the other hand, hBtab\_MEAM1 possesses a gene lost in hBtab\_MED-Q1 (*elbB*) which could favor isoprenoid synthesis and thus enhance carotenoid production (Hemmi *et al.*, 1998). Thus, in the two *H. defensa* strains of *B. tabaci*, in addition to the common inability to synthesize heme, two strategies have emerged to preserve/enhance isoprenoid production (**Figure 5**).



**Figure 5: Isoprenoid fate in the hBtab clade.** The rectangle on the left represents the hBtab bacteria, the rectangle on the right *P. aleyrodidarum*. Dark arrows indicate biosynthetic pathways. Green arrows indicate metabolite exchanges. The red cross indicates that the two bacteria of the hBtab clade are unable to use isoprenoid to produce heme, and have to uptake the latter. In hBtab\_MEAM1, the dashed arrow illustrates the enhancement of the synthesis of isoprenoids by the *elbB* gene. hBtab\_MED1 depends on the chorismate furnished by *P. aleyrodidarum* to synthesize ubiquinones. Isoprenoids are produced by the two bacteria of the hBtab clade, and the surplus could be furnished to the other partner for the production of carotenoids.

## 5. Functional losses in hApis\_5AT

### a) Type 2 Secretion System (T2SS)

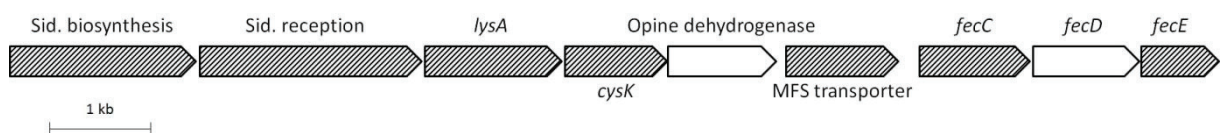
Contrary to the hBtab clade, hApis\_5AT has an inactive T2SS with inactivated *gspE* and *gspF* genes combined with a split of the T2SS in two regions of the genome. The T2SS system is also associated with virulence (Korotkov *et al.*, 2012), and controlled in some species by quorum-sensing (Sandkvist, 2001). In the closely related genus of *Yersinia*, T2SS has been much less studied than T3SS but could also be associated with virulence and cell invasion, and/or with survival outside of the mammalian hosts (von Tils *et al.*, 2012).

### b) Alpha-fimbriae synthesis

The fimbriae of the alpha-fimbriae family are involved in the adherence of enterotoxigenic *Escherichia coli* and the colonization of the human intestine (reviewed in Nuccio and Bäumlner, 2007). Several cryptic fimbriae are found in *E. coli*, and are differentially expressed depending on the environmental conditions. hBtab could have two such systems that are inactive in hApis\_5AT. Indeed, one pseudogene of hApis\_5AT is localized in a cluster of five genes identified as encoding proteins involved in the formation of alpha-fimbriae in *Yersinia pestis* CO92 (Parkhill *et al.*, 2001). This cluster would encode a fimbrial major subunit, a fimbrial minor subunit, two fimbrial chaperones and a fimbrial usher protein. The minimum required set of proteins is composed of the major subunit, the chaperone and the usher protein (Nuccio and Bäumlner, 2007), the latter being inactive in hApis\_5AT. A second cluster contains four genes that could be involved in another alpha-fimbriae formation. It would encode a protein with an unknown function, an usher protein, a major subunit and a chaperone. The two first genes, and therefore also the usher protein, are inactive in hApi\_5AT, leading to a non-functional system.

### c) Siderophore synthesis

hApis\_5AT underwent several inactivations of genes involved in iron uptake that are intact in hBtab. These inactivations include two genes involved in the synthesis of siderophore, a gene involved in the reception of siderophore, and two genes – *fecC* and *fecE* – annotated as iron transporters (Degnan *et al.*, 2009). All these genes are in the same island (**Figure 6**) which has only been characterized in the pathogen *Proteus mirabilis* (Himpsl *et al.*, 2010), where it synthesizes a newly described kind of siderophore, the proteobactin (*ptb*). The Fec-like iron transporter would serve to the ferri-siderophore uptake (i.e. the siderophore and the chelated iron; Himpsl *et al.*, 2010). This cluster is upregulated in iron-limiting conditions, but mutations in this cluster do not decrease the bacterium's ability to infect hosts (Himpsl *et al.*, 2010).



**Figure 6: Siderophore biosynthesis gene cluster in the hBtab clade.** The inactivated orthologs in hApis\_5AT are represented by hatches.

#### **d) Metabolism of amino acids and isoprenoids**

As seen in **Figure 6**, the *pbt* island contains two genes, *cysK* and *lysA*, involved in the metabolism of cysteine and lysine respectively, which are inactive in hApis\_5AT. A second *lysA* gene is present in hBtab, but is also inactive in hApis\_5AT. *B. aphidicola*, the P-symbiont associated with hApis\_5AT, is able to produce both cysteine and lysine. Thus, the inactivation of such genes in hApis\_5AT could be explained by ancestral metabolic redundancies. On the other hand, both hBtab bacteria retained the ability to synthesize these amino acids, which has been lost in *P. aleyrodidarum*.

#### **6. Common conservation and loss of virulence systems**

Two virulence systems, the Tad gene cluster and the T1SS, have undergone similar evolution in both clades. First, the Tad gene cluster that encodes a type IVb pilus notably involved in adhesion and host colonization (Kachlany *et al.*, 2001) has been independently inactivated by pseudogenization in both clades: *rcpA* in hBtab (plus *tadC* in hBtab\_MED-Q1) and *tadV* in hApis\_5AT. These independent inactivations would be one of the rare convergences observed between the hBtab clade and hApis\_5AT.

Second, the T1SS is intact both in hApis\_5AT (Degnan *et al.*, 2009) and in the hBtab clade. This system is involved in the secretion of the RTX toxins (Holland *et al.*, 2005), of which several copies are present in the *H. defensa* genomes (**Table 3**). While quite a number of different genes encode RTX toxins in all three genomes, many of them are pseudogenized. Considering the antique pool of RTX toxins probably present in the ancestor of the three genomes, some differential inactivations occurred (**Table 3**). The three bacteria only retained two common toxins, a putative serralyisin and *rtxA\_09*, which is fragmented in hApis\_5AT but whose largest fragment is longer than the threshold we used to characterize pseudogenes.

Gene	hApis_5AT	hBtab_MEAM1	hBtab_MED-Q1
HDEF_0400	Intact	Intact	Intact
rtxA_08 - rtxA_09	Intact (Fg)	Intact	Intact
rtxA_03	Intact	Contig	Contig
rtxA_27	Intact (Fg)	Intact	Contig
rtxA_14	Intact (Fg)	ψ	ψ
rtxA_07	Intact (Fg)	ψ	ψ
rtxA_16	Intact		
rtxA_22 to HDEF_1932	ψ	Intact	Intact
rtxA_28 - rtxA_29	ψ	Contig	Contig
rtxA_21	ψ	Contig	Contig
HBTMEAM1_v1_820005		Intact	ψ
rtxA_33		Contig	ψ
HBTMEAM1_v1_1390001		Contig	ψ
HDEF_2014 to rtxA_25	ψ	Intact	ψ
rtxA_30 - rtxA_32	ψ	ψ	Intact
rtxA_15	ψ	ψ	ψ
rtxA_17	ψ	ψ	ψ
HDEF_0620 to rtxA_05	ψ		
HDEF_0633 to rtxA_06	ψ		
HDEF_1149	ψ		
HDEF_1222	ψ		
HDEF_1264 to			
HDEF_1267	ψ		
HDEF_1389	ψ		
rtxA_01 - rtxA_02	ψ		
rtxA_10 to rtxA_12	ψ		
rtxA_13	ψ		
rtxA_18 to rtxA_20	ψ		
rtxA_26	ψ		

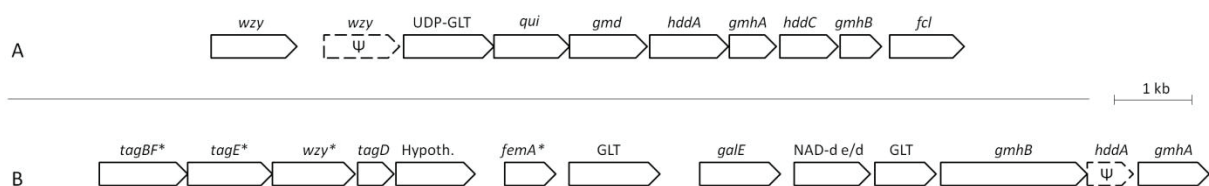
**Table 3: RTX toxins in the three genomes of *H. defensa*.** **Fg:** fragmented gene, but whose longest fragment is still considered active according to our parameter (see Material and Methods). **ψ:** pseudogene. **Contig:** gene with no visible deleterious mutation but not entire due to assembly. Grey rectangles indicate RTX toxin absent from a genome.

## 7. Rapid evolution and divergence of cell wall components

Cell wall components are highly different in hBtab and hApis\_5AT due to both specific gene losses, and the presence of two operons that probably originate from specific acquisitions.

This is especially visible as concerns sugar biosynthesis. hApis\_5AT harbors an entirely original gene cluster dedicated to the synthesis of cell wall components (HDEF\_0185

to HDEF\_0197, **Figure 7B**). It notably contains the *tagD* gene and two genes with *tagE* and *tagB/tagF* signatures. The *tag* gene cluster allows the synthesis of teichoic acids (TA) in Gram-positive bacteria, a heterogeneous class of phosphate-rich polymers covalently linked to the cell wall peptidoglycan. TAs may serve as a reservoir of phosphates to prevent starvation (Grant, 1979) and as phage receptors (Chatterjee, 1969). They also allow bacterial adhesiveness and antigenic power (Wicken and Knox, 1975). However, the genome of hApis\_5AT does not contain the gene *tagA* that is present in all described pathways of TA synthesis (reviewed in Swoboda *et al.*, 2010). Interestingly, the entire gene cluster of hApis\_5AT is found in *Arsenophonus nasoniae* (a wasp parasite) with the same gene organization. We can therefore make the assumption of an original gene pathway which differs from those found in Gram-positive bacteria. Apart from the above-mentioned gene cluster, hApis\_5AT possesses an active copy of *gne*, inactivated in hBtab, which encodes a UDP-N-acetylgalactosamine 4-epimerase that catalyzes the production of UDP-N-acetyl- $\alpha$ -D-galactosamine (GalNAc). GalNAc has been found in *Enterobacteriaceae* as a component of the O-antigen of the lipopolysaccharide, or LPS (Bengoechea *et al.*, 2002; Liu *et al.*, 2008), and in the teichoic acid of *Bacillus subtilis* (Young, 1966).



**Figure 7: Gene clusters involved in the biosynthesis of sugars and cell components specific to the hBtab clade (A) and hApis\_5AT (B).** The names of the genes are accompanied by an asterisk when the attribution is only based on the presence of conserved domains (and some syntenic clues for the *tag* genes). Pseudogenes are indicated by dashed lines and the  $\Psi$  letter. **GLT**: glycosyltransferase; **Hypoth.**: hypothetical; **NAD-d e/d**: NAD-dependent epimerase/dehydratase.

hBtab also possesses original genes for sugar biosynthesis. In particular, one hBtab operon may have been acquired through horizontal transfer, as no trace of orthologs genes has been found in hApis\_5AT. It contains two genes involved in the biosynthesis of GDP-L-fucose (L-Fuc) from GDP-D-mannose (*gmd* and *fcl*) (**Figure 7A**). The hBtab bacterium also encodes the *qui* gene, presumed to be involved in the biosynthesis of GDP-L-quinovose (L-Qui) from GDP-L-Fuc (De Castro *et al.*, 2012). Only two bacteria are known to possess L-Qui,

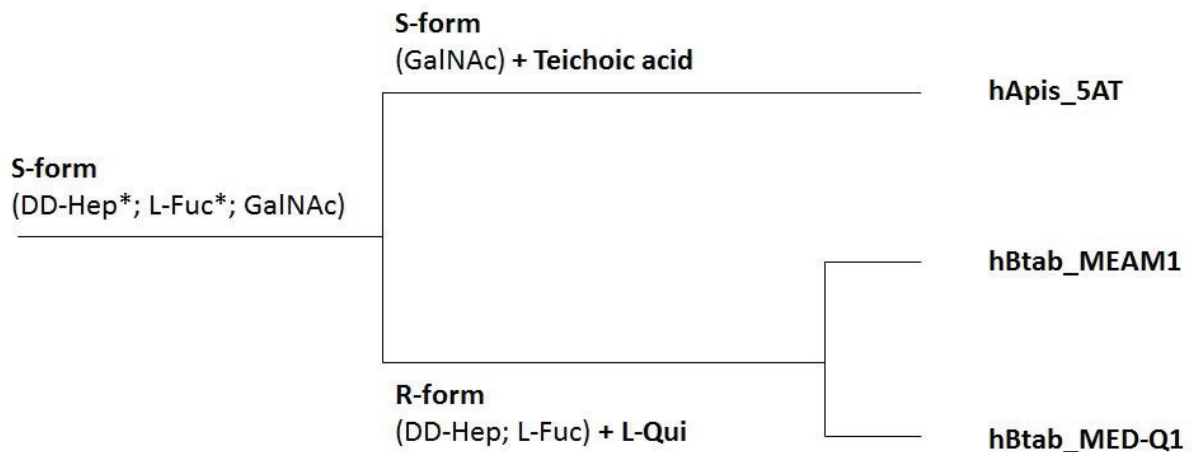


both in their O-antigen fraction: *Yersinia pseudotuberculosis* serotype O:12 (De Castro *et al.*, 2013) and *Providencia alcalifaciens* O:44 (Kocharova *et al.*, 2005). All the genes (*gmhA*, *gmhB*, *hddA* and *hddC*) allowing the synthesis of D-glycero-D-manno-heptose (DD-Hep) are also present in this operon (**Figure 7A**). While hApis\_5AT contains intact *gmhA* and *gmhB* genes (distantly related to the hBtab copies), it has an inactive *hddC* and no *hddA* genes, so that DD-Hep production is not possible in this bacterium.

Another potentially important difference between the two clades concerns the O-antigen, which is the more distal component of LPS, and is composed of different sugars linked together. It is a common factor of the Gram-negative bacteria that is important for host-pathogen interactions (Skurnik *et al.*, 2000). In addition to the pathways of sugar synthesis, which are extremely variable even between closely related strains, four genes are necessary to perform its assembly/export: *wzx*, *wzy*, *wzz* and *rfaL* (Skurnik *et al.*, 2000). In hApis\_5AT, we identified putative candidates for these four genes: (i) HDEF\_0181, annotated as “O-antigen conversion protein”, which possesses a weak similarity to the *wzx* flippase; (ii) HDEF\_0195, homolog of *wzy*; (iii) *fepE* (HDEF\_0537), which has a role analogue to *wzz* in *Salmonella typhimurium* (Murray *et al.*, 2003); and (iv) HDEF\_2072, which contains a *rfaL*-like O-antigen ligase domain. These results suggest that hApis\_5AT is able to produce O-antigen. As hApis\_5AT also contains genes involved in the synthesis of TA, it could, as other Gram-negative bacteria, possess a TA-like O-antigen (Jann *et al.*, 1980; Perepelov *et al.*, 2006).

In contrast, in hBtab, an inactivation occurred in the *rfaL*-like gene, which generally leads to rough-type bacteria (R-form), i.e. to bacteria without O-antigen. Moreover, in hBtab\_MED-Q1, no homolog of *wzy* was found. The presence of both O-antigen assembly genes and different pathways of sugar biosynthesis (**Figure 8**), suggests that the ancestor of *H. defensa* was an S-form bacterium, i.e. harboring an O-antigen, and that this O-antigen has been specifically lost in the hBtab clade. This result is surprising as hBtab is able to produce L-Qui, L-Fuc and DD-Hep, which are sugars of the O-antigen of Enterobacteriaceae, and especially of the *Yersinia* species (reviewed in Bruneteau and Minka, 2003). However, they are also found in other cell components, as could be the case in hBtab. For example, in *Y. pestis*, which is an R-type bacterium devoid of O-antigen, fucose is found in the antigen capsule that the bacterium forms when cultivated (Głównicka and Gruszkiewicz, 1980).

Fucose can also be found in the extracellular polysaccharides of the Rhizobiales bacteria (Gharzouli *et al.*, 2013), which are key elements for the symbiosis and are secreted into the cellular environment (González *et al.*, 1996). DD-Hep can be found in the core fraction of the LPS of *Yersinia* (Knirel and Anisimov, 2012) and many different species like the symbionts *Photorhabdus luminescens* (Kondakova *et al.*, 2012) and *Vibrio fischeri* (Post *et al.*, 2012).



**Figure 8: O-antigen and putative LPS-associated sugars in *H. defensa*.** The ancestor was a putative S-form bacterium, while the hBtab clade could group together R-form bacteria due to the inactivation of a *rfaL*-like gene. The sugars that generally bind to the O-antigen fraction are indicated in the corresponding branch and are detailed in the section “Rapid evolution and divergence of cell wall components”. Asterisks represent sugar pathways that could be ancestral (i.e. with close and best homologs in *Yersinia* species or in *R. insecticola*), but that could also have been horizontally acquired by the hBtab clade.

### III. Discussion

Recent studies have suggested that *H. defensa* could have a nutritional role in *B. tabaci*. Indeed, both genomic and experimental results sustain a metabolic role of this symbiont, which could occur through the complementation of the P-symbiont *P. aleyrodidarum* for both provision of vitamins and cofactors, but also for production of essential amino acids. Through comparison with the genome sequence of the defensive *H. defensa* infecting aphids, our goal was to study the impact of the ecology of the host-symbiont interaction (extended phenotype, within-host localization...) on the evolution of the gene content of the symbiont. The sequence of the *H. defensa* genome from *B. tabaci* MEAM1 allowed comparison with two other genomes of *H. defensa* in aphids and another species of *B. tabaci*, and to infer the gene content of their ancestor.

### **1. A Yersinia-like arsenal for a pathogenic ancestor**

The genome of hApis\_5AT already revealed an important number of virulence factors (Degnan *et al.*, 2009), and our analysis allowed extension of the characterization of the virulence islands present in the ancestor of *H. defensa*. The ancestral genome probably harbored at least three intact secretion systems (T1SS, T2SS, T3SS), several pili/fimbriae (type IV pili, alpha-fimbriae), different toxins and effectors, two extra-chromosomal partners (the APSE phage and a plasmid), different systems of environmental perception, a siderophore and a putative O-antigen linked to different sugars. Many of these systems have been well described in the *Yersinia* species and are significant virulence factors for this genus. This is in agreement with the theory that various virulence factors associated with pathogenic bacteria are part of the genomic repertoire of symbionts as they are necessary to overcome the host immune system (Hentschel *et al.*, 2000; Degnan *et al.*, 2009).

### **2. Divergence through gene losses**

An initial observation is that genome evolution has mainly been driven by gene loss in both lineages. Indeed, very few genes seem to have been acquired through horizontal gene transfer. This trend is even stronger in the hBtab clade, where the possibility of DNA uptake could be reduced, due to the loss and/or inactivation of the extra-chromosomal elements which participate in gene transfers. Importantly, the losses are highly divergent between the hBtab clade and hApis\_5AT, except for some RTX toxins and the Tad cluster. This suggests that the ancestor of *H. defensa* was already a symbiont having experienced genome reduction. Indeed, many studies have highlighted the fact that the ecological switch from free-living bacteria to a symbiotic lifestyle generally leads to convergences in the gene losses, due to a similar relaxation of the selective pressures (McCutcheon and Moran, 2010; Lamelas *et al.*, 2011). These divergent losses thus reflect the specialization of *H. defensa* to two different clades of insects – aphids and whiteflies.

Specialization of endosymbionts to different hosts has already been documented, especially concerning the well-documented *B. aphidicola* species. Charles *et al.* (2011) highlighted that the set of transporters in the genomes of *B. aphidicola* depends on the host plant use: the more specialized the host, the more devoid of transporters the symbiont

genome. Moreover, a comparative genomic study of four *Buchnera* strains, each associated with a different host, revealed 80% of divergent gene losses (Latorre *et al.*, 2005), in agreement with the retention of metabolic capabilities related to the host diet/life cycle. In *H. defensa*, we did not find a massive divergence of the metabolism of the bacteria, except that hApis\_5AT lost genes involved in the lysine and cysteine pathways. This could be the consequence of the metabolic ability of the associated P-symbiont, *B. aphidicola*, to produce these amino acids, in contrast to the situation in *B. tabaci*, where *P. aleyrodidarum* has lost part of these pathways. While only a few genes are involved in this metabolic difference between the two genomes, they might be extremely important for the global evolution of the system. Indeed, in *B. tabaci*, the contribution to the synthesis of amino acids might be one of the major contributions of the symbiont to the extended phenotype of the holobiont (i.e. the host and its symbiotic community). However, the most obvious divergence between the two genomes is related to specific losses of virulence factors.

### **3. Two bacteria, two arsenals**

As just mentioned, the majority of the genes that diverge between hApis\_5AT and the hBtab clades are classified as virulence factors. Virulence factors in *H. defensa* might be related to two different processes. First, they can be involved in the interaction with the host and notably the escape of the immune system by the symbiont. Second, they can target natural enemies of the host, especially in *H. defensa* that is a protective symbiont in aphids (Moran *et al.*, 2005b). Given that *H. defensa* could be a nutritional symbiont in whiteflies and that it colonizes only a single tissue devoted to symbiont hosting, attenuation of the virulence arsenal, which according to our reconstruction was already present in the *H. defensa* ancestors, was expected in this lineage.

A first indication of this attenuation is the loss of the two ancestral T3SS in hBtab, in contrast to their retention in hApis\_5AT. T3SS are probably used by hApis\_5AT to colonize the different tissues of the insect, but also to invade new hosts during the phases of horizontal transfers. Consistently, it has been shown that aphid-infecting *H. defensa* are able to resist phagocytosis in the hemolymph (Schmitz *et al.*, 2012). T3SS is the best candidate to confer this phenotype. More generally, T3SS is probably among the most pathogenic weapons of *Enterobacteriaceae*, and especially of the close *Yersinia* species (Cornelis, 2002).

In addition, almost all known S-symbionts use T3SS to establish symbiosis with their host (Dale and Moran, 2006). It has thus been proposed that the loss of T3SS could follow the evolution towards a mutualistic and obligate relationship (Dale *et al.*, 2005; Sloan and Moran, 2012a), even though some P-symbionts still possess a T3SS (Dale *et al.*, 2002). Another difference between the two *Hamiltonella* lineages concerns the composition of the cell wall. Not only the O-antigen would have been lost in the hBtab clade, but the lineages would have independently acquired different pathways for cell wall sugar synthesis. These variations may account for important differences in the immune dialog between the host and the symbiont. In particular, absence of the O-antigen could limit the ability of *H. defensa* to live outside the bacteriocytes in *B. tabaci*, as it may participate in the evasion of phagocytosis (Saldías *et al.*, 2009).

While the above examples highlight an attenuation of the virulence in hBtab, other systems have been conserved in this lineage, while lost in hApis5\_AT, such as the T2SS system, the production of siderophores and of fimbriae. However, in several symbiotic associations, “virulence” factors that are classically used for pathogenicity induction play an important role in the establishment and maintenance of symbiosis (Hentschel *et al.*, 2000). As such, T2SS is mandatory for gut colonization by the mutualist leech *Aeromonas veronii* (Maltz and Graf, 2011). The fact that few virulence factors are shared between the two lineages (except for the T1SS) makes it probable that communication between the host and the symbiont is highly different in the two systems.

Contrary to what is found in aphids, *H. defensa* in whiteflies does not possess an active APSE phage, suggesting it does not induce protection against parasitoids. However, the genes encoding for the CdtB toxin are still present in the hBtab clade and exhibit a perfect conservation of the important proteic residues (Degnan and Moran, 2008). If sequestration within the bacteriome does not preclude protection, as suggested in other symbiotic systems (Nakabachi *et al.*, 2013), secretion of this toxin could allow the targeting of parasitoids. Another possibility, explaining the maintenance of this toxin, is that in addition to its protective role, it could distend the host cytoplasm to facilitate the symbiont’s establishment (Moran *et al.*, 2005a).

Finally, even though virulence attenuation seems important in hBtab, it still conserves many virulence factors the function of which remains elusive. Two hypotheses could account for their maintenance. First, phylogenetic inertia can be sufficient to explain the presence of many virulence factors. Second, they might be involved in the molecular dialog between the host and the symbiont.

#### **4. Versatility and self-sufficiency**

Confinement to a single specialized cell-type, i.e. the bacteriocytes of whiteflies, was expected to favor loss of genes involved in environmental perception in the hBtab clade. The loss of genes encoding two-component systems and quorum-sensing in hBtab indeed reveals symbionts with weakened abilities to cope with environmental variations compared to hApis\_5AT. Strict intracellular localization can also explain the relaxation of selective pressure on pathways involved in the synthesis of elements available within the cell. Loss of the ability to synthesize heme could be such an example. Most of the accessible heme is indeed found intracellularly, and is thus directly available for the symbiont. In contrast, hApis\_5AT, which invades different tissues and the haemolymph, has retained the ability to synthesize it.

All these elements indicate that, compared to the ancestor and the actual hApis\_5AT, hBatb is much more dependent on the stable environment provided by the bacteriocyte and on metabolites procured by the host. Such increased dependence could allow tighter control of the symbiotic population by the host, in terms of both localization and density, as has been proposed for *B. aphidicola* in aphids (Thomas *et al.*, 2009; Shigenobu and Wilson, 2011).

#### **5. Mutualism and the evolution of evolvability**

Reductive evolution of P-symbionts has often been associated with the idea that these organisms have reduced evolvability (Silva *et al.*, 2003). A comparison of the genomes of the two lineages, both mutualistic but where the advantage provided by the symbiont is either nutritional or protective, provides an interesting opportunity to study how the extended phenotype constrains the evolution of evolvability.

In aphids, the protective phenotype provided by *H. defensa* requires the ability of the symbiont to both circulate within the host body and to kill the parasitoid, which maintains a high selective pressure on the virulence systems. In addition, because the symbiont is engaged in an arms race with the parasitoid, the maintenance of mechanisms allowing a rapid adaptation could also be under strong selection. The conservation of extra-chromosomal elements favoring horizontal gene transfers and the rapid turnover of genes encoding toxins and virulence effectors in the protective *H. defensa* of aphids support this view. In agreement with the arms race theory, the different APSE phages have been involved in the exchange of toxins between close strains of *H. defensa* in aphids (Moran *et al.*, 2005a; Degnan and Moran, 2008). All these elements may have different evolutionary consequences. First, maintenance and turnover of virulence factors required for the protective effect may avoid a complete reduction of the infection cost for the host, which could also be associated with the impossibility to reach high frequency for the symbiont. Second, the versatility of the symbiont may allow it to regularly acquire new hosts through horizontal transmission.

In contrast, the transition towards nutritional symbiosis in whiteflies might have considerably modified the selective pressures acting on the symbiont. Indeed, nutritional provision must have been associated with a dramatic increase in the prevalence of the symbiont within populations, thus increasing the selection on the holobiont for a reduction in the infection cost. Such reduction may occur through the loss of virulence factors and/or through the restriction of the localization of the symbiont within the host. Importantly, this could also be associated with the loss of the protective phenotype if it was ancestral. Both the reduction of the arms race with the parasitoid and the confinement within bacteriocytes could have then led to the degradation of extra-chromosomal elements. Confinement can also favor the evolution of an increased dependency on host metabolites. All these elements would rapidly lead to organisms with highly reduced evolvability and ability to switch hosts.

#### **IV. Conclusions**

Both phylogenetic and phylogenomic analyses revealed a clear separation of the *H. defensa* of *B. tabaci* and aphids. This is congruent with the genome content of the bacteria,

with specific gene acquisitions and losses. Interestingly, the evolution is mainly driven by gene inactivations and losses, while horizontal gene transfers seem restricted to some toxins, effectors and sugars linked to the cell wall. The different gene inactivations that occurred in the symbionts of *A. pisum*, on one side, and *B. tabaci*, on the other, can reflect the specialization of the bacteria in two distinct host clades as well as the different phenotypes and life conditions. Indeed, hApis\_5AT would need some highly virulent factors, like T3SS, and different environmental perceptors, in order to invade host tissues, transfer to the next generation and protect the host against parasitoids. These systems would be obsolete in the bacteriome-associated symbionts of the hBtab clade. Nevertheless, the latter retained different virulence factors lost in hApis\_5AT. Because a protective phenotype has not yet been demonstrated in the hBtab clade, these factors could have an important role for the communication of *H. defensa* with the host and/or the primary endosymbiont, and/or to settle in the host cells. Globally, while both lineages have mutualistic interactions with their hosts, the evolution of their genomes shows different trends that could have important consequences on the evolvability of these organisms.

## **V. Material and methods**

### **1. Whitefly strains**

A population of *B. tabaci* MEAM1 species reared in laboratory conditions at Newe Ya'ar Research Center (Israel) for many years was used for this study. The population was kept on cotton plants inside insect growth chambers (14h–10h light–dark photoperiod). This strain harbored the P-symbiont *P. aleyrodidarum* as well as the S-symbionts *H. defensa* and *Rickettsia* sp.

### **2. DNA extraction and amplification**

Approximately 2,000 adult whiteflies were collected. A modified endosymbiont enrichment extraction method was used (Harrison *et al.*, 1989). Whiteflies were homogenized with a Dounce homogenizer in cold Ringer-Krebs buffer (Sigma-Aldrich). The homogenate was passed through a set of decreasing sized porous nylon membranes (1 mm, 80 µm, 60 µm, 20 µm, 10 µm and 5 µm) (4°C), centrifuged at 8,000 r.p.m. for 15 minutes (4°C) and then endosymbiont pellet was washed with Ringer-Krebs buffer (three times). The



pellet containing the endosymbionts were resuspended in 250  $\mu$ l of Ringer-Krebs buffer and subjected to DNase digestion (50  $\mu$ l of 10X DNase I Buffer and 5  $\mu$ l DNase I) to degrade the host genomic DNA (Harrison *et al.*, 1989), following the manufacturer's instructions (TURBO™ DNase, Life Technologies). The final endosymbiont enriched sample was centrifuged at 8,000 r.p.m. for 15 minutes (4°C) and washed. The pellet was used for DNA extraction with the JETFLEX Genomic DNA Purification Kit (Genomed) following the manufacturer's instructions.

GenomiPhi V2 (GE Healthcare) was used for genomic amplification following the manufacturer's recommendations. In order to diminish chimera biases in sequencing data, six independent amplifications were performed and 1  $\mu$ l of DNA (< 10 ng/ $\mu$ l) was used as input for each reaction. Amplified DNA reactions were pooled and used to construct different sequencing libraries.

### **3. Assembly and annotation**

Two libraries were constructed at Genoscope (France): a 454 paired-end library and an Illumina HiSeq 2000 (100 nt). The first assembly was made with mira4 (Chevreux *et al.*, 1999), using the two libraries (both 454 and Illumina reads). Blast searches against the NCBI databases were performed for each contig to identify which organism it belonged to. Reads coming from other sources (the host *B. tabaci*, the mitochondria, the other symbionts *P. aleyrodidarum*, *Wolbachia sp.* and *Rickettsia sp.*) were removed. Then, three assemblies were performed keeping only the reads from *H. defensa*, with three levels of stringency. CISA (Lin and Liao, 2013) was then used to merge the three obtained assemblies. This step significantly improved the length and gene content of the assembly (data not shown). Mapping was performed with mira4 and soap2 on the final assembly in order to check the quality of the CISA merging. Contig editing and visualization were performed using Gap5 (Bonfield and Whitwham, 2010) and Tablet (Milne *et al.*, 2010).

The assembly was deposited on the MaGe platform in order to be annotated. Genes coding for proteins with a length inferior to 100 amino acids were considered as being artifacts if they had no significant matches against MaGe databases. For the rest, the annotations coming from Swissprot were privileged. Syntenic information was important for

inference of the presence of some islands like operons dedicated to the synthesis of fimbriae or siderophores (see Results). The detected CDSs were annotated as pseudogenes when their length was less than 80% of the one of their orthologs in the other genomes, as suggested by Lerat and Ochman (2005), except if all the ancestral domains were preserved. The inactivations and losses specific to our strain of *H. defensa* were checked using SOAPaligner/soap2 software and the Illumina data. The genes broken by contig edges and with no apparent stop mutation were considered as active. The assembly and annotations are available at the European Nucleotide Archive under accession number PRJEB7127.

#### 4. Comparative genomics and outgroup genome

We propose a host-based nomenclature to denote the three sequenced genomes of *H. defensa* used in this study (**Table 4**). This nomenclature concatenated the “h” of *Hamiltonella* with the different information about the host taxonomy. The genome we sequenced will thus be referred to as hBtab\_MEAM1 and the one infecting the MED species of *B. tabaci* as hBtab\_MED-Q1 (Rao *et al.*, 2012). We called hApis\_5AT the first genome of *Candidatus H. defensa* from *A. pisum* almost entirely sequenced (chromosome with gaps) (Degnan *et al.*, 2009). The two genomes of *H. defensa* of *B. tabaci* will be referred to as the hBtab clade, based on our phylogenetic analysis (see above in the Material and Methods, but also in the Results part).

Reference	Host genus	Host species	Strain (Cytotype)	Nomenclature
Degnan <i>et al.</i> (2009)	<i>Acyrtosiphon</i> (A)	<i>pisum</i> (pis)	5AT	hApis_5AT
Rao <i>et al.</i> (2012)	<i>Bemisia</i> (B)	<i>tabaci</i> (tab)	MED (Q1)	hBtab_MED-Q1
This study	<i>Bemisia</i> (B)	<i>tabaci</i> (tab)	MEAM1	hBtab_MEAM1

**Table 4: Nomenclature for the *H. defensa* genomes.**

To study the evolution of the genome in *H. defensa*, several related bacteria were used as outgroups: two draft genomes of *R. insecticola* (Degnan *et al.*, 2010; Hansen *et al.*, 2012), and the genomes of *Yersinia enterocolitica* strain 8081 (Thomson *et al.*, 2006), *Serratia proteamaculans* and *E. coli* strain K12 (Blattner *et al.*, 1997). These genomes represent different levels of divergence with the genus *H. defensa*, all belonging to Gammaproteobacteria: *R. insecticola* is a clade of facultative endosymbionts and the sister species of *H. defensa*, and the three other outgroups are free-living bacteria.

## 5. Gene families

Blastall was used to blast each protein of a given genome against all the proteins of the seven other genomes. First, large families were built by putting a given protein in the same family of its best hit in each organism, and using a transitive relation (i.e. the respective clustering of A and B on one hand, and A and C on the other hand, means that B and C are in the same family). Thus, two proteins coming from the same genome can be placed in the same family. It allowed the following placements in each family: (i) the fragments of the same protein; (ii) the proteins coming from recent duplication events; and (iii) the different paralogs coming both from lateral gene transfers and old duplication events. MUSCLE software (Edgar, 2004) was used to perform an alignment of the proteins of each family; PhyML v3.0 (Guindon *et al.*, 2010) was used with default parameter to build a rapid family tree. These large families were then manually curated, based on several types of information such as alignments and trees. This allowed the obtainment of more accurate orthologs, and gave a precise idea of the different gene histories, which helped in annotating hBtab\_MEAM1. Pseudogenes were defined in families as truncated genes with a length of less than 80% of the one of their homologs. Because we focused on the genomic history of the genus *Hamiltonella*, the start codons of the *H. defensa* pseudogenes were checked when the truncation was in the 5' region in order to eliminate any false positives. For the putative pseudogenes shared by the three *H. defensa* strains, the seqinR package (R) was used to calculate the Ka/Ks ratio between hApis\_5AT and the hBtab bacteria in order to determine if they were still under purifying selection. It was not possible to determine this ratio between the two hBtab bacteria, because of their similarity (close to 100%).

## 6. Reconstruction of the ancestor of the genus *Hamiltonella*

The ancestor of the genus *H. defensa* was inferred based on the gene families. A gene was considered as present in the ancestor in two distinct conditions: (i) if the gene is present in both the hBtab clade and in hApis\_5AT, and active in at least one of the three genomes of *H. defensa* (a gene present in an inactive state in the three genomes was considered as ancestrally inactivated); (ii) if a gene was intact in either hApis\_5AT or the hBtab clade, and present in at least one outgroup genome. This definition means that a transfer between one of the genomes used as outgroup and a genome of *H. defensa* would be erroneously

considered as ancestral. To avoid such problems, the mobile elements (phages, plasmidic islands, toxin/antitoxin systems, transposases) were removed when studying the ancestor, except for the APSE phage and the genes belonging to the non-integrating plasmid of hApis\_5AT.

## 7. Phylogeny of the genus *Hamiltonella*

To go further, we built the phylogeny of the genus *Hamiltonella* to determine the relationship between *H. defensa* of whiteflies and aphids. Eight genes (*accD*, *dnaA*, *gyrB*, *hrpA*, *murE*, *ptsI*, *recJ*, *rpoS*) have already been sequenced in 19 *H. defensa* of aphids and in one *H. defensa* of the MEAM1 species of *B. tabaci* (Telesnicki *et al.*, 2012). Among these eight genes, six are 100% identical to their respective ortholog in the genome of hBtab\_MEAM1 sequenced in this study (*accD*, *dnaA*, *gyrB*, *hrpA*, *murE*, *ptsI*), and five of these six genes are 100% identical to their orthologs in hBtab\_MED-Q1. Moreover, these six genes are clearly distinct from the orthologs of the *H. defensa* of aphids. The two remaining genes (*recJ* and *rpoS*) exhibit a different pattern: in the MEAM1 strain of *H. defensa* sequenced in Telesnicki *et al.* (2012), they are totally identical to their orthologs in some *H. defensa* of aphids, and distinct from their orthologs in the genome of the two assembled genomes of *H. defensa* of *B. tabaci* (hBtab\_MEAM1 and hBtab\_MED-Q1). These two genes are unlikely to be submitted to frequent transfers (see for example their individual HOGENOM gene trees HOG000018414 and HOG000270273, Penel *et al.*, 2009), and the similarity between the *H. defensa* of *B. tabaci* and aphid species for these two loci could reflect sequencing issues rather than horizontal transfers. In all cases, they probably do not reflect the phylogenetic history of the species, and were thus removed from the analysis.

Thus, six genes were used to study the phylogenetic relationships between the genomes of *H. defensa* in aphids and *B. tabaci* (*accD*, *dnaA*, *gyrB*, *hrpA*, *murE*, *ptsI*). We used both the sequences from Telesnicki *et al.* (2012), to which we added the corresponding sequences in the three assembled genomes of *H. defensa*, and two outgroups (*R. insecticola* strain 5.15 and *Y. enterocolitica* strain 8081). Genes were individually considered and then concatenated.

For each set of genes, alignments were initially generated using the MUSCLE software (Edgar, 2004) and then implemented in CLC Main Workbench v6.7.1 (CLC Bio). Phylogenetic analyses were performed using maximum likelihood inference with PhyML v3.0 (Guindon *et al.*, 2010). The appropriate model of evolution was evaluated with jModeltest v0.1.1 (Posada, 2008) for each set of sequences and the best likelihood score was evaluated with the Akaike information criterion for the concatenated data sets and AIC corrected for small sample size for data of gene separately. The models selected were TPM+I+G for *accD*, TVM+I+G for *gyrB*, TVM+G for *hrpA* and *dnaA*, TPM3+G for *murE*, TIM3+G for *ptsI*, and GTR+G for the concatenated data sets. Robustness of the nodes was assessed with 100 bootstrap replicates. Finally, the trees were edited with Figtree v1.4.0 (A. Rambaut, <http://tree.bio.ed.ac.uk/software/figtree>). *Y. enterocolitica* strain 8081 was used to root the tree.

## ACKNOWLEDGMENTS

These results were achieved within the framework of the first call on Mediterranean agriculture carried out by the Agricultural Research In the Mediterranean Area Network, with funding from Ministry of Agriculture and Rural Development (IL), Agence Nationale de la Recherche (FR), Instituto Nacional de Investigaciones Agrícolas (ES), National Agricultural Research Foundation - DEMETER (GR) and General Directorate of Agricultural Research and Policy (TR). This work was supported by grants from the National Natural Science Foundation of China (Project 31321063) and BFU2012-39816-C02-01 (co-financed by FEDER funds and Ministerio de Economía y Competitividad, Spain) to A. Latorre. P. A. Rollat-Farnier is a recipient of a grant from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement n° [247073]10 SISYPHE. Santos-Garcia D. was recipient of a Prometeo 92/2009 contract. We thank the Genoscope (Evry, France) for sequencing the 454 and Illumina libraries.

## REFERENCE

- Andersson K., Carballeira N., Magnusson K. E., Persson C., Stendahl O., Wolf-Watz H., et Fällman M. **1996**. YopH of *Yersinia pseudotuberculosis* interrupts early phosphotyrosine signalling associated with phagocytosis. *Molecular Microbiology*, **20**, 1057-1069.
- Baumann P. **2005**. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annual Review of Microbiology*, **59**, 155-189.
- Bengoechea J.A., Pinta E., Salminen T., Oertelt C., Holst O., Radziejewska-Lebrecht J. et al. **2002**. Functional characterization of Gne (UDP-N-acetylglucosamine-4-epimerase), Wzz (chain length determinant), and Wzy (O-antigen polymerase) of *Yersinia enterocolitica* serotype O:8. *Journal of Bacteriology*, **184**, 4277-4287.
- Blattner F.R., Plunkett G., Bloch C.A., Perna N.T., Burland V., Riley M. et al. **1997**. The complete genome sequence of *Escherichia coli* K-12. *Science*. **277**, 1453-1462.
- Boykin L.M., Bell C.D., Evans G., Small I. et De Barro P.J. **2013**. Is agriculture driving the diversification of the *Bemisia tabaci* species complex (Hemiptera: Sternorrhyncha: Aleyrodidae)? Dating, diversification and biogeographic evidence revealed. *BMC Evolutionary Biology*, **13**, 228.
- Bonfield J.K. et Whitwham A. **2010**. Gap5—editing the billion fragment sequence assembly. *Bioinformatics*, **26**, 1699-1703.
- Bruneteau M. et Minka S. **2003**. Lipopolysaccharides of bacterial pathogens from the genus *Yersinia*: a mini-review. *Biochimie*, **85**, 145-152.
- Buchner P. **1965**. *Endosymbiosis of animals with plant microorganisms*. New York: John Wiley.
- Charles H., Balmand S., Lamelas A., Cottret L., Pérez-Brocal V., Burdin B. et al. **2011**. A genomic reappraisal of symbiotic function in the aphid/*Buchnera* symbiosis: reduced transporter sets and variable membrane organisations. *PLoS One*, **6**, e29096.
- Chatterjee A.N. **1969**. Use of bacteriophage-resistant mutants to study the nature of the bacteriophage receptor site of *Staphylococcus aureus*. *Journal of Bacteriology*, **98**, 519-527.
- Cheng Q., Ruel T.D., Zhou W., Moloo S.K., Majiwa P., O'Neill S.L. et al. **2000**. Tissue distribution and prevalence of *Wolbachia* infections in tsetse flies, *Glossina* spp. *Medical and Veterinary Entomology*, **14**, 44-50.
- Chevreur B., Wetter T. et Suhai S. **1999**. Genome sequence assembly using trace signals and additional sequence information. In *German Conference on Bioinformatics*, 45-56.
- Chiel E., Zchori-Fein E., Inbar M., Gottlieb Y., Adachi-Hagimori T., Kelly S.E. et al. **2009**. Almost there: transmission routes of bacterial symbionts between trophic levels. *PLoS One*, **4**, e4767.
- Cornelis G.R. **2002**. *Yersinia* type III secretion send in the effectors. *Journal of Cell Biology*, **158**, 401-408.
- Dale C., Plague G.R., Wang B., Ochman H. et Moran N.A. **2002**. Type III secretion systems and the evolution of mutualistic endosymbiosis. *Proceedings of the National Academy of Sciences*, **99**, 12397-12402.
- Dale C., Jones T. et Pontes M. **2005**. Degenerative evolution and functional diversification of type-III secretion systems in the insect endosymbiont *Sodalis glossinidius*. *Molecular Biology and Evolution*, **22**, 758-766.
- Dale C. et Moran N.A. **2006**. Molecular interactions between bacterial symbionts and their hosts. *Cell*, **126**, 453-465.
- Darby A.C., Birkle L.M., Turner S.L. et Douglas A.E. **2001**. An aphid-borne bacterium allied to the secondary symbionts of whitefly. *FEMS Microbiology Ecology*, **36**, 43-50.

- De Castro C., Kenyon J.J., Cunneen M.M., Molinaro A., Holst O., Skurnik M. *et al.* **2013**. The O-specific polysaccharide structure and gene cluster of serotype O:12 of the *Yersinia pseudotuberculosis* complex, and the identification of a novel l-quinovose biosynthesis gene. *Glycobiology*, **23**, 346-353.
- Degnan P.H. et Moran N.A. **2008**. Diverse phage-encoded toxins in a protective insect endosymbiont. *Applied and Environmental Microbiology*, **74**, 6782-6791.
- Degnan P.H., Yu Y., Sisneros N., Wing R.A. et Moran N.A. **2009**. *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proceedings of the National Academy of Sciences*, **106**, 9063-9068.
- Degnan P.H., Leonardo T.E., Cass B.N., Hurwitz B., Stern D., Gibbs R.A. *et al.* **2010**. Dynamics of genome evolution in facultative symbionts of aphids. *Environmental Microbiology*, **12**, 2060-2069.
- Edgar R.C. **2004**. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792-1797.
- Fukatsu T., Nikoh N., Kawai R. et Koga R. **2000**. The secondary endosymbiotic bacterium of the pea aphid *Acyrtosiphon pisum* (Insecta: Homoptera). *Applied and Environmental Microbiology*, **66**, 2748-2758.
- Gharzouli R., Carpené M.A., Couderc F., Benguedouar A. et Poinot V. **2013**. Relevance of fucose-rich extracellular polysaccharides produced by *Rhizobium sultae* strains nodulating *Hedysarum coronarium* L. legumes. *Applied and Environmental Microbiology*, **79**, 1764-1776.
- Gil R., Silva F.J., Peretó J. et Moya A. **2004**. Determination of the core of a minimal bacterial gene set. *Microbiology and Molecular Biology Reviews*, **68**, 518-537.
- Głosnicka R. et Gruszkiewicz E. **1980**. Chemical composition and biological activity of the *Yersinia pestis* envelope substance. *Infection and Immunity*, **30**, 506-512.
- González J.E., Reuhs B.L. et Walker G.C. **1996**. Low molecular weight EPS II of *Rhizobium meliloti* allows nodule invasion in *Medicago sativa*. *Proceedings of the National Academy of Sciences*, **93**, 8636-8641.
- Gottlieb Y., Ghanim M., Gueguen G., Kontsedalov S., Vavre F., Fleury F. *et al.* **2008**. Inherited intracellular ecosystem: symbiotic bacteria share bacteriocytes in whiteflies. *The FASEB Journal*, **22**, 2591-2599.
- Grant W.D. **1979**. Cell wall teichoic acid as a reserve phosphate source in *Bacillus subtilis*. *Journal of Bacteriology*, **137**, 35-43.
- Grosdent N., Maridonneau-Parini I., Sory M.P. et Cornelis G.R. **2002**. Role of Yops and adhesins in resistance of *Yersinia enterocolitica* to phagocytosis. *Infection and Immunity*, **70**, 4165-4176.
- Gueguen G., Vavre F., Gnankine O., Peterschmitt M., Charif D., Chiel E. *et al.* **2010**. Endosymbiont metacommunities, mtDNA diversity and the evolution of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. *Molecular Ecology*, **19**, 4365-4378.
- Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W. et Gascuel O. **2010**. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307-321.
- Hansen A.K., Vorburger C. et Moran N.A. **2012**. Genomic basis of endosymbiont-conferred protection against an insect parasitoid. *Genome Research*, **22**, 106-114.
- Harrison C.P., Douglas A.E. et Dixon A.F.G. **1989**. A rapid method to isolate symbiotic bacteria from aphids. *Journal of Invertebrate Pathology*, **53**, 427-428.



- Hemmi H., Ohnuma S.I., Nagaoka K. et Nishino T. **1998**. Identification of genes affecting lycopene formation in *Escherichia coli* transformed with carotenoid biosynthetic genes: candidates for early genes in isoprenoid biosynthesis. *Journal of Biochemistry*, **123**, 1088-1096.
- Hentschel U., Steinert M. et Hacker J. **2000**. Common molecular mechanisms of symbiosis and pathogenesis. *Trends in Microbiology*, **8**, 226-231.
- Himpl S.D., Pearson M.M., Arewång C.J., Nusca T.D., Sherman D.H. et Mobley H.L. **2010**. Proteobactin and a yersiniabactin-related siderophore mediate iron acquisition in *Proteus mirabilis*. *Molecular Microbiology*, **78**, 138-157.
- Holland I.B., Schmitt L. et Young J. **2005**. Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway. *Molecular Membrane Biology*, **22**, 29-39.
- Jann K., Jann B., Schmidt M.A. et Vann W.F. **1980**. Structure of the *Escherichia coli* K2 capsular antigen, a teichoic acid-like polymer. *Journal of Bacteriology*, **143**, 1108-1115.
- Kachlany S.C., Planet P.J., DeSalle R., Fine D.H. et Figurski D.H. **2001**. Genes for tight adherence of *Actinobacillus actinomycetemcomitans*: from plaque to plague to pond scum. *Trends in Microbiology*, **9**, 429-437.
- Knirel Y.A. et Anisimov A.P. **2012**. Lipopolysaccharide of *Yersinia pestis*, the cause of plague: structure, genetics, biological properties. *Acta Naturae*, **4**, 46.
- Kocharova N.A., Ovchinnikova O.G., Toukach F.V., Torzewska A., Shashkov A.S., Knirel Y.A. et al. **2005**. The O-polysaccharide from the lipopolysaccharide of *Providencia stuartii* O44 contains l-quinovose, a 6-deoxy sugar rarely occurring in bacterial polysaccharides. *Carbohydrate Research*, **340**, 1419-1423.
- Kondakova A.N., Kirsheva N.A., Shashkov A.S., Shaikhutdinova R.Z., Arbatsky N.P., Ivanov S.A. et al. **2012**. Structure of the O-polysaccharide of *Photobacterium luminescens* subsp. *laumondii* containing D- glycerol-3-phospho-D-manno-heptose and 3,6-dideoxy-3-formamido-D-glucose. *Carbohydrate Research*, **351**, 134-137.
- Korotkov K.V., Sandkvist M. et Hol W.G. **2012**. The type II secretion system: biogenesis, molecular architecture and mechanism. *Nature Reviews Microbiology*, **10**, 336-351.
- Lamelas A., Pérez-Brocal V., Gómez-Valero L., Gosalbes M.J., Moya A. et Latorre A. **2008**. Evolution of the secondary symbiont "*Candidatus Serratia symbiotica*" in aphid species of the subfamily Lachninae. *Applied and Environmental Microbiology*, **74**, 4236-4240.
- Lamelas A., Gosalbes M.J., Manzano-Marín A., Peretó J., Moya A. et Latorre A. **2011**. *Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont. *PLoS Genetics*, **7**, e1002357.
- Lapouge K., Schubert M., Allain F.H. et Haas D. **2008**. Gac/Rsm signal transduction pathway of gamma-proteobacteria: from RNA recognition to regulation of social behaviour. *Molecular Microbiology*, **67**, 241-253.
- Latorre A., Gil R., Silva F.J. et Moya A. **2005**. Chromosomal stasis versus plasmid plasticity in aphid endosymbiont *Buchnera aphidicola*. *Heredity*, **95**, 339-347.
- Lerat E., Daubin V. et Moran N.A. **2003**. From gene trees to organismal phylogeny in prokaryotes: the case of the  $\gamma$ -Proteobacteria. *PLoS Biology*, **1**, e19.
- Lerat E. et Ochman H. **2005**. Recognizing the pseudogenes in bacterial genomes. *Nucleic Acids Research*, **33**, 3125-3132.
- Lin S.H. et Liao Y.C. **2013**. CISA: contig integrator for sequence assembly of bacterial genomes. *PLoS One*, **8**, e60843.



- Liu B., Knirel Y.A., Feng L., Perepelov A.V., Senchenkova S.Y.N., Wang Q. *et al.* **2008**. Structure and genetics of *Shigella* O antigens. *FEMS Microbiology Reviews*, **32**, 627-653.
- López-Madrugal S., Latorre A., Porcar M., Moya A. *et al.* **2011**. Complete genome sequence of “*Candidatus Tremblaya princeps*” strain PCVAL, an intriguing translational machine below the living-cell status. *Journal of Bacteriology*, **193**, 5587-5588.
- Lu S., Killoran P.B., Fang F.C. *et al.* **2002**. The global regulator ArcA controls resistance to reactive nitrogen and oxygen intermediates in *Salmonella enterica* serovar Enteritidis. *Infection and Immunity*, **70**, 451-461.
- Maltz M. *et al.* **2011**. The type II secretion system is essential for erythrocyte lysis and gut colonization by the leech digestive tract symbiont *Aeromonas veronii*. *Applied and Environmental Microbiology*, **77**, 597-603.
- Manzano-Marín A. *et al.* **2014**. Settling down: The genome of *Serratia symbiotica* from the aphid *Cinara tujafilina* zooms in on the process of accommodation to a cooperative intracellular life. *Genome Biology and Evolution*, **6**, 1683-1698.
- Martinez-Antonio A. *et al.* **2003**. Identifying global regulators in transcriptional regulatory networks in bacteria. *Current Opinion in Microbiology*, **6**, 482-489.
- McCutcheon J.P. *et al.* **2010**. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biology and Evolution*, **2**, 708-718.
- Milne I., Bayer M., Cardle L., Shaw P., Stephen G., Wright F. *et al.* **2010**. Tablet—next generation sequence assembly visualization. *Bioinformatics*, **26**, 401-402.
- Moran N.A. **1996**. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences*, **93**, 2873-2878.
- Moran N.A., Degnan P.H., Santos S.R., Dunbar H.E. *et al.* **2005a**. The players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. *Proceedings of the National Academy of Sciences*, **102**, 16919-16926.
- Moran N.A., Russell J.A., Koga R. *et al.* **2005b**. Evolutionary relationships of three new species of *Enterobacteriaceae* living as symbionts of aphids and other insects. *Applied and Environmental Microbiology*, **71**, 3302-3310.
- Moran N.A., McCutcheon J.P. *et al.* **2008**. Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics*, **42**, 165-190.
- Moya A., Pereto J., Gil R. *et al.* **2008**. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nature Reviews Genetics*, **9**, 218-229.
- Murray G.L., Attridge S.R. *et al.* **2003**. Regulation of *Salmonella typhimurium* lipopolysaccharide O antigen chain length is required for virulence; identification of FepE as a second Wzz. *Molecular Microbiology*, **47**, 1395-1406.
- Nakabachi A., Ueoka R., Oshima K., Teta R., Mangoni A., Gurgui M. *et al.* **2013**. Defensive bacteriome symbiont with a drastically reduced genome. *Current Biology*, **23**, 1478-1484.
- Niemann G.S., Brown R.N., Gustin J.K., Stufkens A., Shaikh-Kidwai A.S., Li J. *et al.* **2011**. Discovery of novel secreted virulence factors from *Salmonella enterica* serovar Typhimurium by proteomic analysis of culture supernatants. *Infection and Immunity*, **79**, 33-43.
- Nogge G. **1982**. Significance of symbionts for the maintenance of an optional nutritional state for successful reproduction in hematophagous arthropods. *Parasitology*, **82**, 101-104.

- Nuccio S.P. et Bäumler A.J. **2007**. Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. *Microbiology and Molecular Biology Reviews*, **71**, 551-575.
- Ochman H. et Moran N.A. **2001**. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science*, **292**, 1096-1099.
- Okan N.A., Bliska J.B. et Karzai A.W. **2006**. A role for the SmpB-SsrA system in *Yersinia pseudotuberculosis* pathogenesis. *PLoS Pathogens*, **2**, e6.
- Okan N.A., Mena P., Benach J.L., Bliska J.B. et Karzai A.W. **2010**. The smpB-ssrA mutant of *Yersinia pestis* functions as a live attenuated vaccine to protect mice against pulmonary plague infection. *Infection and Immunity*, **78**, 1284-1293.
- Orth K. **2002**. Function of the *Yersinia* effector YopJ. *Current Opinion in Microbiology*, **5**, 38-43.
- Pan H., Li X., Ge D., Wang S., Wu Q., Xie W. et al. **2012**. Factors affecting population dynamics of maternally transmitted endosymbionts in *Bemisia tabaci*. *PLoS One*, **7**, e30760.
- Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G., Prentice M.B. et al. **2001**. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature*, **413**, 523-527.
- Penel S., Arigon A.M., Dufayard J.F., Sertier A.S., Daubin V., Duret L. et al. **2009**. Databases of homologous gene families for comparative genomics. *BMC Bioinformatics*, **10**, S3.
- Perepelov A.V., Wang Q., Senchenkova S.Y.N., Shevelev S.D., Zhao G., Shashkov A.S. et al. **2006**. Structure of a teichoic acid-like O-polysaccharide of *Escherichia coli* O29. *Carbohydrate Research*, **341**, 2176-2180.
- Pérez-Brocal V., Gil R., Ramos S., Lamelas A., Postigo M., Michelena J.M. et al. **2006**. A small microbial genome: the end of a long symbiotic relationship? *Science*, **314**, 312-313.
- Posada D. **2008**. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253-1256.
- Post D.M., Yu L., Krasity B.C., Choudhury B., Mandel M.J., Brennan C.A. et al. **2012**. O-antigen and core carbohydrate of *Vibrio fischeri* lipopolysaccharide composition and analysis of their role in *Euprymna scolopes* light organ colonization. *Journal of Biological Chemistry*, **287**, 8515-8530.
- Rao Q., Wang S., Su Y.L., Bing X.L., Liu S.S. et Wang X.W. **2012a**. Draft genome sequence of “*Candidatus* Hamiltonella defensa,” an endosymbiont of the whitefly *Bemisia tabaci*. *Journal of Bacteriology*, **194**, 3558-3558.
- Russell J.A., Latorre A., Sabater-Muñoz B., Moya A. et Moran N.A. **2003**. Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. *Molecular Ecology*, **12**, 1061-1075.
- Saldías M.S., Ortega X. et Valvano M.A. **2009**. *Burkholderia cenocepacia* O antigen lipopolysaccharide prevents phagocytosis by macrophages and adhesion to epithelial cells. *Journal of Medical Microbiology*, **58**, 1542-1548.
- Sandkvist M. **2001**. Biology of type II secretion. *Molecular Microbiology*, **40**, 271-283.
- Santos-Garcia D., Farnier P.A., Beitia F., Zchori-Fein E., Vavre F., Mouton L. et al. **2012**. Complete genome sequence of “*Candidatus* Portiera aleyrodidarum” BT-QVLC, an obligate symbiont that supplies amino acids and carotenoids to *Bemisia tabaci*. *Journal of Bacteriology*, **194**, 6654-6655.
- Schmitz A., Anselme C., Ravallec M., Rebuf C., Simon J.C., Gatti J.L. et al. **2012**. The cellular immune response of the pea aphid to foreign intrusion and symbiotic challenge. *PLoS One*, **7**, e42114.

- Shigenobu S., Watanabe H., Hattori M., Sakaki Y. et Ishikawa H. **2000**. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature*, **407**, 81-86.
- Shigenobu S. et Wilson A.C. **2011**. Genomic revelations of a mutualism: the pea aphid and its obligate bacterial symbiont. *Cellular and Molecular Life Sciences*, **68**, 1297-1309.
- Silva F.J., Latorre A. et Moya A. **2001**. Genome size reduction through multiple events of gene disintegration in *Buchnera* APS. *Trends in Genetics*, **17**, 615-618.
- Silva F.J., Latorre A. et Moya A. **2003**. Why are the genomes of endosymbiotic bacteria so stable? *Trends in Genetics*, **19**, 176-180.
- Skaljic M., Zanic K., Ban S.G., Kontsedalov S. et Ghanim M. **2010**. Co-infection and localization of secondary symbionts in two whitefly species. *BMC Microbiology*, **10**, 142.
- Skurnik M., Peippo A. et Ervelä E. **2000**. Characterization of the O-antigen gene clusters of *Yersinia pseudotuberculosis* and the cryptic O-antigen gene cluster of *Yersinia pestis* shows that the plague bacillus is most closely related to and has evolved from *Y. pseudotuberculosis* serotype O: 1b. *Molecular Microbiology*, **37**, 316-330.
- Sloan D.B. et Moran N.A. **2012a**. Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. *Molecular Biology and Evolution*, **29**, 3781-3792.
- Sloan D.B. et Moran N.A. **2012b**. Endosymbiotic bacteria as a source of carotenoids in whiteflies. *Biology Letters*, **8**, 986-989.
- Stouthamer R., Breeuwer J.A. et Hurst G.D. **1999**. *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. *Annual Reviews in Microbiology*, **53**, 71-102.
- Su Q., Xie W., Wang S., Wu Q., Liu B., Fang Y. et al. **2014**. The endosymbiont *Hamiltonella* increases the growth rate of its host *Bemisia tabaci* during periods of nutritional stress. *PLoS One*, **9**, e89002.
- Swoboda J.G., Campbell J., Meredith T.C. et Walker S. **2010**. Wall teichoic acid function, biosynthesis, and inhibition. *ChemBiochem*, **11**, 35-45.
- Taylor B.L. et Zhulin I.B. **1999**. PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiology and Molecular Biology Reviews*, **63**, 479-506.
- Telesnicki M.C., Ghera C.M., Martínez-Ghera M.A. et Arneodo J.D. **2012**. Identificación molecular del endosimbionte secundario *Hamiltonella defensa* en el pulgón amarillo de los cereales, *Metopolophium dirhodum*. *Revista Argentina de Microbiología*, **44**, 255-258.
- Thomas G.H., Zucker J., Macdonald S.J., Sorokin A., Goryanin I. et Douglas A.E. **2009**. A fragile metabolic network adapted for cooperation in the symbiotic bacterium *Buchnera aphidicola*. *BMC Systems Biology*, **3**, 24.
- Thomson N.R., Howard S., Wren B.W., Holden M.T., Crossman L., Challis G.L. et al. **2006**. The complete genome sequence and comparative genome analysis of the high pathogenicity *Yersinia enterocolitica* strain 8081. *PLoS Genetics*, **2**, e206.
- von Tils D., Blädel I., Schmidt M.A. et Heusipp G. **2012**. Type II secretion in *Yersinia*—a secretion system for pathogenicity and environmental fitness. *Frontiers in Cellular and Infection Microbiology*, **2**.
- Tsuchida T., Koga R., Meng X.Y., Matsumoto T. et Fukatsu T. **2005**. Characterization of a facultative endosymbiotic bacterium of the pea aphid *Acyrtosiphon pisum*. *Microbial Ecology*, **49**, 126-133.
- Tsuchida T., Koga R., Meng X.Y., Matsumoto T. et Fukatsu T. **2005**. Characterization of a facultative endosymbiotic bacterium of the pea aphid *Acyrtosiphon pisum*. *Microbial Ecology*, **49**, 126-133.

Wicken A.J. et Knox K.W. **1975**. Lipoteichoic acids: a new class of bacterial antigen. *Science*, **187**, 1161-1167.

Wu M., Sun L.V, Vamathevan J., Riegler M., Deboy R., Brownlie J.C. *et al.* **2004**. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biology*, **2**, 0327-0341.

Young F.E. **1966**. Fractionation and partial characterization of the products of autolysis of cell walls of *Bacillus subtilis*. *Journal of Bacteriology*, **92**, 839-846.

## BILAN

Nous avons clairement mis en évidence l'existence de deux clades d'*H. defensa*, assortie à une spécialisation de ces bactéries à leur hôte respectif. La divergence entre ces deux clades de bactéries provient principalement de pertes différentielles dans leur matériel génétique ancestral.

Les facteurs « virulents », héritage d'ancêtres pathogènes potentiels, en sont les premières victimes, révélant une réduction globale du coût associé à la présence de la bactérie pour l'hôte. Cette avirulence représente l'évolution attendue du fait de la transmission verticale des symbiotes : les hôtes infectés par les bactéries les moins coûteuses seront sélectionnés à l'échelle de la population insecte. Toutefois, la nature du bénéfice apporté par la bactérie peut nécessiter une certaine forme d'agressivité de sa part. Ainsi, la conservation par le symbiote du puceron de nombreux systèmes particulièrement virulents (systèmes de sécrétion de type III, antigène-O) est en conformité avec son phénotype protecteur. À l'inverse, l'avirulence semble beaucoup plus importante chez les symbiotes de *Bemisia tabaci*. Cela peut être mis en relation directe avec son phénotype, puisque la bactérie ne paraît pas concernée par la protection de l'hôte. Sa séquestration dans les bactériocytes rend en outre dispensable tout pouvoir invasif. L'écologie et le phénotype sont d'ailleurs probablement reliés, puisque cette invasion des tissus de l'hôte permettrait le contact avec les ennemis de celui-ci.

Un second aspect intéressant concerne la capacité des symbiotes à acquérir du matériel génétique. Le bactériophage APSE et le plasmide pHD5AT sont tous deux en piètre état dans le génome des symbiotes de *B. tabaci*. Du fait de leur implication dans l'acquisition de nouveaux gènes, et donc dans la course aux armements engagée contre le parasitoïde, leur absence semble peu pénalisante dans le cadre d'une symbiose nutritive d'*H. defensa* chez l'aleurode. La capacité à acquérir de nouveaux gènes pourrait permettre également l'invasion de nouveaux hôtes chez les endosymbiotes secondaires. Cette capacité pourrait être déjà malmenée chez les symbiotes de *B. tabaci*, du fait on le rappelle de leur inaptitude à synthétiser le chorismate.

En conclusion, il est difficile de faire un lien précis entre évolution génomique d'une part, écologie et/ou phénotype de l'autre part. La génomique comparative broserait toutefois le portrait de deux symbiotes distincts, l'un virulent et protecteur, associé à deux partenaires impliqués dans l'acquisition de nouveaux gènes, et l'autre avirulent et peu enclin à l'évolvabilité.

Existe-t-il d'autres endosymbiotes chez *B. tabaci* qui pourraient s'avérer être des protecteurs ? Dans le cadre de notre collaboration, nous avons analysé le génome de *Cardinium hertigii* cBtQ1, symbiote de l'espèce MED-Q1 (**Annexe 2**). De manière intéressante, ce symbiote pourrait produire des protéines insecticides, et son génome code pour un système potentiellement impliqué dans la mobilité qui mérite que l'on s'y intéresse, d'autant plus que ces gènes sont présents en deux copies dans le génome. Nous aurions donc affaire à un endosymbiote mobile et virulent, comme *H. defensa* chez le puceron. De là à dire qu'il assume le même rôle chez l'aleurode...



# ARTICLE TROISIEME

---

## Différentiel de taux d'évolution chez les endosymbiotes : Influence du contenu génique et de l'écologie sur l'évolution génomique

---

### Contexte

Nous avons étudié l'évolution du genre *Hamiltonella*. Pour ce faire, nous avons utilisé différents génomes d'organismes références. Parmi ces génomes, ceux de *Regiella insecticola* ont retenu particulièrement notre attention. En effet, *Hamiltonella* et *Regiella* sont deux genres frères, qui forment un groupe monophylétique apparenté aux pathogènes *Yersinia* (Degnan *et al.*, 2009 ; Husnik *et al.*, 2011). Tous deux renferment de nombreuses souches protectrices des pucerons, que ce soit contre les parasitoïdes ou les champignons pathogènes (Oliver *et al.*, 2013). De la sorte, leur ancêtre supposé serait un endosymbiote facultatif qui aurait déjà fait de la protection sa stratégie de maintien dans les populations d'insectes (Baumann, 2005).

Lors de notre précédente analyse, nous avons toutefois noté que la divergence (au niveau des séquences protéiques) entre *Yersinia* et *Regiella* semblait moins prononcée qu'entre *Yersinia* et *Hamiltonella*. Quelles peuvent en être les raisons ? Quelles autres caractéristiques distinguent les génomes des deux clades de symbiotes ?

### Attente et objectifs

Il nous est venu à l'idée que l'ancêtre d'*H. defensa* et *R. insecticola* pourrait ne pas être un endosymbiote, mais une bactérie libre depuis laquelle auraient émergé deux clades de bactéries intracellulaires. Si *H. defensa* était un clade plus ancien d'endosymbiotes, cela expliquerait le patron observé. Les taux de mutations pourraient être différents entre les deux lignées. Par exemple, la perte de gènes liés à la réparation et la recombinaison conduit à l'apparition d'organismes à l'évolution accélérée. La probabilité de fixer ces mutations peut également différer, par exemple si les tailles efficaces des populations diffèrent entre les symbiotes.



Nous avons voulu tester ces différentes hypothèses au moyen de la génomique comparative. Tout d'abord, les endosymbiotes ont une évolution caractéristique (pertes de gènes, réarrangements synténiques, accumulation de mutations etc.) qui laisse une empreinte forte dans leurs génomes (Moran *et al.*, 2008). Si le dernier ancêtre des deux espèces était un endosymbiote, son génome aurait subi de tels événements, dont les stigmates devraient être partagés par ses descendants. Ensuite, les gènes perdus dans les génomes de ces derniers, et notamment ceux liés à la réparation et à l'écologie, pourraient avoir eu un impact, respectivement, sur les taux de mutations subis, ainsi que sur la probabilité que ces mutations même délétères se soient fixées.

---

Differential in evolutionary rates in *Hamiltonella defensa* and *Regiella insecticola* since the last ancestor:

Influence of gene content and ecology on genomic evolution

---

Pierre-Antoine Rollat-Farnier, Marie-France Sagot, Fabrice Vavre et Laurence Mouton

Cet article sera soumis à *Genome Biology and Evolution*

## I. Introduction

Insects are privileged partners of intracellular residents, called endosymbionts (Baumann, 2005 ; Moran *et al.*, 2008). It is estimated that more than 50% of the insect species are infected by such partners, generally bacterial (Van Dijk, 2008). Some hot spots exist among the insects. For example, the species that have an unbalanced diet (blood, sap etc.) generally depend on the presence of primary endosymbionts, which are by definition obligate for their eukaryotic host (Buchner, 1965 ; Nogge, 1982). These symbionts supply the host with essential nutrients that are not found in their diet, and are transmitted maternally from mother to progeny (Baumann, 2005 ; Moran *et al.*, 2008). As a result, a co-cladogenesis between the two partners is often observed lasting hundreds of millions of years, as revealed by the perfect congruence of their respective phylogenies (Baumann, 2005 ; Moran *et al.*, 2008). Insects are also infected by secondary endosymbionts, which use both mutualistic and parasitic strategies to maintain themselves in the host populations, and that are generally facultative for their hosts (Baumann, 2005 ; Moran *et al.*, 2008). Almost no congruence is detectable between the phylogenies of secondary endosymbionts and of their hosts, indicating that co-cladogenesis between these species is limited, and that horizontal transmission between species is frequent (Baumann, 2005 ; Moran *et al.*, 2008).

The intracellular lifestyle shared by primary and secondary symbionts has drastic and irreversible genomic consequences. Independent lineages of endosymbionts all exhibit AT enriched and tiny genomes, with a reduced gene content, high substitution rates and a gene organization totally remodeled compared to their free-living relatives (Anderson and Kurland, 1998; Moran *et al.*, 2008 ; Shigenobu *et al.*, 2000 ; Nakabachi *et al.*, 2006 ; Santos-

Garcia *et al.*, 2012 ; Sloan and Moran, 2012). Two main reasons explain these convergences. First, the cell is rich in nutrients and confers stable conditions of life to these organisms (Shigenobu *et al.*, 2000). As a consequence, endosymbionts lose an important part of the genes ancestrally involved in motility, regulation, environmental perception etc. (Moran *et al.*, 2005a ; Moran *et al.*, 2008). A second explanation is that maternal transmission imposes frequent bottlenecks to the endosymbionts, leading to intra-host populations with reduced effective sizes (Moran, 1996). This alters the efficiency of purifying selection, and favors the action of drift (Moran, 1996). The endosymbionts thus accumulate mildly deleterious mutations in their genes, which generate pseudogenes and participate to the erosion of their genomes.

This global evolution of endosymbionts presents important variations among the endosymbiotic lineages. The endosymbiotic characteristics are more pronounced in primary endosymbionts, which have much smaller genomes compared to secondary endosymbionts (Moran *et al.*, 2008). The reason is mostly ecological. Indeed, while primary endosymbionts are sequestered in only one type of “symbiotic” cells called bacteriocytes, secondary endosymbionts invade diverse organs and fluids in the insect body (muscles, hemolymph, saliva, bacteriocytes etc.). Moreover, because they are facultative and can be lost by their host, their long-time survival mainly depends on their ability to spread horizontally, *i.e.* from host to host, of a same species or not (Oliver *et al.*, 2010). This invasion of tissues and hosts requires hundreds of genes which are dispensable in obligate endosymbionts, and thus can be lost in their genomes. The acquisition of new functions through Lateral Gene Transfer (LGT) can facilitate these horizontal transfers of bacteria, and are based on genes involved in DNA repair and recombination, and on phages and plasmids, which allow the integration of new genetic material and are generally absent in primary symbionts (Moran *et al.*, 2005b).

Importantly, the rate of change is not linear in time. Ecological transitions are probably associated with the more impressive accelerations in endosymbiont evolution. Two major transitions exist: from a free-living bacterium towards an endosymbiont (obligate or facultative), and from a facultative to an obligate endosymbiont. The first is followed by an intense process of pseudogenization and deletion of genes, mainly because the stable intracellular environment leads to pressure relaxation (Toh *et al.*, 2006 ; Oakeson *et al.*,

2014). This erosion can be accompanied by gene order reorganization, correlated with the spread of mobile elements (Oakeson *et al.*, 2014). These phenomena can be explained by the ineffective purifying selection and by the presence of pseudogenes that can be used as a template (Oakeson *et al.*, 2014). The transition from facultative to obligate endosymbiosis is correlated with important pseudogenization rates and genome erosion, because the less sophisticated ecology of primary endosymbionts induces relaxed selective pressure on hundreds of genes that can be lost without lethal consequences (Lamelas *et al.*, 2011). After both transitions, the evolutionary processes are assumed to progressively slow down. Consistently, the substitution rate in primary endosymbionts is inversely correlated with the elapsed time since endosymbiosis (Allen *et al.*, 2009). Moreover, some “old” genomes of obligate endosymbionts underwent no break of their synteny in dozens of millions of years (Tamas *et al.*, 2002).

In addition to ecology and the age of symbiosis, some specific genes or genomic properties influence the rates of these evolutionary processes. The presence of repeated elements can favor rearrangements, mainly in secondary endosymbionts, but also in rare cases, in primary endosymbionts (Sloan and Moran, 2013). In primary endosymbionts, the loss of the genes involved in DNA repair leads to an increase in the substitution rates (Sloan and Moran, 2013). Finally, the ecology of the host may also have an influence as mutations in endosymbionts generate inter-host variability. If the action of selection is strong in the insect population, the more beneficial endosymbiotic variants can regularly be fixed in the population, in a process that is called “selective sweep” (Shoemaker *et al.*, 2004; Hurst and Jiggins, 2005). This sweeps will reduce the effective size of the cytoplasmic compartments, and can lead to frequent fixation of mutations present in the genome of the selected variants.

*Hamiltonella defensa* and *Regiella insecticola* form a monophyletic clade of facultative endosymbionts that descend from pathogenic bacteria close to the genus *Yersinia* (Degnan *et al.*, 2009). Both protect aphids against their enemies (parasitoids, fungi) (Oliver *et al.*, 2003 ; Scarborough *et al.*, 2005), and the more parsimonious hypothesis assumes that their last ancestor was already a protective endosymbiont (Baumann, 2005). *H. defensa* is also found in the aleurod *Bemisia tabaci* (species MED and MEAM1), in which it

could have evolved to an obligate symbiont that furnishes lysine to its partners (Rao *et al.*, submitted ; Rollat-Farnier *et al.*, in revision). Interestingly, initial analyses revealed that *H. defensa* and *R. insecticola* present differences in terms of substitution rates, *H. defensa* having more diverged from *Yersinia* than *R. insecticola*. Different hypotheses that are not mutually exclusive could explain why these two lineages present such differences. A first could be that their common ancestor was in reality a free-living bacterium, and that two endosymbioses emerged in an independent fashion. If on the contrary the transition towards endosymbiosis occurred well before the *Regiella-Hamiltonella* divergence, the observed differences could be explained by gene content and/or wider ecology. For example, independent loss of genes involved in DNA repair could have increased the mutation rate in one lineage. Another hypothesis would be that the probability to fix mutations is higher in one lineage because of a strong action of drift in populations with a reduced effective size.

Comparative genomics is a suitable approach to discriminate between these different hypotheses. First, if the last ancestor was an endosymbiont, its genome probably underwent the above-described events: gene losses, accumulation of mutations, and synteny perturbation. In this case, its descendants would share common mutations, which could be detectable by such an approach. Second, comparative genomics allows determining if the increase in substitution rate in one lineage is correlated with other phenomena like the increase of deleterious mutations, which would inform on the pressures exerted on the two lineages. The status of genes involved in DNA repair in the two lineages (presence/absence, pseudogenization) will also be of great interest to us. In conclusion, genomic comparison will furnish a better understanding of the ecology and mutation rates in the two lineages, which could help to understand the forces that constrain them, and explain the observed pattern of substitution rates.

## II. Material and methods

### 1. Choice of an outgroup

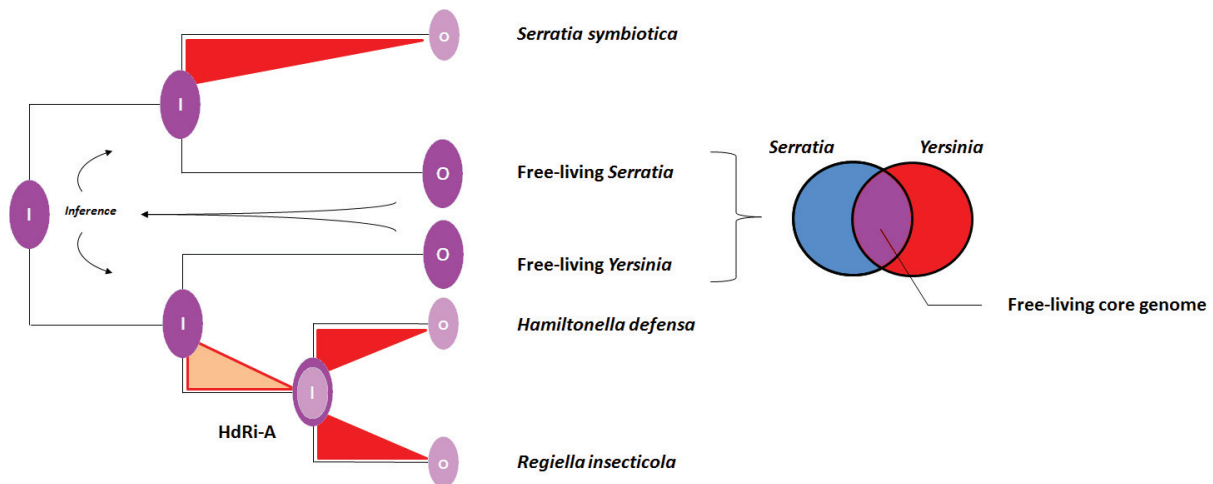
The use of an endosymbiotic outgroup is primordial to determine if the ancestor of *H. defensa* and *R. insecticola* was an endosymbiont. Indeed, convergences being numerous in endosymbionts, surely *H. defensa* and *R. insecticola* will share common characteristics - for example gene losses - whether their ancestor was an endosymbiont or a free-living bacterium. The question is thus to determine if the resemblance between the two bacteria is so important that it excludes the hypothesis of independent transitions towards endosymbiosis.

The species *Serratia symbiotica* is the endosymbiotic clade that is closest to *H. defensa* and *R. insecticola* (**Figure 1**). This is the ideal outgroup for many reasons. First, since they are close relatives, the free-living ancestors of the three endosymbionts shared an important set of common genes. Most of these genes should still be present in the free-living relatives of the two lineages, the genera *Yersinia* and *Serratia*, and the available genomes of these free-living bacteria will allow us to infer the ancestral core genome of the endosymbionts (**Figure 1**).

Second, among the different strains of *S. symbiotica*, some have the same phenotype as *R. insecticola* and *H. defensa*: they are secondary endosymbionts with a protective role in aphids (Oliver *et al.*, 2003). This characteristic will be of great interest. Indeed, if we had used an outgroup with another phenotype, its genome would have undergone a different evolutionary pressure compared to the protective endosymbionts *H. defensa* and *R. insecticola*, which would have increased the risk of clustering them on the basis of the phenotype even if their ancestor was not an endosymbiont. By using a protective outgroup, we minimized this risk, and could therefore be more confident about our conclusions.

The genome of one of these protective *S. symbiotica*, the strain Tucson of the aphid *Acyrtosiphon pisum*, has been assembled and can be used as outgroup (Burke and Moran 2011). The genomes of two other strains have been sequenced but will not interest us

because they followed evolutionary trajectories very distinct from *H. defensa* and *R. insecticola*. These are the primary endosymbiont of *Cinara cedri* (Lamelas *et al.*, 2011), and a symbiont of *Aphis fabae*, which presents the peculiarity of being both symbiotic and free-living (Foray *et al.*, 2014).



**Figure 1: Relationships between *S. symbiotica* and the *H. defensa*-*R. insecticola* clade.** Inferred gene contents are indicated by the letter “I”, observed gene contents by the letter “O”. Free-living *Serratia* and *Yersinia* genomes can be used to determine a Free-Living Core Genome (FLCG, dark purple circles). The FLCG is assumed to have been part of the genome of the ancestor of all represented organisms, but also in the genome of the ancestor of the genus *Serratia* (free-living and endosymbiotic strains), and in the genome of the ancestor of the clade *Yersinia*-*H. defensa*-*R. insecticola*. In endosymbiotic genomes, the FLCG will be extremely reduced (light purple circles). Concerning the ancestor of *H. defensa* and *R. insecticola* (HdRi-A), nothing is known about the fate of the FLCG: if HdRi-A was an endosymbiont, the FLCG in its genome was already drastically reduced; if not, it was probably mostly preserved. The fate of the FLCG in the different endosymbiotic lineages will interest us to understand the evolution of *H. defensa* and *R. insecticola* since their last common ancestor.

## 2. Genome acquisition

The genomes of twelve organisms were used in this study, six free-living bacteria that will be used to infer the ancestral gene content of the clade (**Figure 1**), and six endosymbionts. Two draft genomes were retrieved from the NCBI databases: *R. insecticola* strains LSR1 (Degnan *et al.*, 2010) and 5.15 (Hansen *et al.*, 2012), both symbionts of aphids. The other ten genomes were retrieved from the MaGe platform (Vallenet *et al.*, 2006). Three *Hamiltonella defensa* strains were used: 5AT, symbiont of the aphid *Acyrtosiphon pisum* and whose genome is complete (Degnan *et al.*, 2009; referred to as hApis\_5AT,), MEAM1 and MED-Q1 symbionts of the aleurod *Bemisia tabaci*, whose genomes are draft assemblies

(Rollat-Farnier *et al.*, in revision; Rao *et al.*, submitted; referred to as hBtab\_MEAM1 and hBtab\_MED-Q1). The genome of the outgroup *S. symbiotica* strain Tucson used in this study is also a draft assembly (Burke and Moran, 2005).

The six free-living organisms all presented a complete genome. We used three strains of free-living *Serratia* (*S. proteamaculans* 568 ; *S. marcescens* WW4, Chung *et al.*, 2013 ; *S. plymuthica* AS9, Neupane *et al.*, 2013) and three strains of free-living *Yersinia* (*Y. enterocolitica* 8081, Thomson *et al.*, 2006 ; *Y. pestis* CO92, Parkhill *et al.*, 2001 ; *Y. pseudotuberculosis* IP32953, Chain *et al.*, 2004).

### 3. Ortholog relationship and ancestor reconstruction

To analyze the evolution of the ancestral gene content in endosymbionts, the first step was to find orthologs shared by all free-living bacteria, and to study their fate in the endosymbiotic lineages. The proteins of all twelve genomes were thus extracted, and grouped into families using **SiliX** with default parameters (Miele *et al.*, 2011). The families that contained more than one protein for at least one organism were manually checked and curated. In the case of the draft genomes of the endosymbionts, the genes at the end of the contigs that had been erroneously discarded were manually added back. The default parameters of **SiliX** were not stringent enough for the huge families of ATP-binding components of ABC-transporters: using such parameters, the 738 sequences were all grouped inside a same **SiliX** family. Setting a threshold of 45% for the required proteic similarity enabled however to separate these 738 proteins into different families using **SiliX**. This allowed splitting the ABC-transporters into 88 **SiliX** families, 65 of which contained one sequence per genome. The 23 remaining families were manually checked. In this study, we focused on the evolution of the functional gene content, and were not interested by the families of mobile elements. The latter were thus excluded from the analysis. This concerned toxin/antitoxin systems, plasmid genes, transposases, reverse transcriptases and phage genes. Indeed, these genes are the subject of frequent LGT between bacteria, and could disturb the analysis.



For each protein in each family, COG categories were assigned with **COGnitor** (implemented in the MaGe platform). We then checked that all genes in a given family were homogenous in terms of COG assignments. Some families contained contradictions in such assignment. This was generally due to assignments that were close to the COGnitor threshold. They were manually curated on the basis of information retrieved from **Uniprot** (Uniprot Consortium, 2008). However, 99% of the families were homogenous, which indicates that the **Silix** construction was reliable.

#### **4. Free-living and endosymbiotic core genomes**

The reconstruction gave 2262 families that present one copy for all the six genomes of free-living *Yersinia* and *Serratia*. The 2262 corresponding proteins from the Free-Living Core Genome (FLCG) were considered as ancestrally present in all endosymbiotic lineages before their transition to an intracellular lifestyle (**Figure 1**).

Only 499 families presented one copy for all the twelve organisms. This corresponds to the Total Core Genome (TCG) of all studied bacteria, whose reduced size mostly reflects the different gene losses and inactivations that occurred in symbiotic lineages and strains. Such core genome will be adopted in the following analyses because it allows to use strictly the same genes for all endosymbionts.

#### **5. Determination of the substitution rates**

A phylogenetic tree was built to determine the substitution rates in endosymbiotic lineages. To this purpose, **ClustalW** (Thomopson *et al.*, 2004) was used to align proteins inside each of the 499 families of the TCG. The 499 alignments were then concatenated, leading to one sequence for each organism. The best model for tree reconstruction was determined with the **CLC Workbench Software 7.0.3** (<http://www.clcbio.com>), and a phylogeny was performed with **phymI** (Guindon *et al.*, 2009) using this model. A second alignment was obtained from this concatenation, using a home-made perl script to keep only the codons containing G and C for all concerned genomes. The phylogeny was built following the same procedure, with **CLC** and **phymI**.

## 6. Gene properties

We used **BlastP** inside each of the 499 gene families to determine the proteic similarity among all the organisms. The **seqinr** package (Charif *et al.*, 2007) implemented in the R software was used to determine for each organism the AT composition of the 499 genes corresponding to the proteins.

For each family, the sequences were aligned with **clustalW** at the proteic level, and then at the nucleic level constrained by the protein alignment with a home-made **perl** script. The software **seqinr** was used to determine the dN/dS ratio between *Y. enterocolitica* and the strains of *H. defensa* and *R. insecticola* on the basis of the nucleic alignments.

## 7. Ancestor reconstruction

We also wanted to study the fate of the genes in endosymbiotic lineages. To that purpose, it was necessary to determine the gene content of the ancestor of the genus *Hamiltonella* and the one of the ancestor of the genus *Regiella*. In order to do this, a parsimony principle was applied: a gene from the 2262 genes of the FLCG was assumed to be present in the last ancestor if it was observed in at least one of its descendants. We applied this principle to the ancestors of *H. defensa* (referred to as “Hd-A”) and of *R. insecticola* (referred to as “Ri-A”), and to the ancestor of the two species (referred to as “HdRi-A”). This allowed to compare the repartition of gene losses among the three lineages (*H. defensa*, *R. insecticola* and *S. symbiotica*) in order to solve the question of the ecology of the ancestor HdRi-A.

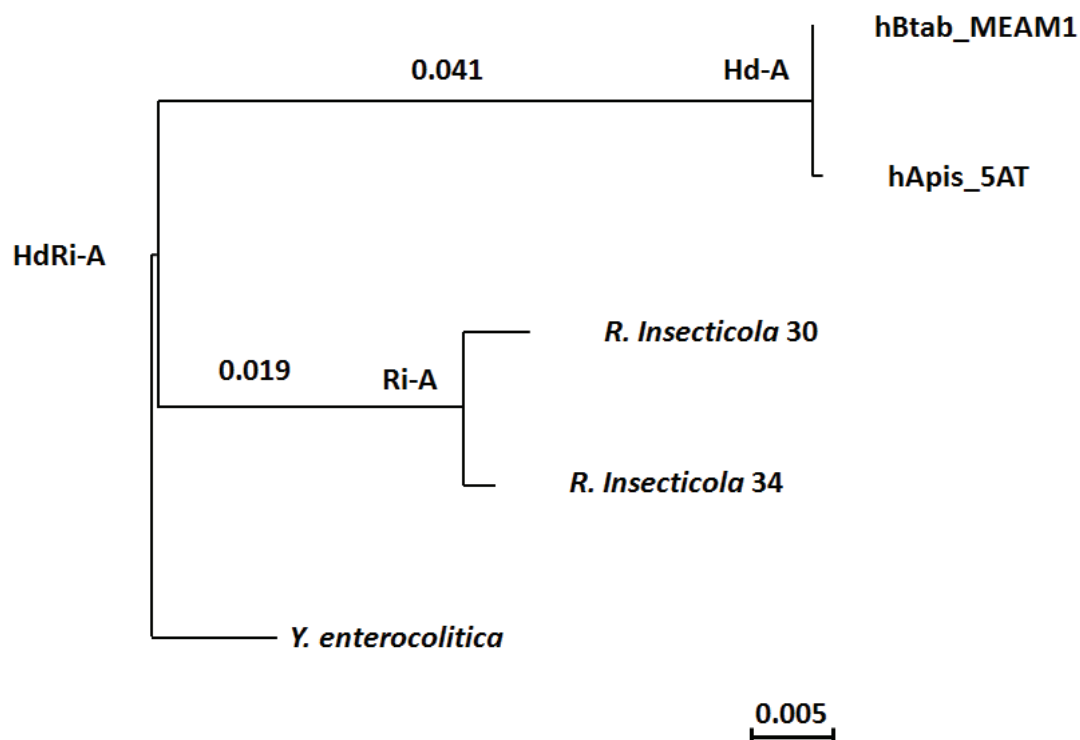
## 8. GO term enrichment

To compare HdA and RiA and better understand their respective ecology, a GO term enrichment analysis was performed using the **Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7** (Dennis *et al.*, 2003). The HdRiA gene content was used as a background, and we applied the **DAVID Functional Annotation Clustering** with medium stringency on the genes lost by HdA on one hand, and the genes lost by RiA on the other.

### III. Results

#### 1. Phylogeny reconstruction and substitution rates

The concatenation of the proteins of the TCG allowed reconstructing the phylogeny shown in **Figure 2**.



**Figure 2: Substitution rates in the *Hamiltonella-Regiella* clade.** Phylogeny based on the concatenation of the 499 proteic sequences of the TCG. Bootstrap values are indicated on the corresponding nodes.

As can be observed in this figure, the distance based on the number of substitutions between *Y. enterocolitica* and HdRi-A is less important than the distance from this ancestral bacterium to any of its current descendants. This is not surprising given the endosymbiotic lifestyle of the latter, which implies important rates of substitutions. What is more interesting is the fact that the distances between the last common ancestor to the sequenced strains of *R. insecticola* are about 2 times smaller than between this same ancestor and the *H. defensa* strains (**Figure 2**). Therefore, the evolutionary rates in the endosymbiotic lineages are higher than in the free-living lineage, but not constant, the *H. defensa* species presenting a faster rate than *R. insecticola*.

	<i>H. defensa</i>		<i>R. insecticola</i>	
	<u>hApis_5AT</u>	<u>hBtab_MEAM1</u>	<u>rApis_LSR1</u>	<u>rMper_5.15</u>
<i>Y. enterocolitica</i>	69.6	69.3	75	74.7
<u>hApis_5AT</u>		97.3	66	65.7
<u>hBtab_MEAM1</u>			65.9	65.5
<u>rApis_LSR1</u>				95.7

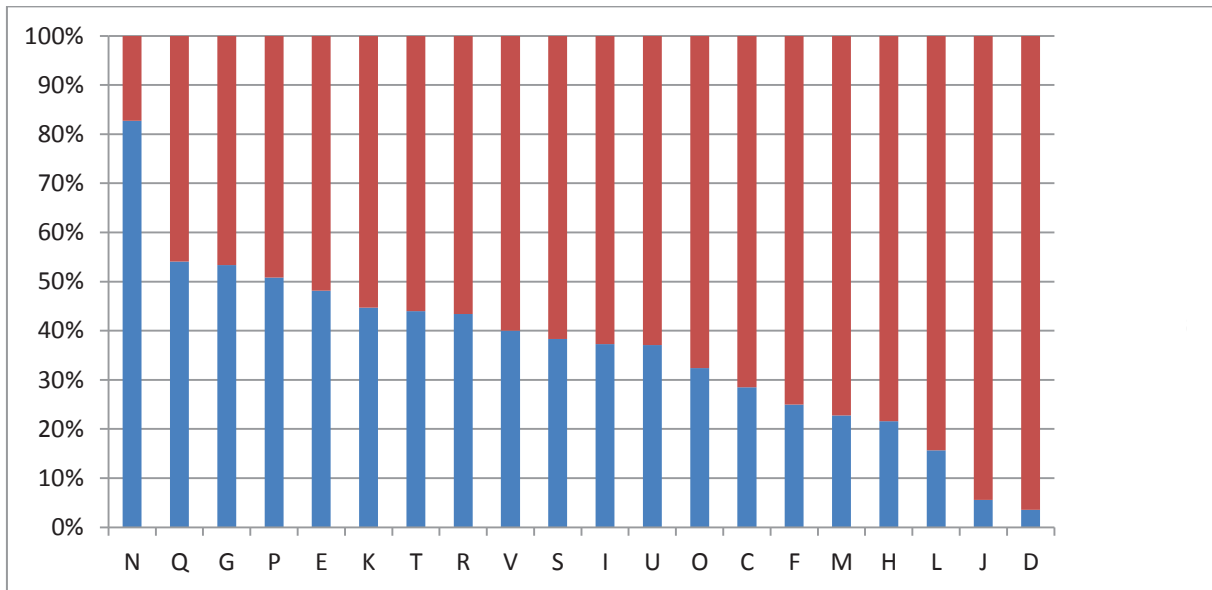
**Table 1: Proteic similarity.** This table gives the similarity at the proteic scale between each couple of genomes.

Consistently, *Y. enterocolitica* presents a proteic similarity of about 69% with *H. defensa*, and 75% with *R. insecticola* (**Table 1**). It is significantly closer to the latter (Pearson's test, p-value < 2e-16). As for the substitution distance, the genera *H. defensa* and *R. insecticola* exhibit less similarity between them than with *Y. enterocolitica* (Pearson's test, p-value < 3e-08).

## 2. Convergences in gene losses

To explain this pattern, a first hypothesis could be that the transition towards endosymbiosis occurred earlier in the *H. defensa* lineage. In this case, the ancestor of this lineage and of *R. insecticola* would not have been an endosymbiont, and three transitions would have taken place independently in the species *R. insecticola*, *S. symbiotica* and *H. defensa*.

We decided to investigate the genes lost in the three endosymbiotic lineages to test this hypothesis. We compared the fate of the FLCG in the three genomes: Ri-A (the ancestor of the *Regiella* genus), Hd-A (the ancestor of the *Hamiltonella* genus) and *S. symbiotica* strain Tucson. From the 2262 genes composing the FLCG, 822 are lost in all endosymbiotic lineages (36%). This number cannot be explained by random losses ( $\chi^2$ , p.value < 0.001), which is consistent with several previous studies that indicated that some gene categories are preferentially lost in endosymbionts (Delmotte *et al.*, 2006 ;Manzano-Marín *et al.*, 2012; **Figure 3**).



**Figure 3: Convergent losses in the different COG categories.** For each category, the percentage of losses is indicated in blue, retained genes are in red. The categories are sorted from the more affected to the less.

For example, the genes involved in cell cycle (D COG category), translation (J category) and recombination/repair (L category) are affected by less than 20% of losses, while those involved in the metabolism/transport of inorganic ions (P category) and carbohydrates (G category), as well as secondary structures (Q category) are affected by more than 50% of losses, the genes involved in motility (N motility) being affected by more than 80% of losses.

### 3. Repartition of the losses

With this goal in mind, we removed the genes lost in all three lineages, and only considered the genes conserved in at least one of them (**Table 2**). Here again, the repartition of gene losses cannot be explained by chance alone ( $\chi^2$ , p.value < 0.001). It appears that HdA and RiA are clearly separated from *S. symbiotica* (**Table 2**). Indeed, the major contribution to the  $\chi^2$  value came from genes lost by both HdA and RiA, which are 5-times more numerous than expected, and genes lost only by *S. symbiotica*, which are 1.5-times more numerous. In contrast, there are fewer genes than expected which are shared by *S. symbiotica* and one of the other two lineages.

Genome(s)	Observed losses	Expected losses	$\chi^2$ contribution
HdA-RiA	103	21	<b>323,0</b>
HdA- <i>S. symbiotica</i>	44	73	11,4
RiA- <i>S. symbiotica</i>	122	158	8,1
HdA	28	29	0,0
RiA	69	63	0,6
<i>S. symbiotica</i>	336	219	<b>62,5</b>

**Table 2: Repartition of gene losses between the three lineages.** For each organism, or couple of organisms, are indicated the observed and expected losses, and the contribution for the value of the  $\chi^2$ .

The most likely hypothesis to explain such a pattern is that most of the losses common to RiA and HdA have occurred before their last ancestor. This clearly indicates that two transitions towards endosymbiosis occurred: one which led to *S. symbiotica*, and one which led to the ancestor of RiA and HdA. This allowed us to reconstruct the ancestor of RiA and HdA, that we called HdRi-A. This ancestor retained about 1100 genes from the FLCG, which brings it closer to the 1200 genes retained by *S. symbiotica* even if this is still significantly less than expected by chance (Pearson's test, p. value < 0.01). This also implies that the substitution rates in the *H. defensa* and *R. insecticola* species cannot be attributed to independent transitions towards endosymbiosis, and need other explanations.

#### 4. DNA repair genes

A first hypothesis is an increase of the mutation rate, which is generally due to the inactivation of genes which are involved in DNA repair. Different such important genes have been lost in the *H. defensa* lineage (**Table 3**). The 6-O-methylguanine-DNA methyltransferases encoded by *ogt* is involved in the protection of DNA against O-6-methylguanine and O-4-methylthymine. These alkylated forms of guanine and thymine are responsible for the G:C to A:T transversions (Mackay *et al.*, 1994). In the same way, *mutM* was lost in the ancestor of *H. defensa*. This gene encodes a formamidopyrimidine-DNA glycosylase whose main function is to recognize and remove oxidized guanines such as 7,8-dihydro-8-oxoguanine (8-oxo-G). In *Escherichia coli*, a knocking out of both *mutM* and *mutY* generates a 1000-fold increase of G:C to A:T transversions (Horst *et al.*, 1999). The loss of these two genes in other reduced genomes has already been proposed as being responsible for AT enrichment (Dufresne *et al.*, 2005). *H. defensa* also lost *recQ*, which has been

correlated with accelerated mutation rates (Kato and Inoue, 2006).

Genes	Hd-A	Ri-A
<i>ogt</i>		Present
<i>mutM</i>		Present
<i>recQ</i>		Present
<i>ada</i>	Present	
<i>recF</i>	Present	
<i>recO</i>	Present	
<i>recC</i>	Present	
<i>phr</i>	Present	

**Table 3: DNA repair genes lost in the *H. defensa* or in the *R. insecticola* genuses.**

In *R. insecticola*, difference DNA repair and recombination genes have also been lost. It lost the *ada* gene, which perform the same function as *ogt*. Nevertheless, this gene in *H. defensa* lost its DNA binding and regulator sites, which suggests a bad functioning in this species. Ri-A also lost the RecFOR pathway, involved in recombination, and *phr*, which encodes a photolyase involved in UV protection. Nevertheless, the loss of such photolysase would not cause an increase in mutation rates in prokaryotes (Lucas-Lledó and Lynch, 2009)

Thus, it seems that the losses that occurred in *H. defensa* may have a greater repercussion on mutation rates. It is however very difficult to conclude with certitude. What is sure is that a few lost genes can have a huge impact on this rate (Sloan and Moran, 2014), and that the different losses in the two lineages will have drastic consequences.

## 5. AT enrichment

Consistently with a putative increase of this rate in *H. defensa*, especially concerning GC to AT mutations, the different assemblies reveal that the genome of *R. insecticola* has an AT enrichment slightly less pronounced than is the case for *H. defensa* (57.5% against 60%, **Table 4**). If we look at the coding sequences using the 499 genes of the TCG, both symbiotic genera have an enrichment significantly higher than observed for the reference genome of *Y. enterocolitica* (**Table 4**; Wilcoxon test, p-value < e-16). Moreover, the enrichment that is observed is stronger in *H. defensa* than in *R. insecticola* (Wilcoxon test, p-value < e-16). The bias is more important at the third-codon position (**Table 4**), which is expected due to the

less efficient action of the purifying selection at this position, especially in *H. defensa* (Wilcoxon test, p-value < e-16).

	<i>Y. enterocolitica</i>	<i>H. defensa</i>		<i>R. insecticola</i>	
		<i>hApis_5AT</i>	<i>hBlab_MEAM1</i>	<i>rApis_LSR1</i>	<i>rMper_515</i>
<u>Genome</u>	52.4	59.9	59.5	57.5	57.4
<u>Coding sequences</u>	51.3	59.5	59.6	57.2	57.2
<u>16S rRNA</u>	45.9	46.2	45.8	45.7	46.9 (partial)
<u>Third codons</u>	52.1	65.3	65.4	62.6	62.8
<u>Intergenic space</u>	62	66	66.2	60.3	61

**Table 4: AT enrichment.** On each line, a same grey color clusters genomes which present no significant differences in AT %.

Nevertheless, the substitution rates remain higher in the *H. defensa* lineage even when all codons containing AT bases in the different genomes are removed. This means that there is a global acceleration of the substitution rate in this species, which in turn leads to the AT enrichment. This thus does not exclude that mutation rates are higher in *H. defensa*.

## 6. Rate of gene losses

The other possibility to explain the acceleration observed in *Hamiltonella* is a different probability to fix mutations (*i.e.* the substitution rate) in these two lineages. For example, if the ecology of the two lineages differed, this could have affected the effective size of the bacterial population, and thus the efficiency of purifying selection. Thus it is primordial to have a look on the losses specific to each lineage.

A first observation is that HdA lost significantly more genes than RiA (Pearson's test, p. value < 0.01). Nevertheless, it is difficult to draw a conclusion because these two ancestral bacteria have probably not the same age. Indeed, at the proteic level, the *R. insecticola* bacteria are less similar among them than are the *H. defensa* bacteria (**Table 1**), which implies a more recent divergence inside the latter species. Consistently, if we look at the phylogeny (**Figure 2**), it appears that the substitution distance between HdA and its descendants is shorter than the distance between RiA and its own descendants. The smaller number of losses in RiA could therefore reflect only the shorter time elapsed since the last common ancestor of *H. defensa* and *R. insecticola*.



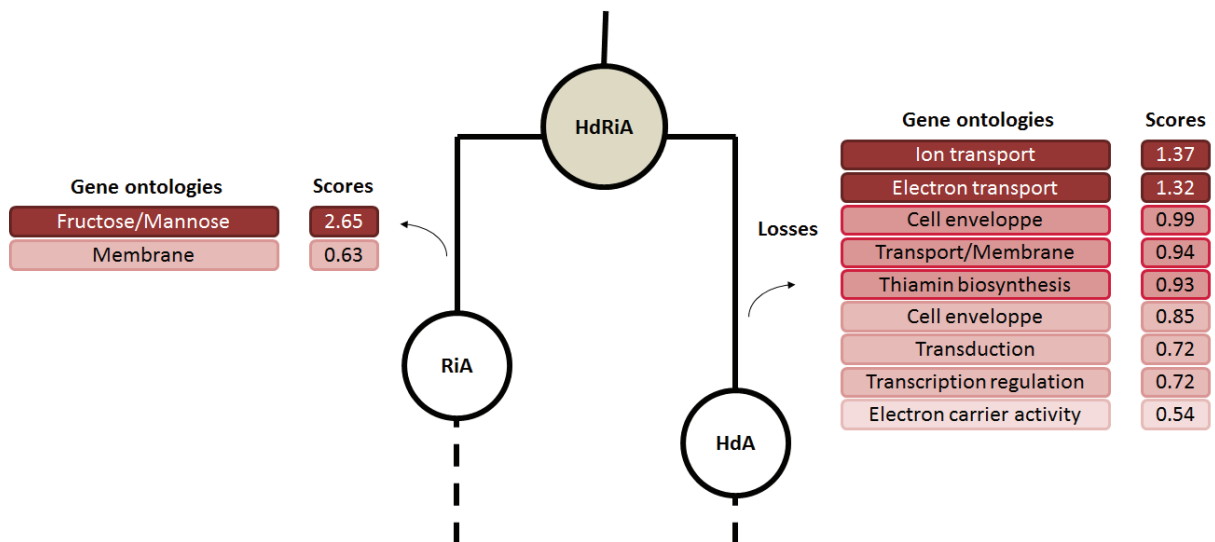
To compare the rates, the length of the branches of the phylogenetic tree represented in **Figure 2** can be useful. If we make the assumption that the rates have been constant along the evolution of each species, we can determine the relative time between HdRi-A and HdA on one hand, and between HdA and its descendants on the other. Based on **Figure 2**, we determined that the time separating HdRi-A from HdA represents 97% of the elapsed time since the HdRi-A ancestor, and the time separating HdA from its descendants represents 3%. Using the same approach for *R. insecticola*, we found that the time separating HdRi-A from RiA represents 93% of the total elapsed time.

Because the elapsed time  $t$  since HdRi-A is the same for all the sequenced strains of *R. insecticola* and *H. defensa*, we can compare the rates of loss in the two lineages: in RiA, we have a rate of 72 losses on  $0.93t$ , and in HdA, a rate of 190 losses on  $0.97t$ . In conclusion, the rate of losses was more than 2-times higher in the branch leading to HdA, compared to the branch leading to RiA. This results highlights that the rate of gene losses was also much more rapid in the *H. defensa* lineage than in the *R. insecticola* lineage. These losses can directly result from an increased rate of mutations in *H. defensa*, which in turn increases the rate of gene inactivation.

## 7. Ecology of the endosymbionts

Another explanation to these losses could be that many genes became dispensable in *H. defensa*, because of a simplification of its niche since HdRi-A. Ecology has an impact on the effective size of the populations and thus on the action of drift, and can explain differences in terms of substitution rate. The differential in gene content between Hd-A and Ri-A could provide some insights into their respective ecology. Interestingly, the losses which are specific to Hd-A are enriched in genes involved in energy production and conversion (corresponding to the C class of COG). This represents approximately 7-times more losses than in the RiA ancestor. The GO term enrichment in losses also abounds in this sense, as three clusters among nine concern genes involved in ion and electron transport (**Figure 4**).

Some of these losses are directly linked to the endosymbiotic ecology. This lineage lost the NADH dehydrogenase II (*ndh*) and different genes involved in the biosynthesis of the cytochrome *bd* quinol oxidase (*cydA*, *B*, *C* and *D*).



**Figure 4: GO term enrichment of the gene losses.** This figure represents the losses that occurred between HdRiA and the ancestors of the genera *Regiella* and *Hamiltonella*. Only gene clusters significantly impacted have been represented.

These two complexes, *ndh* and *cyd*, are directly correlated with the tissue tropism of the bacteria. Because these complexes meet optimal conditions at low oxygen level, both are responsible for the symbiosis between Candidatus *Schmidhempelia bombi* and the bumble bee, allowing the bacterium to survive in the host hindgut and its low-oxygen and anaerobic environments (Martinson *et al.*, 2014). Hd-A also lost the *rnf* and *nqr* operons, which encode two genetically related systems: a Na(+)-translocating ferredoxin:NAD1 oxidoreductase and a Na(+)-translocating NADH:ubiquinone oxidoreductase. Both generate a sodium gradient that is used for ATP production and flagella rotation (Biegel *et al.*, 2011). This variety of electron transport chains probably allowed Ri-A to cope with different environmental conditions related for instance to oxygen concentration, while Hd-A was more limited in this sense.

Thus, the *H. defensa* lineage underwent an early reduction of its potential niche, while Ri-A was probably able to survive in different environment conditions. Differences in terms of niche influence the effective size of the populations. It should have an impact on the action of drift.

## 8. Selective pressures

In order to further test the hypothesis that the effective size was lower in the *H.*

*defensa* lineage, we tried to determine the ratio between non synonymous and synonymous mutations (dN/dS), which is a good indicator of the efficiency of purifying selection. Because of the high substitution rates, it has not been possible to calculate the dN/dS value for the majority of the genes, the dS values being often saturated. Only 251 sequences presented no saturation between *Y. enterocolitica* and the endosymbionts. For these sequences, no significant differences were observed between *Y. enterocolitica* and, respectively, the *H. defensa* and *R. insecticola* bacteria (p-value = 0.3). It is obviously not possible to compare the dN/dS between HdA (or RiA) and *Y. enterocolitica*. Nevertheless, because most of the elapsed time occurred before the divergence between HdA (or RiA) and its descendants, the ratios that we calculated are a good approximation.

Our result implies that the effective size of the populations has been the same during the evolution of the two lineages, and could not explain the substitution rates.

## IV. Discussion

### 1. Parsimony and comparative genomics

If all the members of a clade share a similar phenotype, the principle of parsimony proposes that they have inherited it from their last common ancestor. Following this simple way of thinking, it has been suggested that the ancestor of the *Regiella-Hamiltonella* clade was a protective endosymbiont (Baumann, 2005). In contrast, due to evolutive convergences (or homoplasy), the members of a clade can independently acquire a similar phenotype even if their ancestor did not possess it. In our case, this alternative scenario had to be considered because its likelihood was far from being null. Indeed, these bacteria belong to the clade of Enterobacteriaceae, which established several independent endosymbiotic relationships. Moreover, at least two protective endosymbioses emerged from two relative clades of Enterobacteriaceae, the genus *Serratia* on one hand, and the clade composed of the genera *Yersinia*, *Hamiltonella* and *Regiella* on the other. This could suggest that the bacteria of these clades could be predisposed to a protective endosymbiosis, probably because of the virulence factors that they inherited from their pathogenic ancestor (Degnal *et al.*, 2009; Degnan and Moran, 2011). For example, such a predisposition exists in the genus *Actinobacteria*, which synthesizes an important cocktail of secondary metabolites that have been hypothesized to act as antibiotics and to protect their insect hosts (Kaltenpoth, 2009).

It was therefore not clear whether the ancestor of *H. defensa* and *R. insecticola* was an endosymbiont. We used markers that allowed to determine the more likely scenario between these two; the first proposing that their ancestor already lived inside host cells, the second that this ancestor was a free-living bacterium with a strong predisposition for protective endosymbiosis. This is what we did using the genomes of the endosymbionts, which gave us access to an important mass of information. Indeed, if one gene could have followed a distinct trajectory from the entire genome, this is not the case if one considers thousands of genes. This principle is used in the phylogenetic reconstructions, the concatenation of several genes or proteins allowing approaching the actual evolution of the studied organisms. In our case, the genes lost represented a relevant marker to determine the more likely hypothesis. Indeed, the losses in *H. defensa* and *R. insecticola* are more

similar than the losses between each of them and *S. symbiotica*. The likelihood to have obtained this pattern by chance is close to zero.

## 2. From an ancestor to another

We inferred the gene content of the ancestor of *H. defensa* and *R. insecticola* based on the one of its descendants. This was possible only because this ancestor was proven to be an endosymbiont. If that had not been the case, such reconstruction might have been very far from the real gene content of the ancestor, which would have been closer to the free-living core genome. The ancestor of the two lineages thus retained at least 1100 of the 2262 genes which composed this core. It is important to keep in mind that we missed in the reconstruction all the genes that have been convergently lost by both *H. defensa* and *R. insecticola* after their divergence.

This reconstruction was interesting because it allowed understanding what happened along the branches leading to the two lineages. We thus distinguished several losses that occurred specifically in the *H. defensa* lineage. Even if dozens of genes have been lost, the more interesting concern electron transport chains (at least four systems), which are directly associated to the niches and micro-niches the endosymbiont can occupy. These systems were present in RiA, and have been differently lost in the *R. insecticola* strains that have been sequenced (one loss in the strain 5.15, three in the strain LSR1). Moreover, in both strains of *R. insecticola*, genes composing these systems are still present, which suggests that these losses are recent, while there is no trace of any of them in the *H. defensa* strains, except for a fragment of the *cydB* gene in hApis\_5AT and hBtab\_MEAM1. *R. insecticola* lineage possessed (and still possesses) many systems that allowed it to survive into different environments, and thus a probably more extended niche inside the insect body than *H. defensa*.

This ecology however does not seem to have had any repercussion on the respective actions of selection and drift if we consider the ratio between synonymous and non-synonymous mutations. This is not totally surprising. A more extended niche leads to more important populations and thus to an apparently important effective size. Nevertheless, the action of drift mostly occurs during the bottleneck represented by maternal transmission of

the bacteria, which probably is (and has always been) of the same order of magnitude in *H. defensa* and *R. insecticola*. This differential in terms of potential ecology thus does not seem to explain the observed substitution rates.

### 3. Mutation rates in the two lineages

We can however ask the reverse question: could such substitution rates and the mechanism(s) that underlie them explain the ecological differences? It is possible that deleterious mutations regularly affect the genes involved in the invasive power of a bacterium, and thus lead to a reduction of its tissue tropism. Gene losses are seen by some as an inevitable process, which progressively leads to the extinction of an endosymbiont. If *H. defensa* had a stronger mutation rate, it could accelerate this ineluctable fate.

The analysis of the genes involved in DNA repair goes in the sense of this hypothesis. *Prochlorococcus*, which is a bacterium living in large populations and under strong selective pressures, has an accelerated rate of mutations, probably due to the loss of DNA repair genes (Dufresne *et al.*, 2005). As in our case, the loss of only a few genes involved in DNA repair would be sufficient to explain such a pattern. Only three such genes have been lost in *H. defensa*, and one in *R. insecticola* (*ada*), but this latter gene underwent deletions in *H. defensa* which cast doubts on its efficiency. The loss of DNA repair genes could be sufficient to explain the observed pattern of different substitution rates. The mutation rates in *H. defensa* could be more important, and thus the probability that some of these mutations reach fixation increases. This means that a gene would have a higher probability to be inactivated in *H. defensa*, which would favor its elimination and explain the losses in this lineage. Because of the universal bias towards adenine and thymine, an increased substitution rate also explains the stronger AT enrichment in *H. defensa*.

### 4. Host selection level

If mutation rates are high in *H. defensa*, new variants can frequently appear and get fixed **within one host** because of the action of drift. Insects belonging to a same population can thus carry different variants of *H. defensa*. This inter-host polymorphism allows the more beneficial *H. defensa* to be selected and fixed **within host populations**, which reduces the polymorphism of *H. defensa* at the scale of the entire host population. This frequent

“selective sweeps” could thus derive directly from the increased mutation rates, and participate to the observed pattern of substitution rates.

This is also interesting if we take into account the protective role of these endosymbionts. It has been demonstrated both experimentally and *in silico* that the presence of pathogenic viruses could favor the spread of “mutators”, *i.e.* of fast evolving bacteria (Pal *et al.*, 2007). This is particularly due to the loss of genes involved in DNA repair (Pal *et al.*, 2007). These mutators would be advantaged compared to slow-evolving bacteria in the arms race engaged against the viruses. If we replace viruses by parasitoids, this experience suggests that the “mutator” *H. defensa* could have an advantage over *R. insecticola*. An increased mutation rate which leads to inter-host variability could allow positive selection to act in response to the presence of parasitoids. Nevertheless, this is totally speculative, as both are protective endosymbionts, and no element allows to sustain this hypothesis.

## V. Conclusion

The close bacteria of the species *R. insecticola* and *H. defensa* can be distinguished by their different substitution rates. The more likely hypothesis is a loss of DNA repair genes in the latter, which leads to an increase in the mutation rates. These mutation rates explain the acceleration in *H. defensa* of the processes that occur in all endosymbionts. They could also explain the losses undergone by the genomes of this species, deleterious mutations being so frequent that they participate progressively to the erosion of the genome of *H. defensa*, which could precipitate the extinction of this species. These mutation rates could also play a role in the protection conferred by the bacteria, but no element sustains this idea.



## REFERENCE

- Allen, J.M., Light, J.E., Perotti, M.A., Braig, H.R. & Reed, D.L. **2009**. Mutational meltdown in primary endosymbionts: selection limits Muller's ratchet. *PLoS One*, **4**, e4969.
- Baumann P. **2005**. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annual Review of Microbiology*, **59**, 155-189.
- Buchner P. **1965**. *Endosymbiosis of animals with plant microorganisms*. New York: John Wiley.
- Burke G.R. & Moran N.A. **2011**. Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biology and Evolution*, **3**, 195-208.
- Chain P.S., Carniel E., Larimer F.W., Lamerdin J., Stoutland P.O., Regala W.M. *et al.* **2004**. Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 13826-13831.
- Charif D. & Lobry J.R. **2007**. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In *Structural approaches to sequence evolution* (pp. 207-232). Springer Berlin Heidelberg.
- Chung W.C., Chen L.L., Lo W.S., Kuo P.A., Tu J. *et al.* **2013**. Complete genome sequence of *Serratia marcescens* WW4. *Genome announcements*, **1**, e00126-13.
- Degnan P.H., Yu Y., Sisneros N., Wing R.A. & Moran N.A. **2009**. *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proceedings of the National Academy of Sciences*, **106**, 9063-9068.
- Degnan P.H., Leonardo T.E., Cass B.N., Hurwitz B., Stern D., Gibbs R.A. *et al.* **2010**. Dynamics of genome evolution in facultative symbionts of aphids. *Environmental Microbiology*, **12**, 2060-2069.
- Dennis Jr G., Sherman B.T., Hosack D.A., Yang J., Gao W., Lane H.C. *et al.* (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome biology*, **4**, P3.
- Foray V., Grigorescu A.S., Sabri A., Haubruge E., Lognay G., Francis F. *et al.* **2014**. Whole-genome sequence of *Serratia symbiotica* strain CWBI-2.3 T, a free-living symbiont of the black bean aphid *Aphis fabae*. *Genome announcements*, **2**, e00767-14.
- Guindon S., Delsuc F., Dufayard J.F. & Gascuel O. **2009**. Estimating maximum likelihood phylogenies with PhyML. In *Bioinformatics for DNA Sequence Analysis* (pp. 113-137). Humana Press.
- Hansen A.K., Vorburger C. & Moran N.A. **2012**. Genomic basis of endosymbiont-conferred protection against an insect parasitoid. *Genome Research*, **22**, 106-114.
- Horst J.P., Wu T.H., Marinus M.G. **1999**. *Escherichia coli* mutator genes. *Trends in Microbiology*, **7**, 29-36.
- Hurst G.D.D. & Jiggins F.M. **2005**. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society of London Series B*, **272**, 1525-1534.
- Kato A. & Inoue H. **2006**. Growth defect and mutator phenotypes of RecQ-deficient *Neurospora crassa* mutants separately result from homologous recombination and nonhomologous end joining during repair of DNA double-strand breaks. *Genetics*, **172**, 113-125.

- Lamelas A., Gosalbes M.J., Manzano-Marín A., Peretó J., Moya A. & Latorre A. **2011**. *Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont. *PLoS Genetics*, **7**, e1002357.
- Lucas-Lledó J.I. & Lynch M. **2009**. Evolution of mutation rates: phylogenomic analysis of the photolyase/cryptochrome family. *Molecular Biology and Evolution*, **26**, 1143-1153.
- Mackay W.J., Han S. & Samson L.D. **1994**. DNA alkylation repair limits spontaneous base substitution mutations in *Escherichia coli*. *Journal of Bacteriology*, **176**, 3224-3230.
- Manzano-Marín A., Lamelas A., Moya A. & Latorre A. **2012**. Comparative genomics of *Serratia* spp.: two paths towards endosymbiotic life. *PLoS One*, **7**, e47274.
- Miele V., Penel S. & Duret L. **2011**. Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC bioinformatics*, **12**, 116.
- Moran N.A. **1996**. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences*, **93**, 2873-2878.
- Moran N.A., Dunbar H.E. & Wilcox J.L. **2005a**. Regulation of transcription in a reduced bacterial genome : nutrient-provisioning genes of the obligate symbiont *Buchnera aphidicola*. *Journal of Bacteriology*, **187**, 4229-4237.
- Moran N.A., Degnan P.H., Santos S.R., Dunbar H.E. & Ochman H. **2005b**. The players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. *Proceedings of the National Academy of Sciences*, **102**, 16919-16926.
- Moran N.A., McCutcheon J.P. & Nakabachi A. **2008**. Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics*, **42**, 165-190.
- Nakabachi A., Yamashita A., Toh H., Ishikawa H., Dunbar H.E., Moran N.A. *et al.* **2006**. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science*, **314**, 267.
- Neupane S., Högberg N., Alström S., Lucas S., Han J., Lapidus A. *et al.* **2012**. Complete genome sequence of the rapeseed plant-growth promoting *Serratia plymuthica* strain AS9. *Standards in genomic sciences*, **6**, 54.
- Nogge G. **1982**. Significance of symbionts for the maintenance of an optional nutritional state for successful reproduction in hematophagous arthropods. *Parasitology*, **82**, 101-104.
- Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G., Prentice M.B. *et al.* **2001**. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature*, **413**, 523-527.
- Oliver K.M., Degnan P.H., Burke G.R. & Moran N.A. **2010**. Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annual Review of Entomology*, **55**, 247-266.
- Shoemaker D.D., Dyer K.A., Ahrens M., McAbee K. & Jaenike J. **2004**. Decreased diversity but increased substitution rate in host mtDNA as a consequence of *Wolbachia* endosymbiont infection. *Genetics*, **168**, 2049-2058.
- Sloan D.B. & Moran N.A. **2013**. The evolution of genomic instability in the obligate endosymbionts of whiteflies. *Genome Biology and Evolution*, **5**, 783-793.
- Tamas I., Klasson L., Canbäck B., Näslund A.K., Eriksson A.S., Wernegreen J.J. *et al.* **2002**. 50 million years of genomic stasis in endosymbiotic bacteria. *Science*, **296**, 2376-2379.

Toh H., Weiss B.L., Perkin S.A., Yamashita A., Oshima K., Hattori M. *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.

Thomopson J.D., Higgins D.G. & Gibson T.J. **1994**. ClustalW. *Nucleic Acids Research*, **22**, 4673-4680.

Thomson N.R., Howard S., Wren B.W., Holden M.T., Crossman L., Challis G.L. *et al.* **2006**. The complete genome sequence and comparative genome analysis of the high pathogenicity *Yersinia enterocolitica* strain 8081. *PLoS Genetics*, **2**, e206.

UniProt Consortium. **2008**. The universal protein resource (UniProt). *Nucleic acids research*, **36**, D190-D195.

Van Dijk T. 2008. *Microbial ecology research trends*. Nova Publishers.

## BILAN

Malgré de nombreuses convergences dans les pertes de gènes chez les symbiotes, nous avons tout de même pu écarter l'hypothèse d'un ancêtre libre aux espèces *H. defensa* et *R. insecticola* : la ressemblance entre les deux est telle qu'elle suppose clairement un mode de vie intracellulaire pour leur dernier ancêtre commun. Cela implique que cette bactérie avait perdu de nombreux gènes, ce qui est l'évolution attendue dans pareilles conditions de vie.

Après la divergence entre les deux espèces, *H. defensa* aurait perdu de nombreux gènes liés à la respiration cellulaire, ce qui aurait pu réduire la niche potentielle du symbiote, et donc la taille efficace de ses populations. Nous n'avons pas décelé de différences en termes de dN/dS entre les deux espèces (et donc d'action de la dérive), mais les valeurs proches de la saturation se prêtaient sans doute guère à cette analyse. Les deux lignées ont toutefois perdu différents gènes liés à la réparation de l'ADN, ces pertes étant plus importantes qualitativement chez *H. defensa*. Ainsi, les taux de mutations seraient différents dans ces deux lignées, conduisant au patron que l'on observe.

Cette analyse n'est pas aboutie. Comme on le voit, de nombreuses hypothèses, non exclusives les unes des autres, peuvent être formulées. Il y a tout de même une corrélation entre différents événements (accélération du taux de substitution, enrichissement en bases A et T, pertes de gènes de réparation de l'ADN, réduction de la niche écologique etc.) qui abondent dans le même sens. En ce cas, difficile d'imaginer la cause, ainsi que ses conséquences. La simple perte de gènes liés à la réparation peut-elle amener à cet état de fait ? Cela aurait eu pour conséquence un taux plus important de mutations chez *H. defensa*, conduisant à une accélération globale du taux de substitution, et à un enrichissement en AT. Des taux importants de mutations dans des organismes réduisent l'efficacité de la sélection (Marais *et al.*, 2008). Cela aurait pour conséquence une perte régulière de gènes chez *H. defensa*, réduisant de manière progressive la niche potentielle du symbiote. Cette perte de gènes liés à la réparation de l'ADN aurait-elle pu être sélectionnée ? Peut-on faire un lien avec la protection conférée par la bactérie ?

En effet, les conséquences des taux de substitutions importants chez *H. defensa* n'ont pas encore fait l'objet d'une analyse approfondie. Y a-t-il des catégories davantage marquées par les mutations chez *H. defensa*? Notamment, trouve-t-on des traces de sélection positive en ce qui concerne les facteurs de virulence de la bactérie, potentiellement en lien avec la course aux armements engagée contre les ennemis de l'hôte? En effet, il est certain que ce phénomène a une répercussion sur les associations entre les bactéries et leur hôte. Des taux de mutations importants chez les populations clonales pourraient être une alternative au brassage par recombinaison sexuée (Tenailon *et al.*, 2000), et donc conférer un avantage à la bactérie dans le cadre de la course aux armements.

# DISCUSSION

---

Dans le cadre de cette thèse, nous nous sommes intéressés de manière générale à l'évolution des génomes des bactéries symbiotiques. Nous nous sommes focalisés sur l'espèce *Hamiltonella defensa*, aux deux associations qu'elle forme avec les Hémiptères et aux partenaires avec qui elle interagit dans le cadre de ces deux symbioses. L'hôte bien sûr est un partenaire privilégié de la bactérie, qu'il soit le puceron ou l'aleurode. Cependant, d'autres protagonistes ont attiré notre attention. Nous nous sommes ainsi intéressés aux interactions métaboliques entre les espèces symbiotiques au sein de leur hôte, ainsi qu'aux pressions de sélection exercées par les parasitoïdes sur les génomes des bactéries protectrices.

Deux aspects nous ont davantage concernés : la génomique, qui permet d'obtenir une « image » du présent, de connaître le contenu génique d'un organisme et par extension d'en inférer les capacités (métabolisme, virulence, perception environnementale etc.), et son prolongement direct, la génomique **comparative**, qui permet de replacer cette image dans un contexte plus large, et notamment de comprendre le scénario évolutif ayant mené à la situation actuelle.

Nous allons discuter les résultats obtenus au moyen de ces outils, et voir que de nombreuses hypothèses peuvent être formulées, et des expériences proposées pour les tester. Nous allons partir de la symbiose entre l'aleurode et *H. defensa*, avant d'étendre progressivement notre discussion à l'ensemble de l'espèce *H. defensa*.

# I. Complémentation métabolique entre les endosymbiotes de *Bemisia tabaci*

En premier lieu, nous nous sommes intéressés aux phénomènes de complémentation métabolique, du scénario ayant conduit à leur apparition aux conséquences pour le consortium insectes-bactéries symbiotiques, en passant par les conditions nécessaires à ce que de tels événements se produisent.

Nous avons étudié pour cela les génomes de deux symbiotes colonisant les mêmes cellules de *Bemisia tabaci* MED-Q1 : *Portiera aleyrodidarum* et *Hamiltonella defensa*. Cette reconstruction du génome d'un organisme nous a permis d'obtenir une image assez précise de ses capacités de synthèse, de perception ou de régulation. Il apparaît en effet assez logique de considérer comme fonctionnels tout système strictement conservé et sous sélection, toute voie de synthèse entièrement présente.

## 1. Complémentation inter-voies

Le génome de l'endosymbiote primaire *P. aleyrodidarum* révèle un métabolisme dédié à la synthèse de deux types de métabolites, les caroténoïdes et les acides aminés essentiels. Si ces derniers sont fréquemment synthétisés par les symbiotes primaires (Moran *et al.*, 2008), les caroténoïdes sont plus rarement apportés à leur hôte par un symbiote, et pourraient jouer un rôle antioxydant important chez les Hémiptères, comme nous l'avons déjà évoqué (voir **Article Premier – Bilan**). Contrairement à de nombreux symbiotes obligatoires (Moran *et al.*, 2008), *P. aleyrodidarum* ne synthétise aucun cofacteur, cette tâche étant dévolue au symbiote « secondaire » (je reviendrai sous peu sur l'utilisation de ces guillemets) *H. defensa*.

Nous avons qualifié cette « répartition des tâches » de « complémentations inter-voies ». Elle est intéressante car pourrait justifier à elle seule le maintien des symbiotes secondaires. En effet, ces métabolites synthétisés par *H. defensa* ne sont pas essentiels, mais peuvent conférer un bénéfice à l'hôte dans certaines conditions. Nous avons déjà vu que les différentes espèces de puceron hébergeaient des symbiotes primaires aux capacités de synthèse de cofacteurs variées, ce qui va jouer sur le spectre de plantes consommables

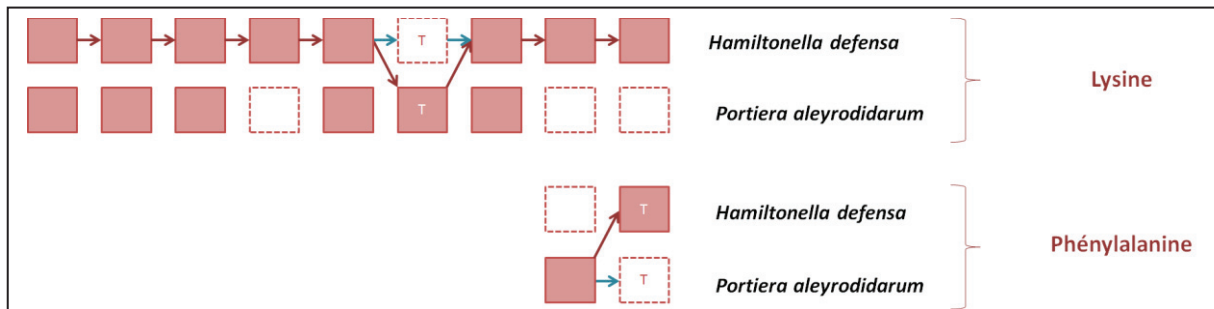
(Nakabachi et Ishikawa, 1999). *B. tabaci* MED-Q1 se caractérise par son importante polyphagie (Byrne et Bellows, 1991), ce qui implique que l'importance de cette synthèse de cofacteurs par son symbiote devrait varier avec les plantes hôtes. En ce cas, la prévalence du symbiote secondaire devrait différer selon les plantes infester. Selon cette hypothèse, la quasi-fixation d'*H. defensa* chez *B. tabaci* MED-Q1 nécessite une autre explication.

Il est toutefois possible que la fréquence à laquelle l'insecte change de plantes hôtes soit très élevée, un même individu consommant diverses plantes au cours de sa vie. Ainsi, les insectes qui perdraient *H. defensa* seraient rapidement contre-sélectionnés, puisque certaines plantes ne leur apporteraient pas l'ensemble des cofacteurs requis. En ce cas, la fixation de la bactérie pourrait être due à ces cofacteurs qu'elle produit, et qui permettent la polyphagie de son hôte.

## 2. Complémentations intra-voies

Une seconde hypothèse, davantage soutenue par les données génomiques, peut expliquer à elle seule la fixation du symbiote. En effet, le génome de *P. aleyrodidarum* démontre que la bactérie n'est pas capable de synthétiser seule plusieurs acides aminés essentiels. Pour la leucine, l'isoleucine et la valine, ce serait l'hôte qui permettrait de compléter les réactions manquantes, comme c'est le cas chez de nombreux Hémiptères (revue dans Sloan et Moran, 2013). Pour la lysine et la phénylalanine, ce serait *H. defensa* le partenaire métabolique de *P. aleyrodidarum* (**Figure 1**). Nous avons qualifié ces phénomènes de complémentations métaboliques « intra-voie ». Il s'agit en ce cas d'une répartition *in silico* des différentes étapes de la synthèse d'un métabolite entre – au moins – deux partenaires.





**Figure 1. Complémentation métabolique potentielle chez *Bemisia tabaci***

*In silico*, deux acides aminés essentiels requièrent les capacités métaboliques des deux endosymbiotes pour être synthétisés, la lysine et la phénylalanine (flèches rouges – Section I-2 de cette discussion). Une hypothèse alternative, représentée cette fois-ci par des flèches bleues, propose que les étapes manquantes puissent être assurées par une autre enzyme (plus précisément, une autre transaminase, d'où le « T » en légende). *H. defensa*, dont le génome encode de nombreuses transaminases, synthétiserait ainsi seule la lysine, tandis que *P. aleyrodidarum* de par ses capacités métaboliques réduites pourrait utiliser une transaminase de l'hôte (Section I-3 de cette discussion).

Cette complémentarité suppose un dialogue métabolique précis entre les partenaires, et nécessite des transporteurs et régulateurs efficaces assurant le transit des métabolites. L'hôte lui-même jouerait un rôle important à la fois en amont et en aval de la synthèse de différents métabolites essentiels, en fournissant les sources nécessaires et en participant à la dernière étape de la synthèse de certaines d'entre elles. Le fait que les deux symbiotes partagent les cellules pourrait être une des clés nécessaires à ce dialogue, la promiscuité favorisant logiquement les échanges. Ainsi, la complémentarité métabolique est importante entre les deux endosymbiotes primaires *Tremblaya princeps* et *Moranella endobia*, le second ayant colonisé le cytoplasme du premier, véritable « matriochka » bactérienne (Lopez-Madrigo *et al.*, 2011). La complémentarité métabolique (inter et intra-voies) semble toutefois permise dans un système où les symbiotes ne cohabitent pas, comme c'est le cas chez le puceron *Cinara cedri* qui héberge ses deux symbiotes primaires dans deux types distincts de bactériocytes (Gómez-Valero *et al.*, 2004).

### 3. Transaminases et plasticité enzymatique

Ben évidemment, rien n'indique qu'une quelconque complémentarité ait lieu. En ce qui concerne la phénylalanine, il est même envisageable que *P. aleyrodidarum* en assume la synthèse sans *H. defensa*. En effet, nous savons grâce au séquençage du génome du symbiote primaire de *Trialeurodes vaporarum*, un aleurode proche du complexe de *B.*

*tabaci*, que l'ancêtre du genre *Portiera* ne possédait pas le gène codant pour la dernière étape de la synthèse de la phénylalanine. Cette perte serait donc antérieure à leur dernier ancêtre commun, et partagée par l'ensemble des symbiotes primaires de *B. tabaci*.

Or, dans certaines espèces du complexe comme l'espèce *B. tabaci* Océan Indien, très proche de nos espèces MED et MEAM1, les symbiotes secondaires présentent des prévalences intermédiaires (Gueguen *et al.*, 2010 ; Thierry *et al.*, 2011). La plupart des individus dans ces espèces ne peuvent donc pas compter sur quelque symbiote secondaire que ce soit pour assurer la fonction manquante. De même, il n'est pas fait mention de la présence d'autres symbiotes dans l'article présentant le génome de *P. aleyrodidarum* chez *T. vaporarium* (Sloan et Moran, 2013). Il est donc possible que le génome de *P. aleyrodidarum* ou celui de son hôte encode en réalité une enzyme capable d'assurer la dernière étape de la synthèse de la phénylalanine, tant chez *B. tabaci* que chez *T. vaporarium*. Cela est d'autant plus probable que la réaction manquante est assurée par une enzyme appartenant à la classe des transaminases, enzymes peu spécifiques et aux nombreux substrats (Rothman et Kirsch, 2003).

De manière analogue, de nombreuses transaminases absentes des génomes des symbiotes primaires sont (ou seraient) compensées par des enzymes de l'hôte. C'est le cas chez la cochenille, où les transaminases de l'hôte assureraient des fonctions perdues par les deux symbiotes primaires *M. endobia* et *T. princeps* (Husnik *et al.*, 2013), ou chez le puceron, où de nombreuses enzymes de l'hôte complètent le métabolisme de la bactérie (Wilson *et al.*, 2010). Quant à l'enzyme permettant la dernière réaction de la synthèse de la phénylalanine, elle a été perdue par convergence dans de nombreux génomes de symbiotes primaires (*B. aphidicola*, *C. ruddii*, *T. princeps*, *Ishikawella capsulata* etc.), et Hansen et Moran (2014) proposent que leurs hôtes respectifs assurent cette fonction.

Le même raisonnement peut être appliqué à la synthèse de la lysine par *H. defensa*. L'ancêtre du genre *Hamiltonella* avait en effet déjà « perdu » le gène codant pour la réaction manquante, là encore une transaminase. Il est possible qu'une autre enzyme assure la réaction manquante, et qu'en réalité *H. defensa* chez l'aleurode assure seule la synthèse de la lysine.

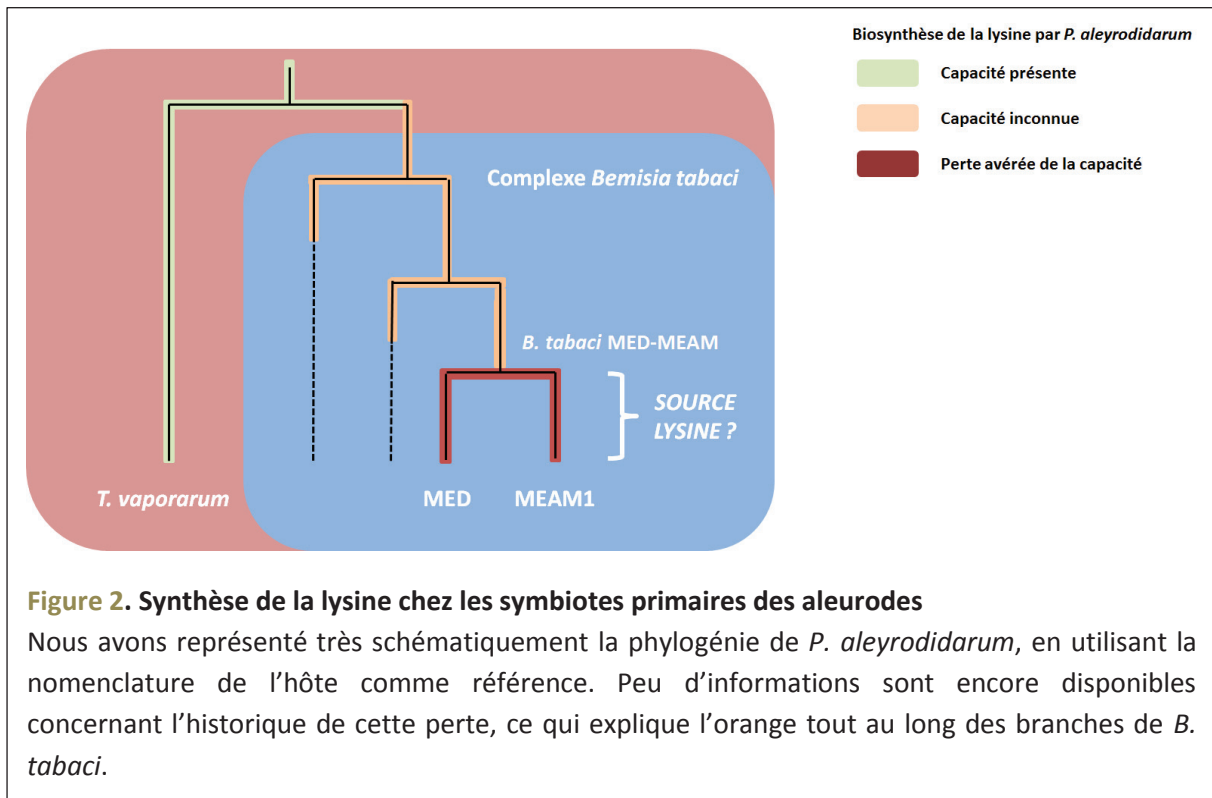
## II. *Hamiltonella defensa* : un endosymbiote co-primaire de l'aleurode ?

*H. defensa* est donc impliquée dans la synthèse de la lysine chez *B. tabaci* MED-Q1 (voir **Article Premier**) et MEAM1 (voir **Article Second**), qu'elle en assume seule ou non la production. Cela est très intéressant au regard du caractère essentiel de ce métabolite. Cette incapacité du symbiote primaire à le synthétiser seul signifie que le supposé symbiote « facultatif » *H. defensa* serait en réalité une composante obligatoire pour les holobiontes *B. tabaci* MED-Q1 et MEAM1. Ce statut est totalement plausible si l'on se réfère à la fixation de ce symbiote dans les populations d'aleurodes MED-Q1 et MEAM1 (Gueguen *et al.*, 2010). En outre, comme nous l'avons déjà évoqué, le caractère nutritif de cette symbiose corrobore une étude de Lu et de ses collaborateurs (2014), qui démontrent l'importance d'*H. defensa* en cas de stress alimentaire.

Ce qui est encore plus intéressant est le scénario qui a conduit à ce statut obligatoire. Nous avons en effet maintes et maintes fois abordé la perte de gènes inéluctables chez les organismes endosymbiotiques. Ces gènes perdus par les endosymbiotes sont à l'origine de leur **dépendance** à l'hôte et à l'environnement intracellulaire. De manière intéressante, dans notre exemple, cette propension à la perte de gènes a conduit à la probable dépendance **de l'hôte** à un nouveau partenaire. C'est ce scénario que nous allons décrire pour expliquer la probable dépendance de *B. tabaci* à *H. defensa*.

### 1. Pertes des capacités biosynthétiques chez *Portiera aleyrodidarum*

Revenons sur la perte de la capacité à synthétiser la lysine par *P. aleyrodidarum*. Contrairement à ce qu'on observe pour la phénylalanine, le symbiote primaire de *T. vaporarium* était tout à fait capable, lui, de synthétiser la lysine. Cela signifie que des inactivations ont eu lieu de manière spécifique chez *B. tabaci*. En outre, les mêmes gènes sont affectés par les mêmes mutations dans les génomes de *P. aleyrodidarum* MED-Q1 et MEAM1. Cela signifie que la perte de cette capacité aurait eu lieu chez une bactérie ancestrale à ces deux symbiotes, et ne serait pas le fait de convergences (**Figure 2**). Hélas, nous ne disposons pas des génomes des symbiotes primaires pour les autres espèces de *B. tabaci*, qui nous permettraient de connaître plus précisément l'historique de cette perte.



## 2. Source de lysine dans l'holobionte ancestral

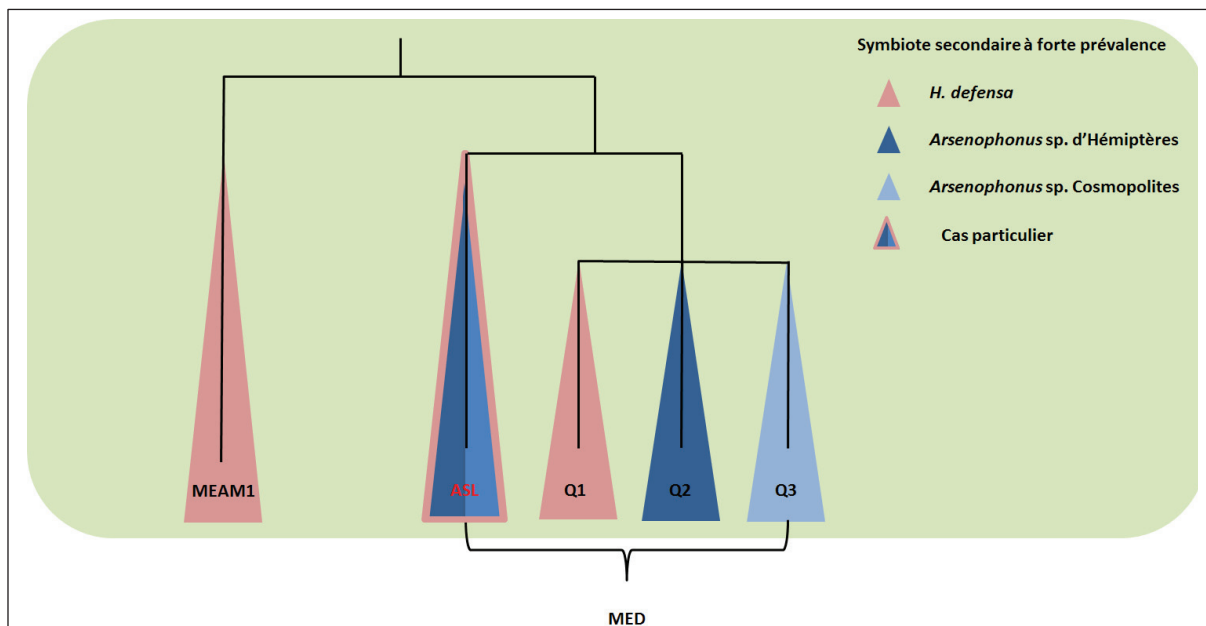
Depuis le dernier ancêtre commun aux espèces MED et MEAM1, que nous appellerons *B. tabaci* MED-MEAM, la lysine n'est donc plus fournie par le symbiote primaire. Puisque cet acide aminé essentiel n'est pas synthétisé par les Métazoaires, l'hôte ancestral disposait obligatoirement d'une autre source pour son approvisionnement. Plusieurs hypothèses sont envisageables : un transfert de gènes dans le génome nucléaire de l'insecte, un spectre de plantes hôtes riches en lysine, ou encore la présence d'un second partenaire, présent dans le tube digestif ou les cellules de l'hôte, qui ait pris le relais de *P. aleyrodidarum* dans la synthèse de ce nutriment.

La dernière hypothèse est la plus vraisemblable selon les données disponibles. En effet, des analyses du transcriptome de l'hôte n'ont révélé aucune trace de gènes codant pour la synthèse de la lysine (Wang *et al.*, 2011). De plus, les deux espèces sont extrêmement polyphages et ne ciblent donc pas particulièrement les plantes selon leur contenu en acides aminés. Il est possible que cette caractéristique commune soit un héritage depuis leur dernier ancêtre, et donc que ce dernier ait eu un régime polyphage. Toutefois,

certaines cytotypes de l'espèce MED ne sont pas polyphages : le cytotype Q3, par exemple, se distingue en Afrique de l'Ouest par un spectre de plantes hôtes réduit, puisqu'on ne l'a retrouvé que sur la tomate et le lanturier (Gnankiné *et al.*, 2013). Il est donc difficile de discriminer entre une polyphagie ancestrale, perdue dans certains cytotypes, et une convergence évolutive chez les cytotypes polyphages.

Toutefois, il faudrait plus vraisemblablement aller chercher du côté de la symbiose la source en lysine de cet holobionte ancestral. Une donnée est intéressante : pratiquement toutes les populations des espèces *B. tabaci* MED et *B. tabaci* MEAM1 hébergent un symbiote – présumé – secondaire, mais dont la prévalence avoisine les 100% : soit *Arsenophonus* sp., soit notre susmentionnée *H. defensa* (Gueguen *et al.*, 2010 ; **Figure 3**). Si la nature de ce symbiote peut varier, la présence de symbiotes secondaires aux fortes prévalences donne du crédit à une compensation de la perte de capacité de *P. aleyrodidarum* par un second partenaire. Contrairement au cas d'*H. defensa*, nous ne connaissons malheureusement rien des capacités métaboliques d'*Arsenophonus* sp. Il serait intéressant d'en séquencer le génome, ou tout du moins les gènes impliqués dans la lysine. Nous assurer de ses capacités biosynthétiques donnerait plus de poids à cette théorie d'un symbiote secondaire nutritif dans les espèces *B. tabaci* MED et MEAM1. En blastant les gènes permettant la synthèse de la lysine contre les séquences du genre *Arsenophonus*, il apparaît que cette voie semble bien préservée chez *Arsenophonus nasoniae*, l'endosymbiote de *Nasonia vitripennis*.

Il est intéressant de rappeler que hormis MED et MEAM1, les autres espèces de *B. tabaci* ne présentent pas toutes ces fameux symbiotes secondaires proches de la fixation, et notamment l'espèce proche dite « Océan Indien » (Gueguen *et al.*, 2010 ; Thierry *et al.*, 2011). Cela implique que leurs symbiotes primaires assument seuls la synthèse de la lysine. Sous cette hypothèse, la perte de la capacité à synthétiser cet acide aminé chez *P. aleyrodidarum* pourrait avoir eu lieu très récemment, dans la branche conduisant aux deux espèces MED et MEAM1. Cela expliquerait pourquoi certains des gènes impliqués dans cette synthèse soient encore présents dans leurs génomes (**Article Premier**).

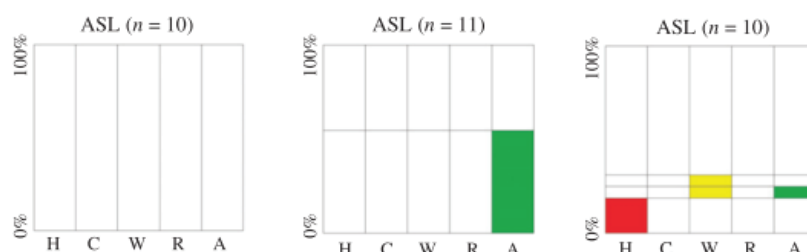


**Figure 3. Symbiotes secondaires « fixés » dans les populations de *B. tabaci* MEAM-MED**

Nous avons représenté une phylogénie simplifiée du clade MED-MEAM de *B. tabaci* (d'après Gnankiné *et al.*, 2013 ; de Baro *et al.*, 2011). Une troisième espèce vient normalement se glisser dans cette phylogénie et forme un clade monophylétique avec MEAM1, l'espèce *B. tabaci* MEAM2. De plus, l'espèce MED comporte 6 cytotypes : ASL, J, L, Q1, Q2 et Q3 (Gnankiné *et al.*, 2013 ; de Baro *et al.*, 2011). Cependant, nous ne connaissons pas le cortège symbiotique infectant *B. tabaci* MEAM2, MED-J et MED-L, ce qui explique que leurs branches soient absentes. Le cytotype ASL est infecté par *Arsenophonus* sp. comme par *H. defensa*, voire, certaines populations ne sont pas infectées par un quelconque symbiote secondaire.

### 3. Synthèse de la lysine chez *Bemisia tabaci* ASL

Revenons plus en détail sur le cas de *B. tabaci* MED-ASL. Ce cytotype pourrait en réalité être une espèce à part entière, génétiquement isolée des insectes du cytotype Q1 vivant en sympatrie (Mouton *et al.*, 2014). L'holobionte ASL est en outre un cas quelque peu particulier. Contrairement aux autres sous-espèces qui hébergent une seule espèce de symbiote secondaire « fixé », on retrouve dans les populations de *B. tabaci* ASL à la fois *H. defensa* et les deux souches d'*Arsenophonus* sp., de rarissimes individus étant infectés à la fois par *Arsenophonus* sp. et *H. defensa* (Gnankiné *et al.*, 2013).

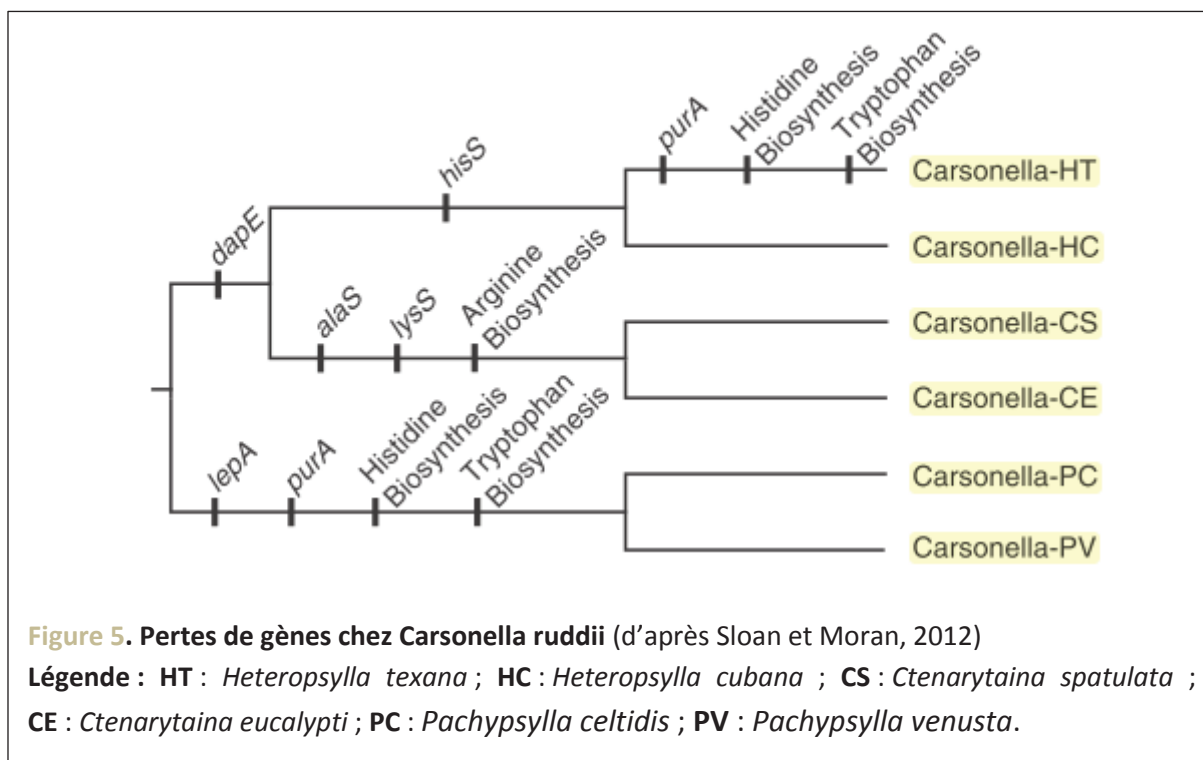


**Figure 4. Exemples de cortèges endosymbiotiques chez *Bemisia tabaci* ASL**

La prévalence de chaque symbiote est indiquée par différentes bandes verticales. Sont représentés *H. defensa* (H), *C. hertigii* (C), *Wolbachia* sp. (W), *Rickettsia* sp. (R) et *Arsenophonus* sp. (A). Les différents cortèges sont séparés par des lignes horizontales, et la taille de chaque bande indique la fréquence des différentes associations : dans la population figurée à gauche, aucun symbiote n'a été détecté parmi les dix individus testés ; six des onze individus constituant la population figurée au milieu étaient infectés par *Arsenophonus* sp. ; dans la population représentée à droite, deux individus se sont avérés être infectés par *H. defensa*, un individu par *Wolbachia* sp., et un individu à la fois par *Wolbachia* sp. et *Arsenophonus* sp.

Aucun de ces symbiotes ne semble pour autant proche de la fixation, et certaines populations de *B. tabaci* ASL sont d'ailleurs dépourvues de tout symbiote secondaire (Figure 4). Cette situation semble incompatible avec notre hypothèse d'une perte ancestrale chez *P. aleyrodidarum* MED-MEAM quant à la synthèse de la lysine. Différentes hypothèses peuvent être formulées quant à la provenance de cet acide aminé dans ce cytotype.

En effet, ce cas a déjà été identifié dans une autre Superfamille d'insecte : chez les psylles, des pertes régulières de fonctions métaboliques ont lieu chez le symbiote primaire, *Carsonella ruddii* (Sloan et Moran, 2012 ; Figure 5). En séquençant le métagénome des holobiontes correspondant, il a été démontré qu'un symbiote secondaire proche du symbiote *Moranella endobia* complémentait ces pertes pour trois des six hôtes analysés. En revanche, les gènes de la synthèse du tryptophane et de l'histidine, perdus par *C. ruddii* chez *Heteropsylla texana*, *Pachyopsylla celtidis*, et *Pachyopsylla venusta*, ne semblent pas compensés par la présence de symbiotes secondaires (Sloan et Moran, 2012). Nous allons revenir sur les hypothèses formulées par les auteurs de cette analyse chez les psylles, et voir si elles sont transposables à *B. tabaci* ASL.



Leur première hypothèse est qu'un régime alimentaire original chez les trois psylles hébergeant un symbiote primaire « déficient » pourrait permettre d'obtenir les acides aminés concernés directement depuis la plante. Notamment, les insectes du genre *Pachypsylla* induisent la formation de galles (Spaulding et von Dohlen, 2001), qui pourraient modifier la composition de la sève à leur avantage (Sloan et Moran, 2012 ;). Ces changements de régime alimentaire sont déjà documentés dans une autre Superfamille d'Hémiptères, les Phylloxeroïdes : ces insectes ne consomment non-plus le phloème mais les tissus parenchymateux des végétaux, véritable réserve de protéines (Vorwerk *et al.*, 2007). La lyse de ces dernières permet d'obtenir directement les acides aminés essentiels requis, de sorte qu'ils ne sont infectés par aucun symbiote primaire (Buchner, 1965). Dans le cas de *B. tabaci* ASL, un pareil changement de régime alimentaire ne semble pas évident, de sorte que cette hypothèse semble peu vraisemblable : on retrouve l'insecte sur les mêmes plantes que la polyphage espèce Q1, c'est-à-dire aussi bien sur la tomate, la courge, l'aubergine, le gombo ou le coton (Mouton *et al.*, 2014 ; Gnankiné *et al.*, 2013).

Une deuxième hypothèse formulée par Sloan et Moran (2012) et transposée à *B. tabaci* ASL impliquerait une acquisition par l'hôte des gènes nécessaires à la synthèse de la



lysine, par transfert horizontal. De pareils transferts ont déjà été mis en évidence chez les Hémiptères. Par exemple, les aleurodes synthétisent les caroténoïdes par le truchement de leur symbiote primaire, tandis que les pucerons les produisent seuls par l'intermédiaire de gènes fongiques incorporés dans le génome nucléaire (Moran et Jarvik, 2010).

Une troisième hypothèse serait que *B. tabaci* ASL n'appartienne pas au groupe monophylétique MED-MEAM. En ce cas, *P. aleyrodidarum* ASL pourrait n'avoir pas perdu sa capacité à synthétiser la lysine. Cela semble peu probable puisque les marqueurs utilisés pour distinguer les sous-espèces du clade MED-MEAM sont mitochondriaux (Khadsan *et al.*, 2005), de sorte que leur phylogénie est censée être parfaitement congruente avec les autres éléments à transmission maternelle stricte. Il serait en tous les cas intéressant d'obtenir le génome de *P. aleyrodidarum* ASL afin de déterminer ses potentielles capacités métaboliques.

### III. Évolution de la dépendance

#### 1. Redondance et dépendance

Selon notre hypothèse, un ancêtre commun à *B. tabaci* MED et *B. tabaci* MEAM1 hébergeait – au moins – deux symbiotes : un symbiote primaire *P. aleyrodidarum*, et un second symbiote, probablement facultatif, tous deux étant capables de synthétiser la lysine. Cette **redondance** fonctionnelle est à la base de la présence d'un second endosymbiote « primaire » dans *B. tabaci* MED-MEAM. En effet, dans pareille situation, que l'un ou l'autre des deux partenaires perde sa capacité à produire cet acide aminé essentiel n'a aucune conséquence létale pour l'holobionte. Cependant, si cette perte se produit chez le symbiote primaire, le second partenaire devient le seul à assumer la synthèse de la lysine : il devient donc **obligatoire** pour le consortium. C'est exactement ce scénario que nous retenons quant à la présence de deux symbiotes potentiellement obligatoires pour *B. tabaci* MED-MEAM. Dans la suite de ce manuscrit et pour simplifier le dialogue, nous qualifierons de symbiote « néo-primaire » cet endosymbiote potentiellement obligatoire (*Arsenophonus* sp. ou *H. defensa*), en prenant garde de ne point oublier le caractère hautement spéculatif de cette hypothèse, surtout en ce qui concerne *Arsenophonus* sp.

#### 2. Fixation du nouveau consortium

Nous avons détaillé le scénario évolutif par pertes de gènes ayant conduit au statut sans doute obligatoire d'un second symbiote chez *B. tabaci* MED-MEAM. Cependant, sa fixation dans la population ancestrale, et donc celle d'un *P. aleyrodidarum* inapte à synthétiser la lysine, reste à élucider.

En effet, ce dernier est un symbiote des aleurodes depuis près de 200 millions d'années (Moran *et al.*, 2008). L'on peut imaginer que la relation qu'il forme avec son hôte ait subi un profilage régulier, ayant conduit à une régulation précise des échanges entre les deux partenaires. Tout l'inverse de la relation entre l'hôte et le symbiote néo-primaire, dont l'âge récent ne permet certainement pas le même niveau d'interactions – on estime à 1 ou 2 millions d'années seulement la divergence entre les espèces MED et MEAM (Silva *et al.*, in prep.). Cette considération vaut très certainement pour la lysine : si les deux symbiotes primaires et secondaires étaient capables de la synthétiser chez l'ancêtre MED-MEAM, il est

toutefois probable que l'hôte accédât plus facilement à celle produite par *P. aleyrodidarum*, du fait de l'ancienneté de leur relation. Dès lors, qu'un second symbiote puisse remplacer le primaire dans cette fonction, mais surtout, que ce nouvel état envahisse la population peut laisser perplexe.

Il peut sembler intuitif que cette fixation ait nécessité que le symbiote néo-primaire apportât un autre bénéfice. D'ailleurs, sa présence initiale dans les populations de *B. tabaci* s'expliquait certainement par une fonction mutualiste telle que la protection contre des ennemis naturels, comme c'est le cas pour *H. defensa* chez le puceron, ou la synthèse de cofacteurs, si l'on se réfère aux capacités métaboliques du symbiote. Cependant, au moment de la perte de fonction chez *P. aleyrodidarum*, deux consortiums distincts bénéficiaient de la présence d'*H. defensa* : les insectes contenant un duo de bactéries toutes deux capables de synthétiser la lysine, et des insectes contenant ce même duo mais pour lesquels seul l'endosymbiote « secondaire » en était capable. Ainsi, les insectes contenant le symbiote primaire « défectueux » n'auraient pas dû être avantagés.

Il est possible que seule la dérive soit responsable de la fixation de ce nouveau consortium, qui aurait ainsi pu « naître » dans une population de petite taille. Autre hypothèse, la redondance métabolique pourrait induire un coût supplémentaire au niveau de l'holobionte. La perte de la capacité à synthétiser la lysine chez *P. aleyrodidarum* aurait en ce cas conféré une valeur adaptative supérieure aux hôtes concernés, par rapport à ceux contenant deux symbiotes capables de synthétiser l'acide aminé essentiel.

### 3. Mutualisme apparent

Ce scénario est en tous les cas tout-à-fait intéressant, car il implique une dépendance potentielle de l'hôte à ce symbiote néo-primaire non pas par **l'apport** d'une nouvelle fonction mais par la **perte** d'une fonction que l'holobionte ancestral assumait seul. Ce paradoxe de la perte conduisant au mutualisme ne correspond pas à la vision « **innovatrice** » de la symbiose, et l'on qualifie parfois ces relations de « mutualisme **apparent** ». Les exemples sont légion. Dans une des deux équipes auxquelles je suis affilié, c'est même une thématique d'intérêt, notamment à travers le modèle *Wolbachia/Asobara tabida*. Chez cet Hyménoptère, l'ovogénèse nécessite la présence de la souche wAtab3 de *Wolbachia*, sans

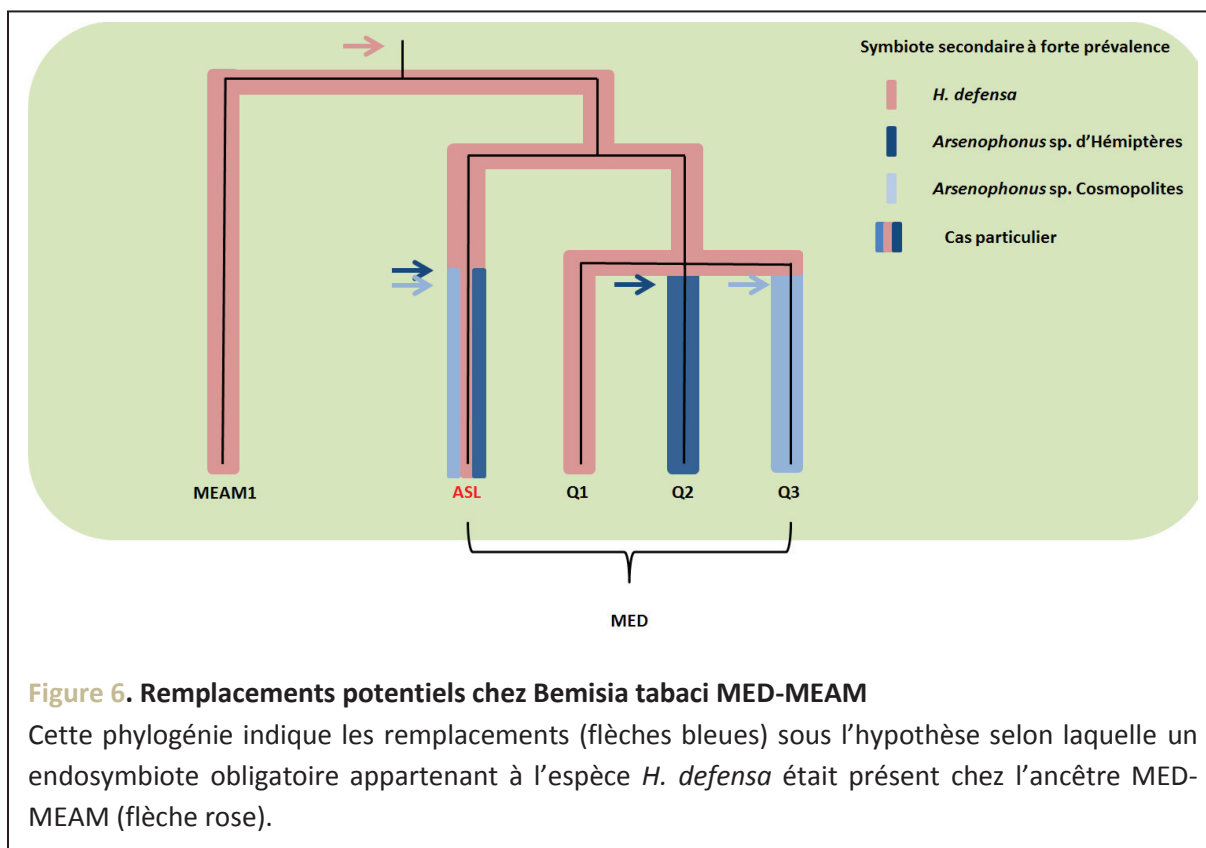
quoi les femelles sont stériles (Dedeine *et al.*, 2005). Pourtant, les espèces apparentées ne sont pas sous la dépendance de bactéries pour leur reproduction, de sorte que cette association obligatoire serait récente. Il y aurait donc eu la perte d'une fonction reproductrice normale chez *A. tabida*, directement imputable à la présence de *Wolbachia*. Sans entrer dans les détails, l'hôte se serait adapté à la présence de ce manipulateur de la reproduction, de sorte qu'il est désormais impossible d'éliminer la bactérie sans affecter la fertilité de son hôte (voir la thèse de Kremer, 2010).

#### 4. Symbiose et remplacements

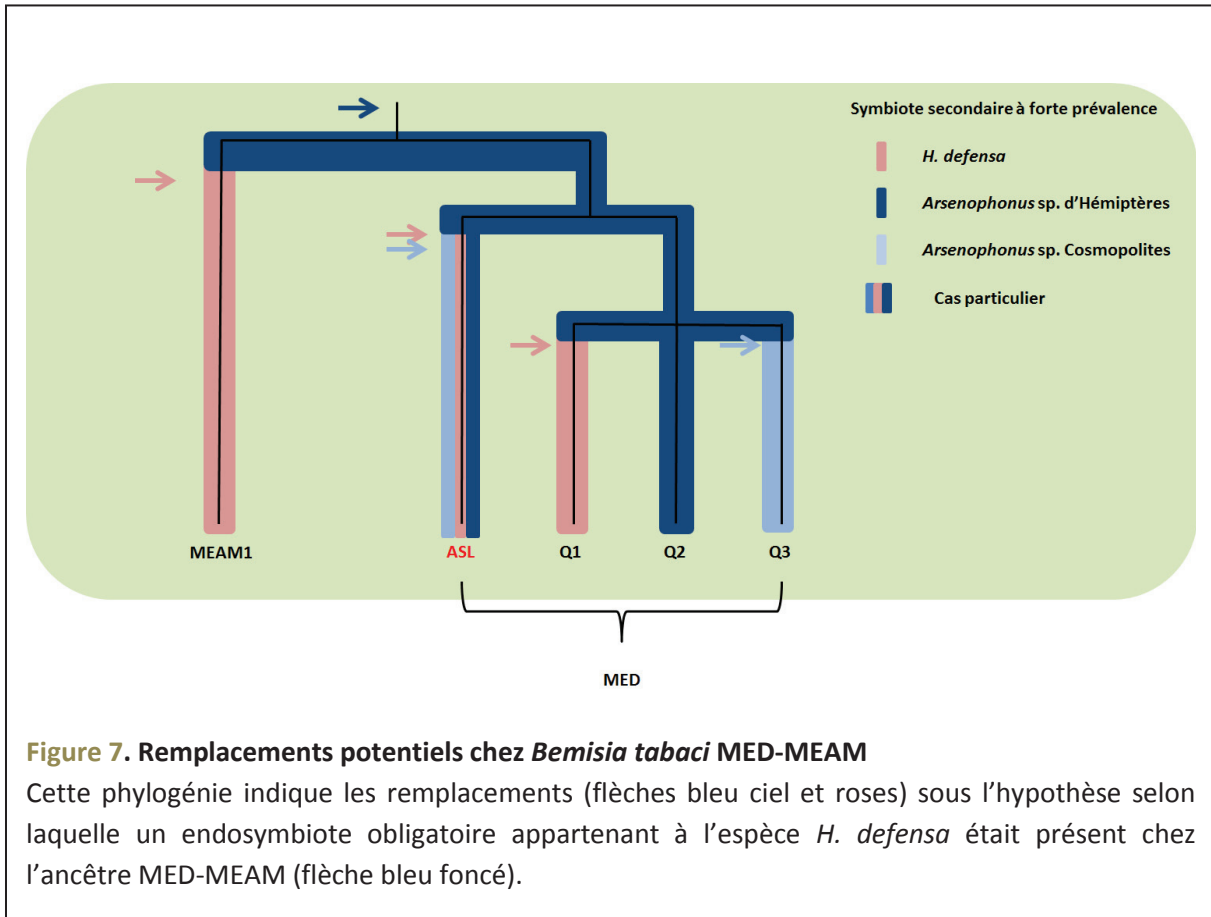
Nous avons évoqué dans les paragraphes précédents un « endosymbiote néo-primaire », sans jamais le nommer. Revenons sur l'identité de ce symbiote ancestral probablement devenu primaire. Dans notre **second article**, nous avons émis l'hypothèse qu'il puisse s'agir du dernier ancêtre d' *H. defensa* MED-Q1 et *H. defensa* MEAM1. En effet, les deux bactéries présentent une divergence nucléotidique qui exclut un transfert récent, et pourrait suggérer une co-spéciation entre *H. defensa* et l'hôte. Mais ces divergences pourraient également provenir de deux anciens transferts horizontaux indépendants d'*H. defensa* dans les espèces MED-Q1 et MEAM1. Dans ce cas, le second symbiote ancestral pourrait être une bactérie du genre *Arsenophonus*, voire d'une toute autre espèce. Dans tous les cas, la distribution des symbiotes dans les espèces et sous-espèces de *B. tabaci* MED et MEAM implique qu'aient eu lieu des **remplacements symbiotiques**.

##### a) Remplacements du symbiote néo-primaire chez *Bemisia tabaci*

Si le symbiote secondaire fournissant la lysine à l'ancêtre MEAM-MED était bien l'ancêtre d'*H. defensa* MEAM1 et MED-Q1, quatre événements de remplacements par *Arsenophonus* sp. auraient eu lieu (**Figure 6**) : chez *B. tabaci* MED-Q2, chez MED-Q3, ainsi que deux événements distincts chez *B. tabaci* MED-ASL – ce cytotype est infecté par deux souches distinctes d'*Arsenophonus* sp (Mouton *et al.*, 2012), sur lesquelles nous revenons ci-après.

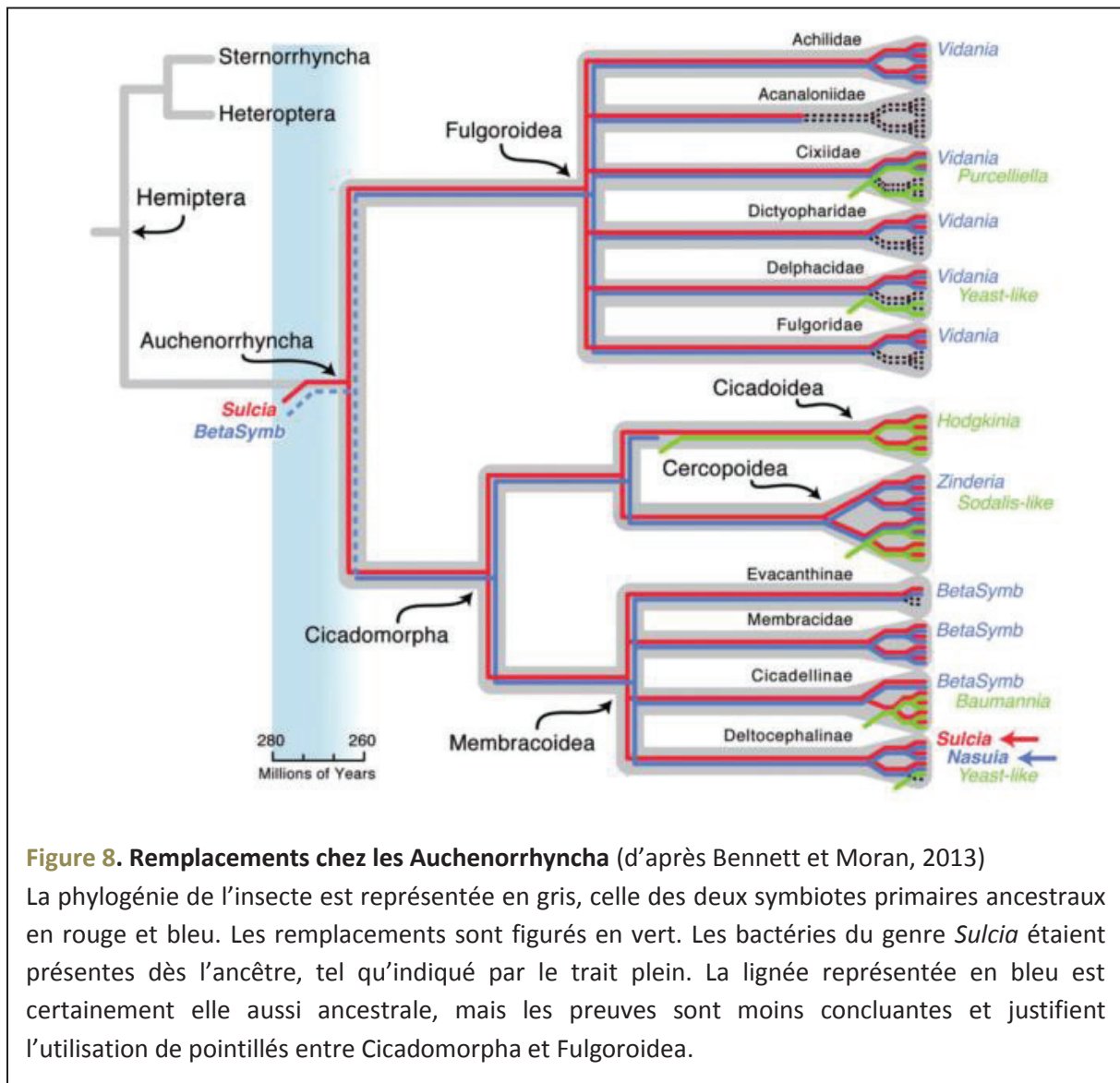


Si le symbiote ancestral était *Arsenophonus* sp., les choses auraient été plus complexes. Tout d'abord, trois remplacements par *H. defensa* au moins auraient eu lieu, chez *B. tabaci* MEAM1, MED-Q1, et MED-ASL (**Figure 7**). De plus, deux souches d'*Arsenophonus* sp. existent dans le complexe *B. tabaci*, et qui proviennent de deux clades différents : un clade uniquement retrouvé chez les Hémiptères, et un clade également retrouvé chez les Diptères et les Hyménoptères (dit « Cosmopolite » sur les **Figures 6 et 7**), dont la présence chez *B. tabaci* se cantonne à l'Afrique et pourrait être récente (Mouton *et al.*, 2012). Ainsi, les bactéries de ce clade ne pourraient pas être la bactérie ancestrale recherchée. À l'inverse, une bactérie du clade inféodé aux Hémiptères pourrait être le néo-primaire chez *B. tabaci* MED-MEAM, et avoir été remplacé par ce clade « africain » dans la sous-espèce MED-Q3 et certains individus de *B. tabaci* MED-ASL (**Figure 7**).



### b) Les remplacements chez les Hémiptères

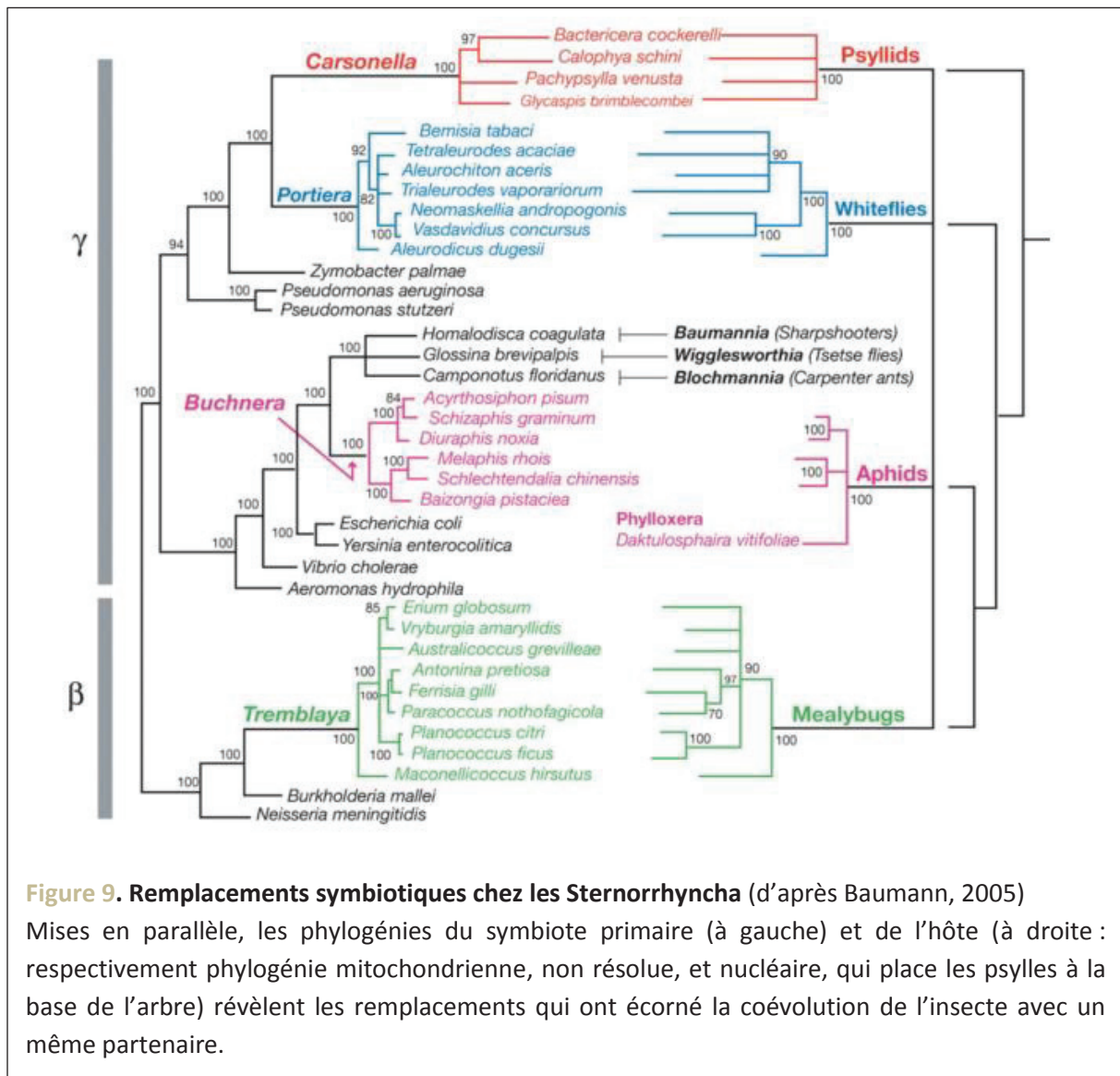
Ce scénario d'un endosymbiote néo-primaire ancestral, suivi de remplacements, n'a absolument rien de saugrenu si l'on se réfère aux connaissances actuelles. L'on retrouve en effet de nombreux remplacements chez les Hémiptères, notamment lorsque l'on a deux symbiotes primaires infectant le même hôte comme c'est le cas chez les Auchenorrhyncha (**Figure 8**). Deux endosymbiotes primaires seraient présents depuis le dernier ancêtre de ce Sous-Ordre : une bactérie de la classe des Bacteroidetes, appartenant au genre *Sulcia*, et une Protéobactérie Béta, l'ancêtre potentiel des genres *Vidania*, *Zinderia* et *Nasuia*.



Pourtant, chez certains Cercopoidea et Cicadellinae, ainsi que chez les Cicadoidea, la protéobactérie Béta a été remplacé respectivement, et indépendamment, par deux Protéobactéries Gamma, l'une apparentée au genre *Sodalis*, l'autre appartenant à l'espèce *Baumannia cicadellincola*, et par une Protéobactérie Alpha, *Hodgkinia cicadicola*. Chez certains Deltocephalinae, l'endosymbiote *Sulcia* a laissé sa place à un symbiote eucaryote. Chez les Delphacidae et certains Cixiida, c'est l'ensemble du consortium ancestral qui a été remplacé par un « nouveau » symbiote.

De la même façon, chacune des quatre Superfamilles du Sous-Ordre des Sternorrhyncha a établi une association avec un endosymbiote différent – seuls les endosymbiotes des psylles et des aleurodes pourraient être apparentés (**Figure 9**). Cette

présence de différents clades de bactéries ne peut s'expliquer que par des remplacements successifs des symbiotes primaires.



Les mécanismes sous-jacents à ces remplacements demeurent bien mal compris. La première condition serait bien évidemment que de nouveaux endosymbiotes arrivent par un transfert horizontal. Ainsi, les clades d'insectes infectés par de nombreuses espèces de bactéries endosymbiotiques seraient plus enclins à pareils remplacements (Moran et Baumann, 1994). Tel est le cas de *B. tabaci*, qui est l'hôte d'au moins une demi-douzaine d'endosymbiotes, et est donc un clade favorable à l'apparition de ce phénomène.



Certains modèles prédisent que la perte massive de gènes chez les symbiotes pourrait expliquer leur remplacement par d'autres endosymbiotes plus « jeunes » : à force de fixer des mutations délétères, ils deviendraient inefficaces pour leur hôte (Koga et al., 2003). Cependant, il existe une sélection au niveau inter-hôte : les insectes, et notamment les phloémophages, vivent dans des populations de grandes tailles et sont donc soumis à sélection, de sorte que toute mutation du symbiote qui affecte la valeur adaptative de son hôte sera contre-sélectionnée. Notre mitochondrie est un bon exemple de « vieil endosymbiote », elle qui assure la respiration des eucaryotes depuis leur apparition sur Terre il y a deux milliards d'années (Margulis, 1981). L'association entre cette organelle et son hôte est toutefois particulièrement « intime », de nombreux gènes présents ancestralement dans le génome de la mitochondrie ayant « migré » dans le génome nucléaire de l'eucaryote.

Enfin, les remplacements pourraient être corrélés avec des transitions écologiques chez l'hôte (Conord *et al.*, 2008). Chez les Auchenorrhyncha, l'infection de *B. cicadellinicola* coïncide avec un changement de régime alimentaire, les insectes ne se nourrissant non plus de phloème mais de xylème. Nous avons également évoqué la modulation du spectre de plantes hôtes induite par la présence de *R. insecticola* (Tsuchida *et al.*, 2004). Des travaux récents suggèrent que les endosymbiotes pourraient, de par leur interaction avec les traits d'histoire de vie de leurs hôtes respectifs, avoir un impact important sur le partitionnement de niche de ces derniers, participant ainsi à leur isolement reproductif (voir par exemple le travail de thèse de Merville, 2013). Toutefois, de nombreux remplacements ne sont pas associés – en apparence – à de telles transitions, ne serait-ce que ceux ayant eu lieu chez les quatre Superfamilles de Sternorrhyncha, toutes majoritairement phloémophages.

Les espèces des clades MED et MEAM1 se caractérisent également par un régime alimentaire phloémophage. Elles diffèrent toutefois dans leur spectre d'hôtes. Par exemple, le cytotype MED-Q3 ne semble pas polyphage et a pour symbiote une souche particulière d'*Arsenophonus* sp. que nous avons désignée par le terme « cosmopolite » sur les **Figures 6** et **7**. Les espèces polyphages sont infectées quant à elles par une autre souche d'*Arsenophonus* ou par *H. defensa*. De plus, les espèces MED et MEAM1 présentent de nombreuses différences en termes de résistance aux insecticides, de fécondité, ou bien de

bénéfices retirés de la présence des virus (Horowitz *et al.*, 2005 ; Luo *et al.*, 2010 ; Pascal et Callejas, 2004 ; Fang *et al.*, 2013). Les cortèges endosymbiotiques jouent-ils un rôle dans cette différenciation entre les espèces de *B. tabaci*, et dans le caractère polyphage ou non de ces insectes ?

## IV. Conséquences de la fixation d'un endosymbiote néo-primaire

### 1. Phénotype étendu : les holobiontes *Bemisia tabaci* MED et MEAM1

De manière intéressante, si les aleurodes des espèces MED et MEAM1 présentent des différences dans leurs traits d'histoire de vie, elles ont surtout pour particularité commune d'être les plus invasives parmi le complexe d'espèces *B. tabaci* (voir « **Modèle biologique** »). Originaires du bassin Méditerranée (MED) et du Moyen-Orient (MEAM1), on les retrouve désormais sur tous les continents. De plus, leur invasion est concomitante avec celle de différents virus phytopathogènes, et notamment le virus TYLC (Tomato Yellow Leaf Curl) avec qui elles entretiennent des relations ambiguës : coûteux pour l'insecte (Rubinstein et Czosnek, 1997), manipulateur de son comportement (Jahan *et al.*, 2014), le virus TYLC pourrait réduire à néant les défenses de la plante et permettrait la propagation de l'insecte en même temps que la sienne (Shi *et al.*, 2014). Elles sont enfin extrêmement polyphages (Guo *et al.*, 2012), toutes ces caractéristiques leur permettant d'exclure les espèces locales (Delatte *et al.*, 2009). Il est intéressant que cette forte capacité invasive soit corrélée avec la présence de symbiotes secondaires à très forte prévalence tels qu'*H. defensa* et *Arsenophonus* sp, phénomène que ne connaissent pas les espèces non invasives. Bien sûr, cette corrélation ne signifie pas pour autant que la présence de ces symbiotes soit la clé de leur succès. Penchons-nous toutefois sur cette hypothèse.

Si nous ne savons rien d'*Arsenophonus* sp., la génomique a révélé qu'*H. defensa* synthétiserait des acides aminés non essentiels, mais surtout différents cofacteurs. Il est possible que leur synthèse élargisse le spectre de plantes cibles pour l'insecte, et soit en lien avec la polyphagie des espèces MED et MEAM1. De plus, il a été démontré qu'*H. defensa* favorisait l'acquisition, la rétention et la transmission de certains virus, tels que le susnommé virus TYLC (Su *et al.*, 2013). Il est donc possible que la présence d'*H. defensa* favorise le mutualisme entre les virus et son hôte, et donc la dispersion de celui-ci (Su *et al.*, 2013). En tous les cas, en interférant avec la relation entre son hôte et les virus, il est évident que la présence d'*H. defensa* aura des conséquences sur les traits d'histoire de vie de l'insecte.

## 2. Phénotype d'*Hamiltonella defensa* chez *Bemisia tabaci*

À l'inverse, la « capture » de ces endosymbiotes secondaires n'est pas sans conséquence sur leur génome, et donc sur leur phénotype. Chez le puceron, nous savons que le genre *Hamiltonella* se caractérise par une propension à la protection contre les ennemis naturels. Ce phénotype probablement ancestral est condamné à disparaître dans le clade hBtab – si ce n'est pas déjà le cas – sous l'hypothèse d'une bactérie néo-primaire pour *B. tabaci* MED-MEAM1. Nous allons tout d'abord expliquer pourquoi une bactérie obligatoire ne peut probablement pas assurer un rôle protecteur efficace.

La dynamique des populations d'endosymbiotes protecteurs et celles de l'ennemi naturel – mettons, un parasitoïde – n'est pas sans rappeler les dynamiques proies-prédateurs de Lotka-Volterra (Takeuchi, 1996). Lorsque la pression parasitaire est élevée, l'infection par la bactérie protectrice devient très avantageuse, et la proportion d'individus infectés augmente. En conséquence de la pression exercée par les symbiotes, la population de parasitoïdes va décroître en même temps que la pression qu'ils induisent, et les parasitoïdes résistants seront sélectionnés. Cela rend moins important le bénéfice conféré par la bactérie, et avantage au contraire les hôtes qui s'affranchissent de la présence du symbiote et de son coût : la population d'individus infectés va diminuer.

Prenons désormais le cas d'une bactérie protectrice, qui devient obligatoire pour son hôte selon le scénario évoqué en amont dans cette **Discussion**. La fixation brutale de ce symbiote va logiquement s'accompagner d'une pression énorme sur les ennemis naturels contre lesquels elle protège. Là encore, la population des parasitoïdes et la pression qu'ils exercent vont diminuer de pair. La protection engendrée par la bactérie va donc devenir moins avantageuse, sans qu'il y ait pour autant de réponse du taux d'infection, du fait de son rôle essentiel.

Il n'y a donc plus désormais de variabilité dans la proportion d'individus infectés en réponse à la pression parasitaire. Cela empêche-t-il la bactérie de protéger son hôte ? Dès lors qu'il y a fixation du symbiote, cela va « forcer » l'interaction entre le parasitoïde et celui-ci, et donc la coévolution entre les deux. Nous arrivons ainsi dans une logique typique de course aux armements : la variabilité dans la population de parasitoïdes va entraîner une

sélection des plus virulents, et en réponse, les holobiontes proposant la meilleure défense pourront être sélectionnés, et ainsi de suite.

Deux composantes peuvent servir de base à cette protection : le compartiment cytoplasmique et le fond génétique de l'hôte. Ces deux composantes se distinguent par leurs modes de reproduction, le premier se répliquant de manière clonale, le second possédant une reproduction sexuée qui va entraîner un brassage génétique à chaque génération. Or, de par les combinaisons nouvelles qu'elle va générer, cette reproduction sexuée est plus efficace que les mutations ponctuelles dans le cadre de la course aux armements (Muller, 1932).

En conclusion, l'on peut s'attendre à ce que ce soit surtout par l'intermédiaire du fond génétique eucaryote que la course aux armements se fasse chez les aleurodes MED et MEAM1 infectés par *H. defensa*.

## V. Niveaux de sélection chez les endosymbiotes

Nous avons donc chez *B. tabaci* un endosymbiote présentant une trajectoire évolutive totalement différente de sa cousine chez le puceron. De plus, leurs écologies dans leur hôte respectif sont totalement distinctes (**Article Second**). Nous allons nous intéresser au lien qui existe entre cette écologie et la rétention de gènes chez les endosymbiotes, et pour cela faire appel aux différents niveaux de sélection qui vont s'exercer sur leurs génomes.

### 1. Rétention de gènes et modalités de la relation mutualiste

Dans notre second article, nous nous sommes ainsi intéressés aux différences entre les souches d'*H. defensa* en fonction de leurs particularités écologiques (**Article Second**). Nous avons pour cela utilisé deux clades de bactéries appartenant à la même espèce *H. defensa*, mais très différentes dans leur contribution pour l'hôte : une bactérie protectrice, qui semble être un symbiote secondaire typique (tropisme tissulaire important, prévalence intermédiaire, capacité à transférer horizontalement), et un symbiote potentiellement nutritif et primaire (séquestration dans les bactériocytes, fixation dans la population).

Ces nuances très marquées dans leur écologie concordent avec les gènes ancestraux différenciellement retenus dans les deux clades. D'un côté, hApis\_5AT, symbiote du puceron, conserve de nombreux facteurs virulents classiques des bactéries pathogènes du genre *Yersinia* auquel *H. defensa* s'apparente (système de sécrétion de type 3, antigène O etc.). Elle est également capable de réagir aux variations environnementales du fait de la rétention de systèmes de perception. Elle dispose en outre de deux partenaires symbiotiques impliqués dans l'acquisition de nouveaux gènes – un plasmide et le bactériophage APSE – et qui interviendraient donc dans la course aux armements engagée contre les ennemis naturels. D'un autre côté, notre clade d'*H. defensa* chez *B. tabaci* se caractérise par la perte de ces systèmes associés à une forte virulence, ainsi que de différents systèmes de régulation. De plus, l'acquisition de nouveaux gènes semble limitée chez ces bactéries, en partie en raison de la perte des partenaires symbiotiques ancestraux.

Il est difficile de se faire une idée sur le rôle précis de ces systèmes de virulence : peut-être ont-ils une implication directe dans le phénotype bactérien (protection, communication avec l'hôte, transmission verticale), peut-être sont-ils impliqués dans le transfert horizontal. Peut-être assurent-ils pour partie les deux fonctions, notamment les facteurs de virulence de hApis\_5AT qui lui permettent probablement à la fois de protéger son hôte, et de transférer d'un insecte à un autre. Nous allons nous pencher davantage sur ces facteurs de virulence et autres gènes impliqués dans le transfert horizontal, sur leur maintien chez les symbiotes facultatifs, et sur leur propension à être perdus ou non selon le phénotype de la bactérie.

## 2. Sélection de clades endosymbiotiques

### a) Généralités

Les endosymbiotes, primaire et secondaires, ont une transmission majoritairement verticale. Durant ces phases, la capacité à transférer horizontalement n'est pas soumise à sélection : un endosymbiote peut la perdre tant qu'il ne sont affectés, ni son phénotype dans l'hôte (mutualisme et/ou manipulation), ni sa capacité à se transmettre à la descendance de l'insecte. Pire, la transmission horizontale reposant sur des systèmes virulents tels que ceux retrouvés dans les génomes d'*H. defensa*, coûteux pour l'hôte, leur perte peut être **sélectionnée** au niveau inter-hôtes, à l'échelle populationnelle. En conséquence, lorsque les symbiotes secondaires sont transmis **verticalement**, ils peuvent rapidement perdre leur capacité à transférer **horizontalement**. Ce faisant, ils sont condamnés à disparaître sur le long terme, du fait de leur caractère facultatif pour l'hôte.

Cette capacité à transférer horizontalement agit donc comme un véritable filtre : seuls les symbiotes qui ne l'ont pas perdue vont être retrouvés sur le long terme. La conservation des systèmes permettant le transfert horizontal dans les génomes des symbiotes facultatifs fait ainsi intervenir un autre niveau de sélection, que l'on appelle « **sélection de clades** ». Cette notion de sélection de clade est surtout utilisée pour expliquer le maintien d'une stratégie particulière que nous avons déjà évoquée, l'incompatibilité cytoplasmique (IC). Cette stratégie efficace peut conduire sous certaines conditions – qui ne nous intéresseront pas ici – à une invasion de la population hôte par le symbiote. Les modèles prédisent toutefois qu'une fois le symbiote fixé, là encore sous certaines conditions

que nous passerons sous silence, les individus qui vont perdre ce symbiote vont être sélectionnés, amenant à une disparition de celui-ci dans la population (Hurst et Mcvean, 1996). Là encore, le maintien de cette capacité rapidement contre-sélectionnée au niveau populationnel ne s'explique que par le filtre de la transmission horizontale.

Ce filtre de la transmission horizontale n'existe pas chez les endosymbiotes à transmission verticale stricte, condamnés par conséquent à voir s'éroder leur contenu génique.

### **b) Conditions pour que la sélection de clade opère**

La vitesse à laquelle les gènes vont être perdus va jouer un rôle important dans la possibilité ou non pour la sélection de clade de s'effectuer. S'ils sont trop virulents, ils vont être rapidement éliminés, de sorte que nul clade ne puisse être sélectionné sur la base de leur conservation. La sélection de clade va donc opérer uniquement si le coût physiologique que ces systèmes vont infliger à l'insecte hôte n'est pas trop important. Une autre possibilité est que le bénéfice retiré par la présence de ces gènes vienne contrebalancer leur virulence pour l'hôte.

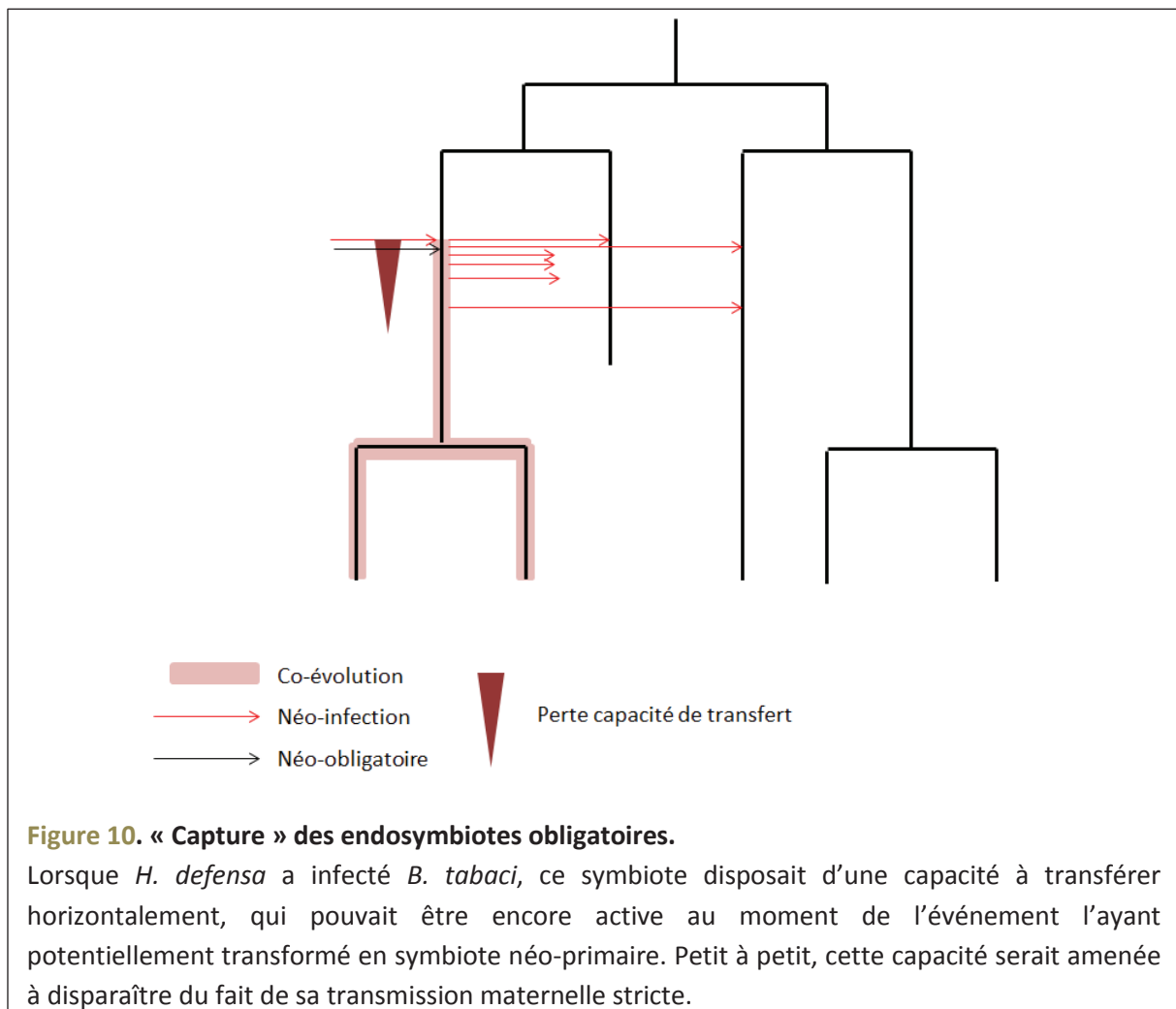
Outre leur rôle dans le transfert horizontal, nous – mais également d'autres avant nous (Degnan *et al.*, 2009 ; Degnan *et al.*, 2010) – avons émis l'hypothèse selon laquelle ces systèmes virulents pouvaient être en lien direct avec le phénotype de protection d'*H. defensa* chez le puceron. Ils lui permettraient tout d'abord de tuer l'ennemi en développement dans l'insecte. En outre, l'infection des différents tissus et fluides de l'hôte, en particulier l'hémolymphe dans laquelle sont pondus les œufs des guêpes parasitoïdes, se baserait sur la capacité du symbiote à contourner les défenses de son hôte (Hansen *et al.*, 2012 ; Dykstra *et al.*, 2014). Là encore, de nombreux gènes seraient impliqués (facteurs de virulence, senseurs, gènes du métabolisme, transporteurs etc.). De manière intéressante, le phénotype de la bactérie aurait ainsi une grande influence sur sa propension à perdre sa faculté à transférer horizontalement : si les systèmes permettant le transfert horizontal sont dans leur grande majorité ceux impliqués dans le phénotype de la bactérie, ils seront préservés de la perte.



### c) « Capture » d'un endosymbiote secondaire

À l'inverse, chez *B. tabaci*, le phénotype apparemment nutritif, combiné à sa séquestration dans les bactériocytes, a probablement conduit à un relâchement des pressions de sélection sur ces facteurs de virulence. C'est pourquoi de nombreux systèmes parmi les plus virulents ont pu être perdus, tels le système de sécrétion de type III, ou encore l'antigène O.

Nous aurions probablement une perte progressive de la capacité de transfert chez ces bactéries. Au moment où *H. defensa* a acquis un statut obligatoire, peut-être possédait-elle encore cette capacité. Aussi est-il envisageable que certaines bactéries aient pu transférer depuis *B. tabaci* vers d'autres hôtes (**Figure 10**). D'ailleurs, sous l'hypothèse d'un ancêtre *B. tabaci* MED-MEAM infecté par une bactérie autre qu'*H. defensa*, la situation actuelle s'expliquerait par trois infections indépendantes dans *B. tabaci* MED, MEAM1 et ASL, et en toute logique par des bactéries capables de transférer horizontalement.



Cependant, pendant que certaines bactéries ont potentiellement pu quitter *B. tabaci*, d'autres se sont retrouvées « capturées » par l'holobionte : de transmission maternelle en transmission maternelle, ces symbiotes vont progressivement perdre, ou ont perdu, leur faculté à transférer horizontalement. La perte de la capacité à synthétiser le chorismate par *H. defensa* MED-Q1 s'insère parfaitement dans cette logique, puisqu'elle rend la bactérie dépendante à un environnement où ce nutriment est présent en quantité suffisante.

#### **d) Dernier ancêtre commun**

Si nous comprenons un peu mieux les forces qui s'exercent sur les *H. defensa* actuelles, en raison de leurs écologies respectives, que peut-on déduire du phénotype de l'ancêtre d'après leurs génomes ?

La génomique comparative nous enseigne que le dernier ancêtre commun à ce genre possédait de nombreux facteurs de virulence. Cet arsenal peut surprendre pour un symbiote de longue date, que la transmission verticale aurait dû rapidement « assagir », ainsi que nous l'avons déjà mentionné. Une première explication fait intervenir directement la relation mutualiste entre cette bactérie et son hôte. Au regard des phénotypes de ses descendants et des bactéries du genre *Regiella* apparentées, l'ancêtre du genre *Hamiltonella* faisait certainement bénéficier un antique Hémiptère phloémophage de sa protection. Il est possible que le spectre d'ennemis de son hôte fût plus important que celui des pucerons et aleurodes. En ce cas, les symbiotes lourdement armés auraient avantagé ce dernier, ce qui aurait directement contribué à la sauvegarde des nombreux facteurs de virulence de la bactérie.

La sélection de clade pourrait être une seconde explication, non exclusive de la première. Du fait de son arsenal, sans doute les capacités de cet ancêtre à transférer horizontalement étaient-elles bien supérieures à celles de ses descendants actuels – en tous les cas ceux dont le génome a été séquencé. Voire, cette transmission pouvait être facilitée par une capacité à survivre hors l'insecte, ce qui pourrait – j'insiste sur le caractère hypothétique de ce qui va suivre – être encore le cas de ses descendants chez le puceron, retrouvé hors de l'insecte (Darby et Douglas, 2003). En contrepartie de cette aptitude au transfert latéral, cette bactérie devait être plus coûteuse pour son hôte. Nous aurions affaire

à un endosymbiote ancestral plus « généraliste », que ce soit en termes d'hôtes à infecter, et/ou d'ennemis contre lesquels défendre ceux-ci. Cela vient compléter l'idée selon laquelle la sélection de clade nécessite que ne soient pas trop rapidement perdus les gènes permettant le transfert horizontal : si la bactérie possède une importante capacité à transférer latéralement, du fait de systèmes de virulence efficace, cela pourrait compenser le coût associé à la conservation de ces systèmes.

Le génome de ce symbiote ancestral, ainsi que celui de ses descendants, révèlent ainsi les différentes forces qui s'exercent chez ces bactéries facultatives. La sélection au niveau de l'holobionte va favoriser l'**avirulence pour l'hôte** (mais pas pour les ennemis de celui-ci chez les symbiotes protecteurs), tandis que la sélection de clades favorisera davantage les bactéries possédant une bonne aptitude à infecter de nouveaux hôtes, et donc potentiellement coûteuses pour leur hôte. La différence majeure entre ces deux forces réside dans l'échelle qui les sépare : l'une agissant à chaque génération, l'autre sélectionnant des clades plutôt que d'autres, agissant donc sur des centaines, milliers, voire millions d'années.

# PERSPECTIVES

---

À travers cette discussion, nous nous sommes intéressés à l'évolution génomique de l'espèce *H. defensa*, et des endosymbiotes facultatifs en général. Nous avons vu que de nombreuses interrogations demeuraient, aussi bien d'un point de vue fonctionnel que conceptuel. Outre l'**Article Troisième**, pour lequel nous avons déjà évoqué les questions qui demeuraient, nous allons revenir sur différentes approches sur lesquelles nous pourrions nous diriger afin de compléter ce travail de thèse.

## I. Interactions métaboliques chez *Bemisia tabaci*

### 1. Limites de la génomique

Il convient d'abord de préciser que la génomique permet de formuler des hypothèses, non pas des certitudes. Ainsi, concernant notre symbiose chez *Bemisia tabaci*, plusieurs expérimentations seraient imaginables afin de confirmer nos premières hypothèses. En premier lieu, une élimination ciblée d'*H. defensa* devrait permettre de démontrer le caractère obligatoire de cette bactérie. Ensuite, l'utilisation de milieux artificiels, de régimes alimentaires riches en phénylalanine et lysine devrait engendrer une restauration de la valeur adaptative des insectes dépourvus d'*H. defensa*, et ainsi confirmer le rôle de celle-ci dans le métabolisme de ces acides aminés. Des études de métabolomique en présence et en absence du symbiote seraient extrêmement intéressantes. Si ces expériences paraissent simples à mener de prime abord, l'élimination du symbiote s'est avérée délicate, et n'a pas encore été réalisée avec succès d'après mes connaissances.

De la même façon, reconstruire le pangénome de l'holobionte est un abus très pratique, mais qui fait fi de la nature plurielle de celui-ci. Rien n'indique que les échanges proposés *in silico* aient réellement lieu. Il est par exemple classique que les nutriments proposés dans les échanges soient les composés « finaux » des voies métaboliques. Rien n'indique que des composés intermédiaires n'en soient pas en réalité les acteurs. Ensuite, il faut que des transporteurs puissent assurer le transit des métabolites entre les partenaires. S'il semble plausible qu'*H. defensa* puisse acquérir un grand nombre de nutriments depuis son environnement, nous avons vu que le symbiote primaire possédait de son côté un

nombre limité de transporteurs. Il est possible que l'hôte remédie à cette situation en fournissant à son partenaire les transporteurs requis. Il serait très intéressant de mettre en évidence la véritable communication qui existe dans cet holobionte, tant au niveau des nutriments échangés qu'au niveau de la régulation de ces échanges.

Le dialogue entre l'hôte et *H. defensa* ne concerne d'ailleurs sans doute pas que les métabolites, et doit s'installer dès les premiers stades larvaires afin de favoriser l'établissement du symbiote dans la cellule. Comparer les processus assurant la transmission verticale d'*H. defensa*, respectivement, chez le puceron et chez *B. tabaci*, permettrait de comprendre l'évolution de cette transmission chez un symbiote en transition vers un mode de vie obligatoire.

## **2. Symbiotes néo-primaires**

Nous avons mis en évidence un rôle potentiel majeur pour *H. defensa* dans la synthèse de la lysine. Nous avons émis l'hypothèse que ce rôle était ancestralement dévolu à *P. aleyrodidarum*. Il serait intéressant de séquencer les génomes des symbiotes primaires pour les espèces apparentées aux espèces MED et MEAM (au moins les gènes concernés par la synthèse de cet acide aminé), ainsi que pour les différents cytotypes de MED (et notamment ASL), afin de mieux connaître l'historique de la perte de capacité à synthétiser la lysine.

Ensuite, nous aimerions tester l'hypothèse selon laquelle *Arsenophonus* sp. synthétiserait cette molécule dans les populations qu'elle infecte. Nous n'avons malheureusement pas pu séquencer le génome de cette bactérie, qui pourrait être très fragile ou très peu représentée dans la cellule comparativement aux autres symbiotes (Diego Santos-Garcia, communication personnelle). L'obtention de ce génome permettrait d'inférer les capacités métaboliques d'*Arsenophonus* sp., et le dialogue qu'il noue avec son symbiote primaire. Comparer les deux symbiotes potentiellement néo-primaires *Arsenophonus* sp. et *H. defensa*, ainsi que les différentes souches d'*Arsenophonus* sp. serait très intéressant, et permettrait peut-être de comprendre pourquoi des remplacements symbiotiques réguliers ont eu lieu dans le clade *B. tabaci* MED/MEAM.

### 3. Tri infections

Nous avons étudié les relations métaboliques s'exerçant entre deux symbiotes fixés dans les populations chinoises de *B. tabaci* MED-Q1, *P. aleyrodidarum* et *H. defensa*. On retrouve les deux mêmes symbiotes chez *B. tabaci* MEAM1, ainsi qu'un troisième partenaire aux prévalences intermédiaires : *Rickettsia* sp.

Nous avons envisagé d'étudier le génome de cette bactérie apparentée aux pathogènes *R. bellii*, en collaboration avec l'équipe chinoise du professeur Liu. Nous disposons du génome non complet de deux symbiotes du genre *Rickettsia* de *B. tabaci* MEAM1 : une bactérie infectant des populations chinoises, d'autres des populations israéliennes. Il n'est pas impossible que l'on dispose bientôt du génome d'un symbiote de populations américaines. Le comparatif des trois symbiotes serait intéressant, puisqu'il est possible qu'ils n'aient pas tout à fait les mêmes associations avec leurs hôtes (Caspi-Fluger *et al.*, 2011 ; Himler *et al.*, 2011 ; Kliot *et al.*, 2014). Le dialogue métabolique qu'ils pourraient établir avec les autres partenaires, ainsi que l'évolution de ces bactéries apparemment mutualistes apparentées aux pathogènes des genres *Rickettsia*, font également partie de nos centres d'intérêt. Le génome de *Wolbachia* sp. chez l'aleurode nous intéresse pour cette même raison, cette bactérie ne présentant pas *a priori* de phénotype manipulateur.

## II. *Hamiltonella defensa* et sélection de clades

Éloignons-nous quelque peu du modèle *Bemisia tabaci* et revenons à l'espèce modèle, *H. defensa*. La notion de sélection de clades est extrêmement intéressante chez les symbiotes facultatifs. Pour l'appréhender, la modélisation est un outil important. L'évolution à court et long termes de bactéries protectrices ou nutritives, leur tendance à l'aviorulence lors de la transmission maternelle, le filtre de la transmission horizontale, sont autant de phénomènes qu'il serait intéressant de modéliser. La sélection de clade peut-elle réellement compenser le déclin *a priori* inéluctable des endosymbiotes facultatifs ?

La génomique peut également apporter sa pierre à l'édifice, et renseigner sur les taux de pertes aux nœuds et aux branches dans les arbres de phylogénie. Par exemple, il existe une diversité énorme de souches chez *H. defensa* chez le puceron. Étudier l'évolution de la virulence chez ces bactéries serait intéressant. Tout d'abord, certaines souches semblent coévoluer avec le genre de pucerons *Uroleucon* : l'évolution de leur génome suit-elle la tendance des génomes de symbiotes de *Bemisia tabaci* ? D'autres souches se retrouveraient dans des hôtes très éloignés. Sont-ce celles ayant conservé davantage de facteurs de virulence ? Existe-t-il une corrélation entre le spectre d'hôtes et la rétention de facteurs de virulence ? Certaines souches ont-elles conservé la quasi-intégralité du lot ancestral de facteurs de virulence, ou bien toutes semblent-elles condamnées à voir leur génome s'éroder malgré la sélection des clades les moins impactés ? Là encore, modéliser cette perte de gènes, qui doit mener vers un amenuisement des capacités à transférer latéralement, serait très intéressant pour comprendre l'évolution de ces symbiotes sur le très long terme.

Pour finir, des pertes de gènes différentes peuvent conduire à des lignées symbiotiques très distinctes à partir d'un même ancêtre commun. L'on peut ajouter à cela les transferts horizontaux de matériel génétique (par exemple, les composants de la paroi cellulaire chez *H. defensa*) qui contribuent à différencier davantage les symbiotes les uns des autres. La sélection peut-elle par la suite opérer pour maintenir cette diversité ? Deux lignées symbiotiques distinctes doivent conférer des bénéfices très différents selon les conditions environnementales. Par exemple, tel facteur de virulence pourrait être plus efficace contre tel ennemi naturel, de sorte qu'en fonction des ennemis présents, les

différentes souches ne conféreraient pas la même intensité dans la protection. Il est possible qu'un phénomène de sélection fréquence-dépendante s'ajoute à cela. Si une souche donnée présente des prévalences importantes dans une population, les parasitoïdes vont être sélectionnés pour contourner la protection qu'elle confère. Il est probable qu'un second symbiote qui viendrait infecter la population et serait inconnu du parasitoïde exclue rapidement le symbiote avec lequel aura co-évolué l'ennemi naturel.





# CONCLUSION

---

Mon travail de thèse a combiné analyse génomique et génomique comparative, pour s'intéresser à la fois aux interactions métaboliques entre symbiotes et au scénario ayant conduit aux observations actuelles. Ces outils se sont également révélés très intéressants pour comprendre les forces agissant sur les symbiotes secondaires selon leur écologie et leur phénotype pour l'hôte. Ces études ont permis de surcroît de formuler de nombreuses hypothèses qui ne demandent qu'à être confirmées par d'autres approches (modélisation, biologie moléculaire, expérimentations *in vitro* etc.).

Enfin, il apparaît que la symbiose chez les insectes est bien moins triviale que de premier abord. De nombreux niveaux de sélection agissent (intra-hôte, inter-holobionte, inter-clades) et se répercutent dans les génomes des symbiotes. Si la sélection inter-hôte favorise le mutualisme, nous avons vu que la sélection de clades favorisait en quelque sorte l'égoïsme des bactéries, qui vont être sélectionnées selon leur rétention de systèmes leur permettant de transférer d'hôte en hôte, bien que ces systèmes puissent être coûteux pour l'insecte.



## RÉFÉRENCES BIBLIOGRAPHIQUES

---

### A

Aksoy S. **1995**. *Wigglesworthia* gen. nov. and *Wigglessworthia glossinidia* sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. *International Journal of Systematic Bacteriology*, **45**, 848-851.

Allen J.M., Light J.E., Perotti M.A., Braig H.R. et Reed D.L. **2009**. Mutational meltdown in primary endosymbionts: selection limits Muller's ratchet. *PLoS One*, **4**, e4969.

Ammar E.D., Shatters Jr R.G., Lynch C. et Hall D.G. **2011**. Detection and relative titer of *Candidatus Liberibacter asiaticus* in the salivary glands and alimentary canal of *Diaphorina citri* (Hemiptera: Psyllidae) vector of citrus huanglongbing disease. *Annals of the Entomological Society of America*, **104**, 526-533.

Anbutsu H., Goto S. et Fukatsu T. **2008**. High and low temperatures differently affect infection density and vertical transmission of male-killing *Spiroplasma* symbionts in *Drosophila* hosts. *Applied and Environmental Microbiology*, **74**, 6053-6059.

Anderson R.M. et May R.M. **1982**. Coevolution of hosts and parasites. *Parasitology*, **85**, 411-426.

Anderson R.P. **2013**. A framework for using niche models to estimate impacts of climate change on species distributions. *Annals of the New York Academy of Sciences*, **1297**, 8-28.

Andersson S.G. et Kurland C.G. **1998**. Reductive evolution of resident genomes. *Trends in microbiology*, **6**, 263-268.

Andersson K., Carballeira N., Magnusson K.E., Persson C., Stendahl O., Wolf-Watz H. et al. **1996**. YopH of *Yersinia pseudotuberculosis* interrupts early phosphotyrosine signalling associated with phagocytosis. *Molecular Microbiology*, **20**, 1057-1069.

Arevalo-Ferro C., Hentzer M., Reil G., Görg A., Kjelleberg S., Givskov M. et al. **2003**. Identification of quorum-sensing regulated proteins in the opportunistic pathogen *Pseudomonas aeruginosa* by proteomics. *Environmental Microbiology*, **5**, 1350-1369.

Attardo G.M., Lohs C., Heddi A., Alam U.H., Yildirim S. et Aksoy S. **2008**. Analysis of milk gland structure and function in *Glossina morsitans*: Milk protein production, symbiont populations and fecundity. *Journal of Insect Physiology*, **54**, 1236-1242.

Augustinos A.A., Santos-Garcia D., Dionyssopoulou E., Moreira M., Papapanagiotou A., Scarvelakis M. et al. **2011**. Detection and characterization of *Wolbachia* infections in natural populations of aphids: is the hidden diversity fully unraveled? *PLoS One*, **6**, e28695.

## B

- Bäckhed F., Ley R.E., Sonnenburg J.L., Peterson D.A. et Gordon J.I. **2005**. Host-bacterial mutualism in the human intestine. *Science*, **307**, 1915-1920.
- Balmand S., Lohs C., Aksoy S. et Heddi A. **2013**. Tissue distribution and transmission routes for the tsetse fly endosymbionts. *Journal of Invertebrate Pathology*, **112**, S116-S122.
- Baumann P. **2005**. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annual Review of Microbiology*, **59**, 155-189.
- Baumann P., Moran N.A. et Baumann L. **2006**. Bacteriocyte-associated endosymbionts of insects. *The Prokaryotes: Volume 1: Symbiotic associations, Biotechnology, Applied Microbiology*, 403-438.
- Bengoechea J.A., Pinta E., Salminen T., Oertelt C., Holst O., Radziejewska-Lebrecht J. et al. **2002**. Functional characterization of Gne (UDP-N-acetylglucosamine-4-epimerase), Wzz (chain length determinant), and Wzy (O-antigen polymerase) of *Yersinia enterocolitica* serotype O: 8. *Journal of Bacteriology*, **184**, 4277-4287.
- Beninati T., Lo N., Sacchi L., Genchi C., Noda H. et Bandi C. **2004**. A novel  $\alpha$ -proteobacterium resides in the mitochondria of ovarian cells of the tick *Ixodes ricinus*. *Applied and Environmental Microbiology*, **70**, 2596-2602.
- Bennett G.M. et Moran N.A. **2013**. Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a Phloem-feeding insect. *Genome Biology and Evolution*, **5**, 1675-1688.
- Berkeley M.J. **1846**. Observations, botanical and physiological, on the potato murrain. *The Journal of the Horticultural Society of London*, **1**, 9-34.
- Bergstrom C.T. et Pritchard J. **1998**. Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. *Genetics*, **149**, 2135-2146.
- Bing X.L., Ruan Y.M., Rao Q., Wang X.W. et Liu S.S. **2013a**. Diversity of secondary endosymbionts among different putative species of the whitefly *Bemisia tabaci*. *Insect Science*, **20**, 194-206.
- Bing X.L., Yang J., Zchori-Fein E., Wang X.W. et Liu S.S. **2013b**. Characterization of a newly discovered symbiont of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Applied and Environmental Microbiology*, **79**, 569-575.
- Bing X.L., Xia W.Q., Gui J.D., Yan G.H., Wang X.W. et Liu S.S. **2014**. Diversity and evolution of the *Wolbachia* endosymbionts of *Bemisia* (Hemiptera: Aleyrodidae) whiteflies. *Ecology and Evolution*, **4**, 2714-2737.
- Blattner F.R., Plunkett G., Bloch C.A., Perna N.T., Burland V., Riley M. et al. **1997**. The complete genome sequence of *Escherichia coli* K-12. *Science*, **277**, 1453-1462.
- Bonfield J.K. et Whitwham A. **2010**. Gap5—editing the billion fragment sequence assembly. *Bioinformatics*, **26**, 1699-1703.
- Bordenstein S.R., O'Hara F.P. et Werren J.H. **2001**. *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. *Nature*, **409**, 707-710.

- Bordenstein S.R. et Wernegreen J.J. **2004**. Bacteriophage flux in endosymbionts (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. *Molecular Biology and Evolution*, **21**, 1981-1991.
- Bordenstein S.R. et Reznikoff W.S. **2005**. Mobile DNA in obligate intracellular bacteria. *Nature Reviews Microbiology*, **3**, 688-699.
- Borenstein E., Kupiec M., Feldman M.W. et Ruppin E. **2008**. Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proceedings of the National Academy of Sciences*, **105**, 14482-14487.
- Bossi L., Fuentes J.A., Mora G. et Figueroa-Bossi N. **2003**. Prophage contribution to bacterial population dynamics. *Journal of Bacteriology*, **185**, 6467-6471.
- Bouchon D., Cordaux R. et Grève P. **2008**. Feminizing *Wolbachia* and the evolution of sex determination in isopods. *Insect Symbiosis*, **3**, 273-294.
- Bourtzis K., Nirgianaki A., Markakis G. et Savakis C. **1996**. *Wolbachia* infection and cytoplasmic incompatibility in *Drosophila* species. *Genetics*, **144**, 1063-1073.
- Boykin L.M., Bell C.D., Evans G., Small I. et De Barro P.J. **2013**. Is agriculture driving the diversification of the *Bemisia tabaci* species complex (Hemiptera: Sternorrhyncha: Aleyrodidae)?: Dating, diversification and biogeographic evidence revealed. *BMC Evolutionary Biology*, **13**, 228.
- Boykin L.M. **2014**. *Bemisia tabaci* nomenclature: lessons learned. *Pest Management Science*, **70**, 1454-1459.
- Brady C.M. et White J.A. **2013**. Cowpea aphid (*Aphis craccivora*) associated with different host plants has different facultative endosymbionts. *Ecological Entomology*, **38**, 433-437.
- Brady C.M., Asplen M.K., Desneux N., Heimpel G.E., Hopper K.R., Linnen C.R. et al. **2014**. Worldwide populations of the aphid *Aphis craccivora* are infected with diverse facultative bacterial symbionts. *Microbial Ecology*, **67**, 195-204.
- Braendle C., Miura T., Bickel R., Shingleton A.W., Kambhampati S. et Stern D.L. **2003**. Developmental origin and evolution of bacteriocytes in the aphid–*Buchnera* symbiosis. *PLoS Biology*, **1**, e21.
- Bressan A. **2014**. Emergence and evolution of *Arsenophonus* bacteria as insect-vectorized plant pathogens. *Infection, Genetics and Evolution*, **22**, 81-90.
- Brodbeck B.V., Mizell III R.F. et Andersen P.C. **1993**. Physiological and behavioral adaptations of three species of leafhoppers in response to the dilute nutrient content of xylem fluid. *Journal of Insect Physiology*, **39**, 73-81.
- Brown J.K., Frohlich D.R. et Rosell R.C. **1995**. The sweetpotato or silverleaf whiteflies: biotypes of *Bemisia tabaci* or a species complex? *Annual Review of Entomology*, **40**, 511-534.
- Brown S.P., Le Chat L., De Paepe M. et Taddei F. **2006**. Ecology of microbial invasions: amplification allows virus carriers to invade more rapidly when rare. *Current Biology*, **16**, 2048-2052.

Brumin M., Kontsedalov S. et Ghanim M. **2011**. Rickettsia influences thermotolerance in the whitefly *Bemisia tabaci* B biotype. *Insect Science*, **18**, 57-66.

Brumin M., Levy M. et Ghanim M. **2012**. Transovarial transmission of *Rickettsia* spp. and organ-specific infection of the whitefly *Bemisia tabaci*. *Applied and Environmental Microbiology*, **78**, 5565-5574.

Bruneteau M. et Minka S. **2003**. Lipopolysaccharides of bacterial pathogens from the genus *Yersinia*: a mini-review. *Biochimie*, **85**, 145-152.

Buchner P. **1965**. *Endosymbiosis of animals with plant microorganisms*. New York: John Wiley.

Burke G.R. et Moran N.A. **2011**. Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biology and Evolution*, **3**, 195-208.

Byrne D.N. et Bellows Jr T.S. **1991**. Whitefly biology. *Annual Review of Entomology*, **36**, 431-457.

## C

Carvalho F.M., Souza R.C., Barcellos F.G., Hungria M. et Vasconcelos A.T. **2010**. Genomic and evolutionary comparisons of diazotrophic and pathogenic bacteria of the order Rhizobiales. *BMC Microbiology*, **10**, 37.

Caspi-Fluger A., Inbar M., Mozes-Daube N., Katzir N., Portnoy V., Belausov E. *et al.* **2011**. Horizontal transmission of the insect symbiont *Rickettsia* is plant-mediated. *Proceedings of the Royal Society B: Biological Sciences*, **rspb20112095**.

Caspi-Fluger A., Inbar M., Mozes-Daube N., Mouton L., Hunter M.S. et Zchori-Fein E. (2011). *Rickettsia* 'in' and 'out': two different localization patterns of a bacterial symbiont in the same insect species. *PloS one*, **6**, e21096.

Casteel C.L., Hansen A.K., Walling L.L. et Paine T.D. **2012**. Manipulation of plant defense responses by the tomato psyllid (*Bactericerca cockerelli*) and its associated endosymbiont *Candidatus Liberibacter psyllae*. *PloS One*, **7**, e35191.

Chain P.S., Carniel E., Larimer F.W., Lamerdin J., Stoutland P.O., Regala W.M. *et al.* **2004**. Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 13826-13831.

Charif D. et Lobry J.R. **2007**. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In *Structural approaches to sequence evolution* (pp. 207-232). Springer Berlin Heidelberg.

Charles H., Balmand S., Lamelas A., Cottret L., Pérez-Brocá V., Burdin B. *et al.* **2011**. A genomic reappraisal of symbiotic function in the aphid/*Buchnera* symbiosis: reduced transporter sets and variable membrane organisations. *PloS One*, **6**, e29096.

Chatterjee A.N. **1969**. Use of bacteriophage-resistant mutants to study the nature of the bacteriophage receptor site of *Staphylococcus aureus*. *Journal of Bacteriology*, **98**, 519-527.

- Chen D.Q. et Purcell A.H. **1997**. Occurrence and transmission of facultative endosymbionts in aphids. *Current Microbiology*, **34**, 220-225.
- Chen D.Q., Montllor C.B. et Purcell A.H. **2000**. Fitness effects of two facultative endosymbiotic bacteria on the pea aphid, *Acyrtosiphon pisum*, and the blue alfalfa aphid, *A. kondoi*. *Entomologia Experimentalis et Applicata*, **95**, 315-323.
- Cheng Q., Ruel T.D., Zhou W., Moloo S.K., Majiwa P., O'Neill S.L. *et al.* **2000**. Tissue distribution and prevalence of *Wolbachia* infections in tsetse flies, *Glossina* spp. *Medical and Veterinary Entomology*, **14**, 44-50.
- Chevreur B., Wetter T. et Suhai S. **1999**. Genome sequence assembly using trace signals and additional sequence information. In *German Conference on Bioinformatics*, 45-56.
- Chiel E., Gottlieb Y., Zchori-Fein E., Mozes-Daube N., Katzir N., Inbar M. *et al.* **2007**. Biotype-dependent secondary symbiont communities in sympatric populations of *Bemisia tabaci*. *Bulletin of Entomological Research*, **97**, 407-413.
- Chiel E., Zchori-Fein E., Inbar M., Gottlieb Y., Adachi-Hagimori T., Kelly S.E. *et al.* **2009**. Almost there: transmission routes of bacterial symbionts between trophic levels. *PLoS One*, **4**, e4767.
- Chung W.C., Chen L.L., Lo W.S., Kuo P.A., Tu J. *et al.* **2013**. Complete genome sequence of *Serratia marcescens* WW4. *Genome announcements*, **1**, e00126-13.
- Clark M.A., Baumann L., Munson M.A., Baumann P., Campbell B.C., Duffus J.E. *et al.* **1992**. The eubacterial endosymbionts of whiteflies (Homoptera: Aleyrodoidea) constitute a lineage distinct from the endosymbionts of aphids and mealybugs. *Current Microbiology*, **25**, 119-123.
- Clayton A.L., Oakeson K.F., Gutin M., Pontes A., Dunn D.M., von Niederhausern A.C. *et al.* **2012**. A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses. *PLoS Genetics*, **8**, e1002990.
- Colgan L.J. et Erbilgin N. **2011**. Tree-mediated interactions between the jack pine budworm and a mountain pine beetle fungal associate. *Ecological Entomology*, **36**, 425-434.
- Conway D.J., Fanello C., Lloyd J.M., Al-Joubori B., Baloch A.H., Somanath S.D. *et al.* **2000**. Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. *Molecular and Biochemical Parasitology*, **111**, 163-171.
- Coombs M.T., Costa H.S., de Barro P. et Rosell R.C. **2007**. Pre-imaginal egg maturation and bacteriocyte inclusion in *Bemisia* aff. *gigantea* (Hemiptera: Aleyrodidae). *Arthropod Biology*, **100**, 736-744.
- Cooper J.E. **2007**. Early interactions between legumes and rhizobia: disclosing complexity in a molecular dialogue. *Journal of Applied Microbiology*, **103**, 1355-1365.
- Cornelis G.R. **2002**. *Yersinia* type III secretion send in the effectors. *The Journal of Cell Biology*, **158**, 401-408.
- Corsaro D. et Greub G. **2006**. Pathogenic potential of novel Chlamydiae and diagnostic approaches to infections due to these obligate intracellular bacteria. *Clinical Microbiology Reviews*, **19**, 283-297.



Costa H.S., Westcot D.M., Ullman D.E. et Johnson M.W. **1993**. Ultrastructure of the endosymbionts of the whitefly, *Bemisia tabaci* and *Trialeurodes vaporariorum*. *Protoplasma*, **176**, 106-115.

Costa H.S., Toscano N.C. et Henneberry T.J. **1996**. Mycetocyte inclusion in the oocytes of *Bemisia argentifolii* (Homoptera: Aleyrodidae). *Annals of the Entomological Society of America*, **89**, 694-99.

Cottret L., Milreu P.V., Acuña V., Marchetti-Spaccamela A., Stougie L., Charles H. et al. **2010**. Graph-based analysis of the metabolic exchanges between two co-resident intracellular symbionts, *Baumannia cicadellinicola* and *Sulcia muelleri*, with their insect host, *Homalodisca coagulata*. *PLoS Computational Biology*, **6**, e1000904.

Cottret L., Wildridge D., Vinson F., Barrett M.P., Charles H., Sagot M.F. et al. **2010**. MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Research*, **38**, W132-W137.

Cryan J.R. **2005**. Molecular phylogeny of Cicadomorpha (Insecta: Hemiptera: Cicadoidea, Cercopoidea and Membracoidea): adding evidence to the controversy. *Systematic Entomology*, **30**, 563-574.

Cryan J.R. et Urban J.M. **2012**. Higher-level phylogeny of the insect order Hemiptera: is Auchenorrhyncha really paraphyletic? *Systematic Entomology*, **37**, 7-21.

Csardi G. et Nepusz T. **2012**. The igraph software package for complex network research. *International Journal of Complex Systems*.

Czosnek H. et Ghanim M. **2011**. *Bemisia tabaci*-tomato yellow leaf curl virus interaction causing worldwide epidemics. In *The whitefly, Bemisia tabaci* (Homoptera: Aleyrodidae) interaction with geminivirus-infected host plants, 51-67, Springer Netherlands.

## D

Dadd R.H. et Krieger D.L. **1968**. Dietary amino acid requirements of the aphid, *Myzus persicae*. *Journal of Insect Physiology*, **14**, 741-764.

Dale J.W., Brittain D., Cataldi A.A., Cousins D., Crawford J.T., Driscoll J. et al. **2001**. Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: recommendations for standardised nomenclature. *The International Journal of Tuberculosis and Lung Disease*, **5**, 216-219.

Dale C., Plague G.R., Wang B., Ochman H. et Moran N.A. **2002**. Type III secretion systems and the evolution of mutualistic endosymbiosis. *Proceedings of the National Academy of Sciences*, **99**, 12397-12402.

Dale C., Jones T. et Pontes M. **2005**. Degenerative evolution and functional diversification of type-III secretion systems in the insect endosymbiont *Sodalis glossinidius*. *Molecular Biology and Evolution*, **22**, 758-766.

Dale C. et Moran N.A. **2006**. Molecular interactions between bacterial symbionts and their hosts. *Cell*, **126**, 453-465.

Darby A.C., Birkle L.M., Turner S.L. et Douglas A.E. **2001**. An aphid-borne bacterium allied to the secondary symbionts of whitefly. *FEMS Microbiology Ecology*, **36**, 43-50.

- Darby A.C. et Douglas A.E. **2003**. Elucidation of the transmission patterns of an insect-borne bacterium. *Applied and Environmental Microbiology*, **69**, 4403-4407.
- Darby A.C., Chandler S.M., Welburn S.C. et Douglas A.E. **2005**. Aphid-symbiotic bacteria cultured in insect cell lines. *Applied and Environmental Microbiology*, **71**, 4833-4839.
- Darby A.C., Choi J.H., Wilkes T., Hughes M.A., Werren J.H., Hurst G.D.D. et al. **2010**. Characteristics of the genome of *Arsenophonus nasoniae*, son-killer bacterium of the wasp *Nasonia*. *Insect Molecular Biology*, **19**, 75-89.
- Dave M., Higgins P.D., Middha S. et Rioux K.P. **2012**. The human gut microbiome: current knowledge, challenges, and future directions. *Translational Research : The Journal of Laboratory and Clinical Medicine*, **160**, 246-257.
- Dayel M.J., Alegado R.A., Fairclough S.R., Levin T.C., Nichols S.A., McDonald K et al. **2011**. Cell differentiation and morphogenesis in the colony-forming choanoflagellate *Salpingoeca rosetta*. *Developmental Biology*, **357**, 73-82.
- De Barro P.J. et Driver F. **1997**. Use of RAPD PCR to distinguish the B biotype from other biotypes of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae). *Australian Journal of Entomology*, **36**, 149-152.
- De Barro P.J., Liu S.-S., Boykin L.M. et Dinsdale A.B. **2011**. *Bemisia tabaci*: a statement of species status. *Annual Review of Entomology*, **56**, 1-19.
- de Bary A. **1879**. Die Erscheinung der Symbiose. *Naturforschung Versammlung Casse*.
- De Caralt S., Uriz M.J. et Wijffels R.H. **2007**. Vertical transmission and successive location of symbiotic bacteria during embryo development and larva formation in *Corticium candelabrum* (Porifera: Demospongiae). *Journal of the Marine Biological Association of the UK*, **87**, 1693-1699.
- De Castro C., Kenyon J.J., Cunneen M.M., Molinaro A., Holst O., Skurnik M. et al. **2013**. The O-specific polysaccharide structure and gene cluster of serotype O:12 of the *Yersinia pseudotuberculosis* complex, and the identification of a novel l-quinovose biosynthesis gene. *Glycobiology*, **23**, 346-353.
- Dedeine F., Bouletreau M. et Vavre F. **2005**. *Wolbachia* requirement for oogenesis: occurrence within the genus *Asobara* (Hymenoptera, Braconidae) and evidence for intraspecific variation in *A. tabida*. *Heredity*, **95**, 394-400.
- Degnan P.H., Lazarus A.B. et Wernegreen J.J. **2005**. Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects. *Genome Research*, **15**, 1023-1033.
- Degnan P.H. et Moran N.A. **2008**. Diverse phage-encoded toxins in a protective insect endosymbiont. *Applied and Environmental Microbiology*, **74**, 6782-6791.
- Degnan P.H. et Moran N.A. **2008**. Evolutionary genetics of a defensive facultative symbiont of insects: exchange of toxin-encoding bacteriophage. *Molecular Ecology*, **17**, 916-929.
- Degnan P.H., Yu Y., Sisneros N., Wing R.A. et Moran N.A. **2009**. *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proceedings of the National Academy of Sciences*, **106**, 9063-9068.

- Degnan P.H., Leonardo T.E., Cass B.N., Hurwitz B., Stern D., Gibbs R.A. *et al.* **2010**. Dynamics of genome evolution in facultative symbionts of aphids. *Environmental Microbiology*, **12**, 2060-2069.
- Delatte H., Duyck P.F., Triboire A., David P., Becker N., Bonato O. *et al.* **2009**. Differential invasion success among biotypes: case of *Bemisia tabaci*. *Biological Invasions*, **11**, 1059-1070.
- Delmotte F., Rispe C., Schaber J., Silva F.J. *et Moya A.* **2006**. Tempo and mode of early gene loss in endosymbiotic bacteria from insects. *BMC Evolutionary Biology*, **6**, 56.
- Denison F.R. *et Kiers T.E.* **2004**. Lifestyle alternatives for rhizobia: mutualism, parasitism, and forgoing symbiosis. *FEMS Microbiology Letters*, **237**, 187-193.
- Dennis Jr G., Sherman B.T., Hosack D.A., Yang J., Gao W., Lane H.C. *et al.* (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome biology*, **4**, P3.
- Dhami M.K., Turner A.P., Deines P., Beggs J.R. *et Taylor M.W.* **2012**. Ultrastructural and molecular characterization of a bacterial symbiosis in the ecologically important scale insect family *Coelostomidiidae*. *FEMS Microbiology Ecology*, **81**, 537-546.
- Dinsdale A., Cook L., Riginos C., Buckley Y.M. *et Barro P.D.* **2010**. Refined global analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. *Annals of the Entomological Society of America*, **103**, 196-208.
- Douglas A.E. **1989**. Mycetocyte symbiosis in insects. *Biological Reviews*, **64**, 409-434.
- Douglas A.E. **1993**. The nutritional quality of phloem sap utilized by natural aphid populations. *Ecological Entomology*, **18**, 31-38.
- Douglas A.E. **1998**. Nutritional interactions in insect-microbial symbioses: Aphids and their symbiotic bacteria *Buchnera*. *Annual Review of Entomology*, **43**, 17-37.
- Douglas A.E. **2006**. Phloem-sap feeding by animals: problems and solutions. *Journal of Experimental Botany*, **57**, 747-754.
- Dowling D.N. *et Broughton W.J.* **1986**. Competition for nodulation of legumes. *Annual Reviews in Microbiology*, **40**, 131-157.
- Dufresne A., Garczarek L. *et Partensky F.* **2005**. Accelerated evolution associated with genome reduction in a free-living prokaryote, **6**, R14.
- Duron O., Bouchon D., Boutin S., Bellamy L., Zhou L., Engelstädter J. *et al.* **2008**. The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC Biology*, **6**, 27.
- Duron O. *et Hurst G.D.* **2013**. Arthropods and inherited bacteria: from counting the symbionts to understanding how symbionts count. *BMC Biology*, **11**, 45.
- Dykstra H.R., Weldon S.R., Martinez A.J., White J.A., Hopper K.R., Heimpel G.E. *et al.* **2014**. Factors limiting the spread of the protective symbiont *Hamiltonella defensa* in *Aphis craccivora* aphids. *Applied and Environmental Microbiology*, **80**, 5818-5827.

## E

Ebert D. et Bull J.J. **2003**. Challenging the trade-off model for the evolution of virulence: is virulence management feasible? *Trends in Microbiology*, **11**, 15-20.

Edgar R.C. **2004**. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792-1797.

Engelstädter J. et Hurst G.D. **2009**. The ecology and evolution of microbes that manipulate host reproduction. *Annual Review of Ecology, Evolution and Systematics*, **40**, 127-149.

Everett K.D., Thao M., Horn M., Dyszynski G.E. et Baumann P. **2005**. Novel chlamydiae in whiteflies and scale insects: endosymbionts 'Candidatus Fritschea bemisiae' strain Falk and 'Candidatus Fritschea eriococci' strain Elm. *International Journal of Systematic and Evolutionary Microbiology*, **55**, 1581-1587.

Ewald P.W. **1983**. Host-parasite relations, vectors, and the evolution of disease severity. *Annual Review of Ecology Evolution and Systematics*, **14**, 465-485.

Ewald P.W. **1987**. Transmission modes and evolution of the parasitism-mutualism continuum. *Annals of the New York Academy of Sciences*, **503**, 295-306.

## F

Fang Y., Jiao X., Xie W., Wang S., Wu Q., Shi X. *et al.* (2013). Tomato yellow leaf curl virus alters the host preferences of its vector *Bemisia tabaci*. *Scientific reports*, **3**.

Feldhaar H. **2011**. Bacterial symbionts as mediators of ecologically important traits of insect hosts. *Ecological Entomology*, **36**, 533-543.

Ferrari J. et Vavre F. **2011**. Bacterial symbionts in insects or the story of communities affecting communities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 1389-1400.

Fine P.E. **1975**. Vectors and vertical transmission: an epidemiologic perspective. *Annals of the New York Academy of Sciences*, **266**, 173-194.

Firdaus S., Vosman B., Hidayati N., Supena E.D.J., Visser R.G.F. et van Heusden A.W. **2013**. The *Bemisia tabaci* species complex: additions from different parts of the world. *Insect Science*, **20**, 723-733.

Foray V., Grigorescu A.S., Sabri A., Haubruge E., Lognay G., Francis F. *et al.* **2014**. Whole-genome sequence of *Serratia symbiotica* strain CWBI-2.3 T, a free-living symbiont of the black bean aphid *Aphis fabae*. *Genome announcements*, **2**, e00767-14.

Forterre P. **2013**. The common ancestor of Archaea and Eukarya was not an archaeon. *Archaea*, **2013**.

Forterre P. et Prangishvili D. **2009**. The great billion-year war between ribosome-and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties. *Annals of the New York Academy of Sciences*, **1178**, 65-77.

Frago E., Dicke M. et Godfray H.C.J. **2012**. Insect symbionts as hidden players in insect-plant interactions. *Trends in Ecology & Evolution*, **27**, 705-711.

Frank A.B. **1877**. Über die biologischen Verhältnisse des Thallus einiger Krustenflechten Beitr. *Biologie der Pflanzen*, **2**, 123-200.

Fukatsu T. et Ishikawa H. **1996**. Phylogenetic position of yeast-like symbiont of *Hamiltonaphis styraci* (Homoptera, Aphididae) based on 18S rDNA sequence. *Insect Biochemistry and Molecular Biology*, **26**, 383-388.

Fukatsu T., Nikoh N., Kawai R. et Koga R. **2000**. The secondary endosymbiotic bacterium of the pea aphid *Acyrtosiphon pisum* (Insecta: Homoptera). *Applied and Environmental Microbiology*, **66**, 2748-2758.

Fukatsu T., Tsuchida T., Nikoh N. et Koga R. **2001**. *Spiroplasma* symbiont of the pea aphid, *Acyrtosiphon pisum* (Insecta: Homoptera). *Applied and Environmental Microbiology*, **67**, 1284-1291.

Fytrou A., Schofield P.G., Kraaijeveld A.R. et Hubbard S F. **2006**. *Wolbachia* infection suppresses both host defence and parasitoid counter-defence. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 791-796.

9

Gehrer L. et Vorburger C. **2012**. Parasitoids as vectors of facultative bacterial endosymbionts in aphids. *Biology Letters*, **rsbl20120144**.

Ghanim M., Sobol I., Ghanim M. et Czosnek H. **2007**. Horizontal transmission of begomoviruses between *Bemisia tabaci* biotypes. *Arthropod-Plant Interactions*, **1**, 195-204.

Ghanim M. et Kontsedalov S. **2009**. Susceptibility to insecticides in the Q biotype of *Bemisia tabaci* is correlated with bacterial symbiont densities. *Pest Management Science*, **65**, 939-942.

Gharzouli R., Carpené M.A., Couderc F., Benguedouar A. et Poinso V. **2013**. Relevance of fucose-rich extracellular polysaccharides produced by *Rhizobium sllae* strains nodulating *Hedysarum coronarium* L. legumes. *Applied and Environmental Microbiology*, **79**, 1764-1776.

Gil R., Silva F.J., Zientz E., Delmotte F., González-Candelas F., Latorre A. et al. **2003**. The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proceedings of the National Academy of Sciences*, **100**, 9388-9393.

Gil R., Silva F.J., Peretó J. et Moya A. **2004**. Determination of the core of a minimal bacterial gene set. *Microbiology and Molecular Biology Reviews*, **68**, 518-537.

Gilbert S.F., Sapp J. et Tauber A.I. **2012**. A symbiotic view of life: We have never been individuals. *The Quarterly Review of Biology*, **87**, 325-341.

- Głosnicka R. et Gruszkiewicz E. **1980**. Chemical composition and biological activity of the *Yersinia pestis* envelope substance. *Infection and Immunity*, **30**, 506-512.
- Gnankiné O., Mouton L., Henri H., Terraz G., Houndeté T., Martin T. *et al.* **2013**. Distribution of *Bemisia tabaci* (Homoptera: Aleyrodidae) biotypes and their associated symbiotic bacteria on host plants in West Africa. *Insect Conservation and Diversity*, **6**, 411-421.
- Gómez-Valero L., Soriano-Navarro M., Pérez-Brocal V., Heddi A., Moya A., García-Verdugo J.M. *et al.* **2004**. Coexistence of *Wolbachia* with *Buchnera aphidicola* and a secondary symbiont in the aphid *Cinara cedri*. *Journal of Bacteriology*, **186**, 6626-6633.
- González J.E., Reuhs B.L. et Walker G.C. **1996**. Low molecular weight EPS II of *Rhizobium meliloti* allows nodule invasion in *Medicago sativa*. *Proceedings of the National Academy of Sciences*, **93**, 8636-8641.
- Gottlieb Y., Ghanim M., Chiel E., Gerling D., Portnoy V., Steinberg S. *et al.* **2006**. Identification and localization of a *Rickettsia* sp. in *Bemisia tabaci* (Homoptera: Aleyrodidae). *Applied and Environmental Microbiology*, **72**, 3646-3652.
- Gottlieb Y., Ghanim M., Gueguen G., Kontsedalov S., Vavre F., Fleury F. *et al.* **2008**. Inherited intracellular ecosystem: symbiotic bacteria share bacteriocytes in whiteflies. *The FASEB Journal*, **22**, 2591-2599.
- Grant W.D. **1979**. Cell wall teichoic acid as a reserve phosphate source in *Bacillus subtilis*. *Journal of Bacteriology*, **137**, 35-43.
- Grasswitz T.R. et Paine T.D. **1992**. Kairomonal effect of an aphid cornicle secretion on *Lysiphlebus testaceipes* (Cresson) (Hymenoptera: Aphidiidae). *Journal of Insect Behavior*, **5**, 447-457.
- Griffiths G.W. et Beck S.D. **1974**. Effects of antibiotics on intracellular symbiotes in the pea aphid, *Acyrtosiphon pisum*. *Cell and Tissue Research*, **148**, 287-300.
- Grosdent N., Maridonneau-Parini I., Sory M.P. et Cornelis G.R. **2002**. Role of Yops and adhesins in resistance of *Yersinia enterocolitica* to phagocytosis. *Infection and Immunity*, **70**, 4165-4176.
- Gruwell M.E., Wu J. et Normark B.B. **2009**. Diversity and phylogeny of *Cardinium* (Bacteroidetes) in armored scale insects (Hemiptera: Diaspididae). *Annals of the Entomological Society of America*, **102**, 1050-1061.
- Gruwell M.E., Flarhety M. et Dittmar K. **2012**. Distribution of the primary endosymbiont (*Candidatus Uzinura diaspidicola*) within host insects from the scale insect family Diaspididae. *Insects*, **3**, 262-269.
- Gueguen G., Vavre F., Gnankine O., Peterschmitt M., Charif D., Chiel E. *et al.* **2010**. Endosymbiont metacommunities, mtDNA diversity and the evolution of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. *Molecular Ecology*, **19**, 4365-4378.
- Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W. et Gascuel O. **2010**. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307-321.
- Guindon S., Delsuc F., Dufayard J.F. et Gascuel O. **2009**. Estimating maximum likelihood phylogenies with PhyML. In *Bioinformatics for DNA Sequence Analysis* (pp. 113-137). Humana Press.



Guo X.J., Rao Q., Zhang F., Luo C., Zhang H.Y. et Gao X.W. **2012**. Diversity and genetic differentiation of the whitefly *Bemisia tabaci* species complex in China based on mtCOI and cDNA-AFLP analysis. *Journal of Integrative Agriculture*, **11**, 206-214.

## H

Hansen A.K., Jeong G., Paine T.D. et Stouthamer R. **2007**. Frequency of secondary symbiont infection in an invasive psyllid relates to parasitism pressure on a geographic scale in California. *Applied and Environmental Microbiology*, **73**, 7531-7535.

Hansen A.K., Trumble J.T., Stouthamer R. et Paine T.D. **2008**. A new huanglongbing species, "*Candidatus Liberibacter psyllauros*," found to infect tomato and potato, is vectored by the psyllid *Bactericera cockerelli* (Sulc). *Applied and Environmental Microbiology*, **74**, 5862-5865.

Hansen A.K., Vorburger C. et Moran N.A. **2012**. Genomic basis of endosymbiont-conferred protection against an insect parasitoid. *Genome Research*, **22**, 106-114.

Hansen A.K. et Moran N.A. **2014**. The impact of microbial symbionts on host plant utilization by herbivorous insects. *Molecular Ecology*, **23**, 1473-1496.

Harrison C.P., Douglas A.E. et Dixon A.F.G. **1989**. A rapid method to isolate symbiotic bacteria from aphids. *Journal of Invertebrate Pathology*, **53**, 427-428.

Hemmi H., Ohnuma S.I., Nagaoka K. et Nishino T. **1998**. Identification of genes affecting lycopene formation in *Escherichia coli* transformed with carotenoid biosynthetic genes: candidates for early genes in isoprenoid biosynthesis. *Journal of Biochemistry*, **123**, 1088-1096.

Hentschel U., Steinert M. et Hacker J. **2000**. Common molecular mechanisms of symbiosis and pathogenesis. *Trends in Microbiology*, **8**, 226-231.

Hertig M. et Wolbach S.B. **1924**. Studies on *rickettsia*-like micro-organisms in insects. *The Journal of Medical Research*, **44**, 329.

Hilgenboecker K., Hammerstein P., Schlattmann P., Telschow A. et Werren J.H. **2008**. How many species are infected with *Wolbachia*?—a statistical analysis of current data. *FEMS Microbiology Letters*, **281**, 215-220.

Himler A.G., Adachi-Hagimori T., Bergen J.E., Kozuch A., Kelly S.E., Tabashnik B.E. et al. **2011**. Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. *Science*, **332**, 254-256.

Himpsl S.D., Pearson M.M., Arewång C.J., Nusca T.D., Sherman D.H. et Mobley H.L. **2010**. Proteobactin and a yersiniabactin-related siderophore mediate iron acquisition in *Proteus mirabilis*. *Molecular Microbiology*, **78**, 138-157.

Hiroki M., Kato Y., Kamito T. et Miura K. **2002**. Feminization of genetic males by a symbiotic bacterium in a butterfly, *Eurema hecabe* (Lepidoptera: Pieridae). *Naturwissenschaften*, **89**, 167-170.

Holland I.B., Schmitt L. et Young J. **2005**. Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway. *Molecular Membrane Biology*, **22**, 29-39.

Honegger R. **2000**. Simon Schwendener (1829–1919) and the Dual Hypothesis of Lichens. *The Bryologist*, **103**, 307-313.

Hongoh Y., Sharma V.K., Prakash T., Noda S., Toh H., Taylor T.D. *et al.* **2008**. Genome of an endosymbiont coupling N<sub>2</sub> fixation to cellulolysis within protist cells in termite gut. *Science*, **322**, 1108-1109.

Horowitz A.R., Kontsedalov S., Khasdan V. et Ishaaya I. **2005**. Biotypes B and Q of *Bemisia tabaci* and their relevance to neonicotinoid and pyriproxyfen resistance. *Archives of Insect Biochemistry and Physiology*, **58**, 216-225.

Horst J.P., Wu T.H., Marinus M.G. **1999**. *Escherichia coli* mutator genes. *Trends in Microbiology*, **7**, 29-36.

Hu J., De Barro P., Zhao H., Wang J., Nardi F. et Liu S.S. **2011**. An extensive field survey combined with a phylogenetic analysis reveals rapid and widespread invasion of two alien whiteflies in China. *PLoS One*, **6**, e16061.

Hurst G.D.D., Jiggins F.M. **2005**. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society of London Series B*, **272**, 1525– 1534.

Hurst L.D. et Mcvean G.T. **1996**. Clade selection, reversible evolution and the persistence of selfish elements: the evolutionary dynamics of cytoplasmic incompatibility. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 97-104.

Husník F., Chrudimský T. et Hypša V. **2011**. Multiple origins of endosymbiosis within the Enterobacteriaceae ( $\gamma$ -Proteobacteria): convergence of complex phylogenetic approaches. *BMC Biology*, **9**, 87.

Husnik F., Nikoh N., Koga R., Ross L., Duncan R.P., Fujie M. *et al.* **2013**. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell*, **153**, 1567-1578.

/

International Aphid Genomics Consortium. **2010**. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biology*, **8**, e1000313.

J

Jahan S.M., Lee G.S., Lee S. et Lee K.Y. **2014**. Upregulation of probing-and feeding-related behavioural frequencies in *Bemisia tabaci* upon acquisition of Tomato yellow leaf curl virus. *Pest Management Science*, **70**, 1497-1502.

Jann K., Jann B., Schmidt M.A. et Vann W.F. **1980**. Structure of the *Escherichia coli* K2 capsular antigen, a teichoic acid-like polymer. *Journal of Bacteriology*, **143**, 1108-1115.



Jeffrey C. **1962**. The origin and differentiation of the archegoniate land-plants. *Botaniska Notiser*, **115**, 446-454.

Jiang Z.F., Xia F., Johnson K.W., Bartom E., Tuteja J.H., Stevens R. *et al.* **2012**. Genome sequences of the primary endosymbiont “*Candidatus portiera aleyrodidarum*” in the whitefly *Bemisia tabaci* B and Q biotypes. *Journal of Bacteriology*, **194**, 6678-6679.

Jiang Z., Jones D.H., Khuri S., Tsinoremas N.F., Wyss T., Jander G. *et al.* **2013a**. Comparative analysis of genome sequences from four strains of the *Buchnera aphidicola* Mp endosymbiont of the green peach aphid, *Myzus persicae*. *BMC Genomics*, **14**, 917.

Jiang Z.F., Xia F., Johnson K.W., Brown C.D., Bartom E., Tuteja J.H. *et al.* **2013b**. Comparison of the genome sequences of “*Candidatus Portiera aleyrodidarum*” primary endosymbionts of the whitefly *Bemisia tabaci* B and Q biotypes. *Applied and Environmental Microbiology*, **79**, 1757-1759.

*K*

Kachlany S.C., Planet P.J., DeSalle R., Fine D.H. *et al.* **2001**. Genes for tight adherence of *Actinobacillus actinomycetemcomitans*: from plaque to plague to pond scum. *Trends in Microbiology*, **9**, 429-437.

Karp P.D., Paley S. *et al.* **2002**. The pathway tools software. *Bioinformatics*, **18**, S225-S232.

Kato A. *et al.* **2006**. Growth defect and mutator phenotypes of RecQ-deficient *Neurospora crassa* mutants separately result from homologous recombination and nonhomologous end joining during repair of DNA double-strand breaks. *Genetics*, **172**, 113-125.

Keeling P.J. *et al.* **2008**. Organelle evolution: what's in a name? *Current Biology*, **18**, R345-R347.

Kevan P.G. *et al.* **1983**. Insects as flower visitors and pollinators. *Annual Review of Entomology*, **28**, 407-453.

Khasdan V., Levin I., Rosner A., Morin S., Kotsedalov S., Maslenin *et al.* **2005**. DNA markers for identifying biotypes B and Q of *Bemisia tabaci* (Hemiptera: Aleyrodidae) and studying population dynamics. *Bulletin of Entomological Research*, **95**, 605–613

Kikuchi Y., Hayatsu M., Hosokawa T., Nagayama A., Tago K. *et al.* **2012**. Symbiont-mediated insecticide resistance. *Proceedings of the National Academy of Sciences*, **109**, 8618-8622.

Kliot A., Cilia M., Czosnek H. *et al.* **2014**. Implication of the bacterial endosymbiont *Rickettsia* spp. in interactions of the whitefly *Bemisia tabaci* with Tomato yellow leaf curl virus. *Journal of virology*, **88**, 5652-5660.

Knirel Y.A. *et al.* **2012**. Lipopolysaccharide of *Yersinia pestis*, the cause of plague: structure, genetics, biological properties. *Acta Naturae*, **4**, 46.

Kocharova N.A., Ovchinnikova O.G., Toukach F.V., Torzewska A., Shashkov A.S., Knirel Y.A. *et al.* **2005**. The O-polysaccharide from the lipopolysaccharide of *Providencia stuartii* O44 contains l-quinovose, a 6-deoxy sugar rarely occurring in bacterial polysaccharides. *Carbohydrate Research*, **340**, 1419-1423.

Koga R., Tsuchida T. et Fukatsu T. **2003**. Changing partners in an obligate symbiosis: a facultative endosymbiont can compensate for loss of the essential endosymbiont *Buchnera* in an aphid. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 2543-2550.

Koga R., Meng X.Y., Tsuchida T. et Fukatsu T. **2012**. Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte–embryo interface. *Proceedings of the National Academy of Sciences*, **109**, E1230-E1237.

Köhler R.H., Cao J., Zipfel W.R., Webb W.W. et Hanson M.R. **1997**. Exchange of protein molecules through connections between higher plant plastids. *Science*, **276**, 2039-2042.

Kolibachuk D. et Baumann P. **1995**. Aromatic amino acid biosynthesis in *Buchnera aphidicola* (endosymbiont of aphids): cloning and sequencing of a DNA fragment containing aroH- thrS-infC-rpmL-rplT. *Current Microbiology*, **30**, 313-316.

Kondakova A.N., Kirsheva N.A., Shashkov A.S., Shaikhutdinova R.Z., Arbatsky N.P., Ivanov S.A. et al. **2012**. Structure of the O-polysaccharide of *Photorhabdus luminescens* subsp. *laumondii* containing D-glycero-D-manno-heptose and 3,6-dideoxy-3-formamido-D-glucose. *Carbohydrate Research*, **351**, 134-137.

Korotkov K.V., Sandkvist M. et Hol W.G. **2012**. The type II secretion system: biogenesis, molecular architecture and mechanism. *Nature Reviews Microbiology*, **10**, 336-351.

Kremer N. **2009**. *Évolution de la dépendance dans les symbioses à Wolbachia: étude du genre Asobara (Hymenoptera: braconidae)* (Doctoral dissertation, Lyon 1).

Kremer N., Charif D., Henri H., Bataille M., Prévost G., Kraaijeveld K. et Vavre F. **2009**. A new case of *Wolbachia* dependence in the genus *Asobara*: evidence for parthenogenesis induction in *Asobara japonica*. *Heredity*, **103**, 248-256.

↳

Lai C.Y., Baumann L. et Baumann P. **1994**. Amplification of trpEG: adaptation of *Buchnera aphidicola* to an endosymbiotic association with aphids. *Proceedings of the National Academy of Sciences*, **91**, 3819-3823.

Lagesen K., Hallin P., Rodland E.A., Staerfeldt H.H., Rognes T., Ussery D.W. **2007**. Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, **35**, 3100-3108.

Lamelas A., Pérez-Brocal V., Gómez-Valero L., Gosalbes M.J., Moya A. et Latorre A. **2008**. Evolution of the secondary symbiont “*Candidatus Serratia symbiotica*” in aphid species of the subfamily Lachninae. *Applied and Environmental Microbiology*, **74**, 4236-4240.

Lamelas A., Gosalbes M.J., Manzano-Marín A., Peretó J., Moya A. et Latorre A. **2011**. *Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont. *PLoS Genetics*, **7**, e1002357.

Lang B.F., Gray M.W. et Burger G. **1999**. Mitochondrial genome evolution and the origin of eukaryotes. *Annual Review of Genetics*, **33**, 351-397.

- Lapouge K., Schubert M., Allain F.H. et Haas D. **2008**. Gac/Rsm signal transduction pathway of gamma-proteobacteria: from RNA recognition to regulation of social behaviour. *Molecular Microbiology*, **67**, 241-253.
- Latorre A., Gil R., Silva F.J. et Moya A. **2005**. Chromosomal stasis versus plasmid plasticity in aphid endosymbiont *Buchnera aphidicola*. *Heredity*, **95**, 339-347.
- Lee K.H. et Ruby E.G. **1994**. Effect of the squid host on the abundance and distribution of symbiotic *Vibrio fischeri* in nature. *Applied and Environmental Microbiology*, **60**, 1565-1571.
- Lengyel S., Gove A.D., Latimer A.M., Majer J.D. et Dunn R.R. **2010**. Convergent evolution of seed dispersal by ants, and phylogeny and biogeography in flowering plants: a global survey. *Perspectives in Plant Ecology, Evolution and Systematics*, **12**, 43-55.
- Leonardo T.E. **2004**. Removal of a specialization-associated symbiont does not affect aphid fitness. *Ecology Letters*, **7**, 461-468.
- Leonardo T.E. et Mondor E.B. **2006**. Symbiont modifies host life-history traits that affect gene flow. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 1079-1084.
- Lerat E., Daubin V. et Moran N.A. **2003**. From gene trees to organismal phylogeny in prokaryotes: the case of the  $\gamma$ -Proteobacteria. *PLoS Biology*, **1**, e19.
- Lerat E. et Ochman H. **2005**. Recognizing the pseudogenes in bacterial genomes. *Nucleic Acids Research*, **33**, 3125-3132.
- Ley R.E., Peterson D.A. et Gordon J.I. **2006**. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*, **124**, 837-848.
- Lin H., Lou B., Glynn J.M., Doddapaneni H., Civerolo E.L., Chen C. **2011**. The complete genome sequence of “*Candidatus Liberibacter solanacearum*”, the bacterium associated with potato zebra chip disease. *PLoS One*, **6**, e19135.
- Lin S.H. et Liao Y.C. **2013**. CISA: contig integrator for sequence assembly of bacterial genomes. *PLoS One*, **8**, e60843.
- Liu B., Knirel Y.A., Feng L., Perepelov A.V., Senchenkova S.Y.N., Wang Q. *et al.* **2008**. Structure and genetics of *Shigella* O antigens. *FEMS Microbiology Reviews*, **32**, 627-653.
- Liu S.S., Colvin J. et De Barro P.J. **2012**. Species concepts as applied to the whitefly *Bemisia tabaci* systematics: how many species are there? *Journal of Integrative Agriculture*, **11**, 176-186.
- López-Madrigal S., Latorre A., Porcar M., Moya A. et Gil, R. **2011**. Complete genome sequence of “*Candidatus Tremblaya princeps*” strain PCVAL, an intriguing translational machine below the living-cell status. *Journal of Bacteriology*, **193**, 5587-5588.
- López-Madrigal S., Latorre A., Porcar M., Moya A. et Gil R. **2013**. Mealybugs nested endosymbiosis: going into the ‘matryoshka’ system in *Planococcus citri* in depth. *BMC Microbiology*, **13**, 74.
- López-Sánchez M.J., Neef A., Peretó J., Patiño-Navarrete R., Pignatelli M., Latorre A. *et al.* **2009**. Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. *PLoS Genetics*, **5**, e1000721.

Lu S., Killoran P.B., Fang F.C. et Riley L.W. **2002**. The global regulator ArcA controls resistance to reactive nitrogen and oxygen intermediates in *Salmonella enterica* serovar Enteritidis. *Infection and Immunity*, **70**, 451-461.

Luisi P.L. **2002**. Toward the engineering of minimal living cells. *The Anatomical Record*, **268**, 208-214.

Łukasik P., van Asch M., Guo H., Ferrari J. et Godfray C.J. **2013**. Unrelated facultative endosymbionts protect aphids against a fungal pathogen. *Ecology Letters*, **16**, 214-218.

Luo C., Jones C.M., Devine G., Zhang F., Denholm I. et Gorman K. **2010**. Insecticide resistance in *Bemisia tabaci* biotype Q (Hemiptera: Aleyrodidae) from China. *Crop Protection*, **29**, 429-434.

Lucas-Lledó J.I. et Lynch M. **2009**. Evolution of mutation rates: phylogenomic analysis of the photolyase/cryptochrome family. *Molecular Biology and Evolution*, **26**, 1143-1153.

Lynch V.J., Leclerc R.D., May G. et Wagner G.P. **2011**. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nature Genetics*, **43**, 1154-1159.

## M

Mackay W.J., Han S. et Samson L.D. **1994**. DNA alkylation repair limits spontaneous base substitution mutations in *Escherichia coli*. *Journal of Bacteriology*, **176**, 3224-3230.

Mahadav A., Gerling D., Gottlieb Y., Czosnek H. et Ghanim M. **2008**. Parasitization by the wasp *Eretmocerus mundus* induces transcription of genes related to immune response and symbiotic bacteria proliferation in the whitefly *Bemisia tabaci*. *BMC Genomics*, **9**, 342.

Maltz M. et Graf J. **2011**. The type II secretion system is essential for erythrocyte lysis and gut colonization by the leech digestive tract symbiont *Aeromonas veronii*. *Applied and Environmental Microbiology*, **77**, 597-603.

Mann R.S., Pelz-Stelinski K., Hermann S.L., Tiwari S. et Stelinski L.L. **2011**. Sexual transmission of a plant pathogenic bacterium, *Candidatus Liberibacter asiaticus*, between conspecific insect vectors during mating. *PloS One*, **6**, e29197.

Manzano-Marín A., Lamelas A., Moya A. et Latorre A. **2012**. Comparative genomics of *Serratia* spp.: two paths towards endosymbiotic life. *PloS One*, **7**, e47274.

Manzano-Marín A. et Latorre A. **2014**. Settling down: The genome of *Serratia symbiotica* from the aphid *Cinara tujafilina* zooms in on the process of accommodation to a cooperative intracellular life. *Genome Biology and Evolution*, **6**, 1683-1698.

Marais G.A.B., Calteau A. et Tenaillon O. **2008**. Mutation rate and genome reduction in endosymbiotic and free-living bacteria. *Genetica*, **134**, 205-210.

Margulis L. **1981**. *Symbiosis in cell evolution*. Freeman, San Francisco

- Margulis L. et Fester R. **1991**. *Symbiosis as a source of evolutionary innovation: speciation and morphogenesis*. MIT Press.
- Martin V.J. et Mohn W.W. **2000**. Genetic investigation of the catabolic pathway for degradation of abietane diterpenoids by *Pseudomonas abietaniphila* BKME-9. *Journal of Bacteriology*, **182**, 3784-3793.
- Martinez-Antonio A. et Collado-Vides J. **2003**. Identifying global regulators in transcriptional regulatory networks in bacteria. *Current Opinion in Microbiology*, **6**, 482-489.
- Matsuura Y., Koga R., Nikoh N., Meng X.-Y., Hanada S. et Fukatsu T. **2009**. Huge symbiotic organs in giant scale insects of the genus *Drosicha* (Coccoidea: Monophlebidae) harbor flavobacterial and enterobacterial endosymbionts. *Zoological Science*, **26**, 448-456.
- McCutcheon J.P. et Moran N.A. **2007**. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proceedings of the National Academy of Sciences*, **104**, 19392-19397.
- McCutcheon J.P., McDonald B.R. et Moran N.A. **2009a**. Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genetics*, **5**, e1000565.
- McCutcheon J.P., McDonald B.R. et Moran N.A. **2009b**. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proceedings of the National Academy of Sciences*, **106**, 15394-15399.
- McCutcheon J.P. et Moran N.A. **2010**. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biology and Evolution*, **2**, 708-718.
- McCutcheon J.P. et von Dohlen C.D. **2011**. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Current Biology*, **21**, 1366-1372.
- McFall-Ngai M.J. **1990**. Cypsis in the pelagic environment. *American Zoologist*, **30**, 175-188.
- Miele V., Penel S. et Duret L. **2011**. Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC bioinformatics*, **12**, 116.
- Milne I., Bayer M., Cardle L., Shaw P., Stephen G., Wright F. et Marshall D. **2010**. Tablet—next generation sequence assembly visualization. *Bioinformatics*, **26**, 401-402.
- Mira A. et Moran N.A. **2002**. Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microbial Ecology*, **44**, 137-143.
- Mittler T.E. **1971**. Dietary amino acid requirements of the aphid *Myzus persicae* affected by antibiotic uptake. *The Journal of Nutrition*, **101**, 1023-1028.
- Montllor C.B., Maxmen A. et Purcell A.H. **2002**. Facultative bacterial endosymbionts benefit pea aphids *Acyrtosiphon pisum* under heat stress. *Ecological Entomology*, **27**, 189-195.
- Moran N.A., Munson M.A., Baumann P. et Ishikawa H. **1993**. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proceedings of the Royal Society B: Biological Sciences*, **253**, 167-171.

- Moran N.A. et Baumann L. **1994**. Phylogenetics of cytoplasmically inherited microorganisms of arthropods. *Trends in Ecology and Evolution*, **9**, 15–20.
- Moran N.A. **1996**. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences*, **93**, 2873-2878.
- Moran N.A. et Plague G.R. **2004**. Genomic changes following host restriction in bacteria. *Current Opinion in Genetics & Development*, **14**, 627-633.
- Moran N.A., Degnan P.H., Santos S.R., Dunbar H.E. et Ochman H. **2005a**. The players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. *Proceedings of the National Academy of Sciences*, **102**, 16919-16926.
- Moran N.A., Dunbar H.E. et Wilcox J.L. **2005b**. Regulation of transcription in a reduced bacterial genome : nutrient-provisioning genes of the obligate symbiont *Buchnera aphidicola*. *Journal of Bacteriology*, **187**, 4229-4237.
- Moran N.A., Russell J.A., Koga R. et Fukatsu T. **2005c**. Evolutionary relationships of three new species of *Enterobacteriaceae* living as symbionts of aphids and other insects. *Applied and Environmental Microbiology*, **71**, 3302-3310.
- Moran N.A. et Dunbar H.E. **2006**. Sexual acquisition of beneficial symbionts in aphids. *Proceedings of the National Academy of Sciences*, **103**, 12803-12806.
- Moran N.A., McCutcheon J.P. et Nakabachi A. **2008**. Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics*, **42**, 165-190.
- Moran N.A. et Jarvik T. **2010**. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science*, **328**, 624-627.
- Moriya Y., Itoh M., Okuda S., Yoshizawa A.C. et Kanehisa M. **2007**. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, **35**, W182-W185.
- Mouton L., Henri H., Bouletreau M. et Vavre F. **2003**. Strain-specific regulation of intracellular *Wolbachia* density in multiply infected insects. *Molecular Ecology*, **12**, 3459-3465.
- Mouton L., Thierry M., Henri H., Baudin R., Gnankine O., Reynaud B. *et al.* **2012**. Evidence of diversity and recombination in *Arsenophonus* symbionts of the *Bemisia tabaci* species complex. *BMC Microbiology*, **12**, S10.
- Mouton L., Gnankiné O., Henri H., Terraz G., Ketoh G., Martin T. *et al.* **2014**. Detection of genetically isolated entities within the Mediterranean species of *Bemisia tabaci*: new insights into the systematics of this worldwide pest. *Pest Management Science*.
- Moya A., Pereto J., Gil R. et Latorre A. **2008**. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nature Reviews Genetics*, **9**, 218-229.
- Muller H.J. **1932**. Some genetic aspects of sex. *American Naturalist*, **66**, 118-138.
- Munson M.A., Baumann P. et Kinsey M.G. **1991**. *Buchnera* gen. nov. and *Buchnera aphidicola* sp. nov., a taxon consisting of the mycetocyte-associated, primary endosymbionts of aphids. *International Journal of Systematic Bacteriology*, **41**, 566-568.



Murray R.G.E. et Schleifer K.H. **1994**. Taxonomic notes: a proposal for recording the properties of putative taxa of prokaryotes. *International Journal of Systematic Bacteriology*, **44**, 174-176.

Murray G.L., Attridge S.R. et Morona R. **2003**. Regulation of *Salmonella typhimurium* lipopolysaccharide O antigen chain length is required for virulence; identification of FepE as a second Wzz. *Molecular Microbiology*, **47**, 1395-1406.

Muscatine L. et Porter J.W. **1977**. Reef corals: mutualistic symbioses adapted to nutrient-poor environments. *Bioscience*, **27**, 454-460.

## N

Nachappa P., Shapiro A.A. et Tamborindeguy C. **2012**. Effect of 'Candidatus Liberibacter solanacearum' on fitness of its insect vector, *Bactericera cockerelli* (Hemiptera: Trioizidae), on Tomato. *Phytopathology*, **102**, 41-46.

Nachappa P., Levy J., Pierson E. et Tamborindeguy C. **2013**. Correlation between "Candidatus Liberibacter solanacearum" infection levels and fecundity in its psyllid vector. *Journal of Invertebrate Pathology*, **115**, 55-61.

Nachappa P., Levy J., Pierson E. et Tamborindeguy C. **2014**. Correlation between "Candidatus Liberibacter solanacearum" infection levels and fecundity in its psyllid vector. *Journal of Invertebrate Pathology*, **115**, 55-61.

Nakabachi A. et Ishikawa H. **1999**. Provision of riboflavin to the host aphid, *Acyrtosiphon pisum*, by endosymbiotic bacteria, *Buchnera*. *Journal of Insect Physiology*, **45**, 1-6.

Nakabachi A., Yamashita A., Toh H., Ishikawa H., Dunbar H.E., Moran N.A. *et al.* **2006**. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science*, **314**, 267.

Nakabachi A., Ueoka R., Oshima K., Teta R., Mangoni A., Gurgui M. *et al.* **2013**. Defensive bacteriome symbiont with a drastically reduced genome. *Current Biology*, **23**, 1478-1484.

Nakamura Y., Kawai S., Yukuhiro F., Ito S., Gotoh T., Kisimoto R. *et al.* **2009**. Prevalence of *Cardinium* bacteria in planthoppers and spider mites and taxonomic revision of "Candidatus Cardinium hertigii" based on detection of a new *Cardinium* group from biting midges. *Applied and Environmental Microbiology*, **75**, 6757-6763.

Nandineni M.R. et Gowrishankar J. **2004**. Evidence for an arginine exporter encoded by yggA (argO) that is regulated by the LysR-type transcriptional regulator ArgP in *Escherichia coli*. *Journal of Bacteriology*, **186**, 3539-3546.

Nardon P. et Heddi A. **2013**. La symbiose chez les insectes phytophages et granivores. *Interactions Insectes-Plantes*, **127**.

Neupane S., Högberg N., Alström S., Lucas S., Han J., Lapidus A. *et al.* **2012**. Complete genome sequence of the rapeseed plant-growth promoting *Serratia plymuthica* strain AS9. *Standards in genomic sciences*, **6**, 54.

Niemann G.S., Brown R.N., Gustin J.K., Stufkens A., Shaikh-Kidwai A.S., Li J. *et al.* **2011**. Discovery of novel secreted virulence factors from *Salmonella enterica* serovar Typhimurium by proteomic analysis of culture supernatants. *Infection and Immunity*, **79**, 33-43.

Nikoh N., Hosokawa T., Moriyama M., Oshima K., Hattori M. et Fukatsu T. **2014**. Evolutionary origin of insect–*Wolbachia* nutritional mutualism. *Proceedings of the National Academy of Sciences*, **111**, 10257-10262.

Nogge G. **1982**. Significance of symbionts for the maintenance of an optional nutritional state for successful reproduction in hematophagous arthropods. *Parasitology*, **82**, 101-104.

Normark B.B. **2004**. Haplodiploidy as an outcome of coevolution between male-killing cytoplasmic elements and their hosts. *Evolution*, **58**, 790-798.

Nováková E., Hypša V. et Moran N.A. **2009**. *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. *BMC Microbiology*, **9**, 143.

Nováková E. et Moran N.A. **2012**. Diversification of genes for carotenoid biosynthesis in aphids following an ancient transfer from a fungus. *Molecular Biology and Evolution*, **29**, 313-323.

Nuccio S.P. et Bäumlner A.J. **2007**. Evolution of the chaperone/usher assembly pathway: fimbrial classification goes Greek. *Microbiology and Molecular Biology Reviews*, **71**, 551-575.

## 0

Oakeson K.F., Gil R., Clayton A.L., Dunn D.M., von Niederhausern A.C., Hamil C. *et al.* **2014**. Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome Biology and Evolution*, **6**, 76-93.

Ochman H. et Moran N.A. **2001**. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science*, **292**, 1096-1099.

Ogata H., Renesto P., Audic S., Robert C., Blanc G., Fournier P.E. *et al.* **2005**. The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biology*, **3**, e248.

Ogata H., La Scola B., Audic S., Renesto P., Blanc G., Robert C. *et al.* **2006**. Genome sequence of *Rickettsia bellii* illuminates the role of *amoebae* in gene exchanges between intracellular pathogens. *PLoS Genetics*, **2**, e76.

Ohtaka C. et Ishikawa H. **1991**. Effects of heat treatment on the symbiotic system of an aphid mycetocyte. *Symbiosis*, **11**, 19-30.

Okan N.A., Bliska J.B. et Karzai A.W. **2006**. A role for the SmpB-SsrA system in *Yersinia pseudotuberculosis* pathogenesis. *PLoS Pathogens*, **2**, e6.

Okan N.A., Mena P., Benach J.L., Bliska J.B. et Karzai A.W. **2010**. The smpB-ssrA mutant of *Yersinia pestis* functions as a live attenuated vaccine to protect mice against pulmonary plague infection. *Infection and Immunity*, **78**, 1284-1293.



Oliver K.M., Russell J.A., Moran N.A. et Hunter M.S. **2003**. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the National Academy of Sciences*, **100**, 1803-1807.

Oliver K.M., Moran N.A. et Hunter M.S. **2005**. Variation in resistance to parasitism in aphids is due to symbionts not host genotype. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 12795-12800.

Oliver K.M., Moran N.A. et Hunter M.S. **2006**. Costs and benefits of a superinfection of facultative symbionts in aphids. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 1273-1280.

Oliver K.M., Campos J., Moran N.A. et Hunter M.S. **2008**. Population dynamics of defensive symbionts in aphids. *Proceedings of the Royal Society B: Biological Sciences*, **275**, 293-299.

Oliver K.M., Degnan P.H., Hunter M.S. et Moran N.A. **2009**. Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science*, **325**, 992-994.

Oliver K.M., Degnan P.H., Burke G.R. et Moran N.A. **2010**. Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annual Review of Entomology*, **55**, 247-266.

Orth K. **2002**. Function of the *Yersinia* effector YopJ. *Current Opinion in Microbiology*, **5**, 38-43.

## P

Pan H., Li X., Ge D., Wang S., Wu Q., Xie W. *et al.* **2012**. Factors affecting population dynamics of maternally transmitted endosymbionts in *Bemisia tabaci*. *PloS One*, **7**, e30760.

Parrella G., Nappo A.G., Manco E., Greco B. et Giorgini M. **2013**. Invasion of the Q2 mitochondrial variant of Mediterranean *Bemisia tabaci* in southern Italy: possible role of bacterial endosymbionts. *Pest Management Science*, **70**, 1514-1523.

Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G., Prentice M.B. *et al.* **2001**. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature*, **413**, 523-527.

Pascual S. et Callejas C. **2004**. Intra-and interspecific competition between biotypes B and Q of *Bemisia tabaci* (Hemiptera: Aleyrodidae) from Spain. *Bulletin of Entomological Research*, **94**, 369-375.

Penel S., Arigon A.M., Dufayard J.F., Sertier A.S., Daubin V., Duret L. *et al.* **2009**. Databases of homologous gene families for comparative genomics. *BMC Bioinformatics*, **10**, S3.

Penz T., Schmitz-Esser S., Kelly S.E., Cass B.N., Müller A., Woyke T. *et al.* **2012**. Comparative genomics suggests an independent origin of cytoplasmic incompatibility in *Cardinium hertigii*. *PLoS Genetics*, **8**, e1003012.

Perepelov A.V., Wang Q., Senchenkova S.Y.N., Shevelev S.D., Zhao G., Shashkov A.S. *et al.* **2006**. Structure of a teichoic acid-like O-polysaccharide of *Escherichia coli* O29. *Carbohydrate Research*, **341**, 2176-2180.

Pérez-Brocal V., Gil R., Ramos S., Lamelas A., Postigo M., Michelena J.M. *et al.* **2006**. A small microbial genome: the end of a long symbiotic relationship? *Science*, **314**, 312-313.

Perotti M.A., Allen J.M., Reed D.L. et Braig H.R. **2007**. Host-symbiont interactions of the primary endosymbiont of human head and body lice. *The FASEB Journal*, **21**, 1058-1066.

Perring T.M. **2001**. The *Bemisia tabaci* species complex. *Crop Protection*, **20**, 725-737.

Perry R.D. et Fetherston J.D. **1997**. *Yersinia pestis*--etiologic agent of plague. *Clinical Microbiology Reviews*, **10**, 35-66.

Petersen L.M. et Tisa L.S. **2013**. Friend or foe? A review of the mechanisms that drive *Serratia* towards diverse lifestyles. *Canadian Journal of Microbiology*, **59**, 627-640.

Plague G.R., Dunbar H.E., Tran P.L. et Moran N.A. **2008**. Extensive proliferation of transposable elements in heritable bacterial symbionts. *Journal of Bacteriology*, **190**, 777-779.

Pontes M.H., Babst M., Lochhead R., Oakeson K., Smith K. et Dale C. **2008**. Quorum sensing primes the oxidative stress response in the insect endosymbiont, *Sodalis glossinidius*. *PLoS One*, **3**, e3541.

Posada D. **2008**. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253-1256.

Post D.M., Yu L., Krasity B.C., Choudhury B., Mandel M.J., Brennan C.A. et al. **2012**. O-antigen and core carbohydrate of *Vibrio fischeri* lipopolysaccharide composition and analysis of their role in *Euprymna scolopes* light organ colonization. *Journal of Biological Chemistry*, **287**, 8515-8530.

Price D.R.G., Feng H., Baker J.D., Bavan S., Luetje C.W. et Wilson A.C.C. **2014**. Aphid amino acid transporter regulates glutamine supply to intracellular bacterial symbionts. *Proceedings of the National Academy of Sciences*, **111**, 320-325.

Provencher L.M., Morse G.E., Weeks A.R. et Normark B.B. **2005**. Parthenogenesis in the *Aspidiotus nerii* complex (Hemiptera: Diaspididae): a single origin of a worldwide , polyphagous lineage associated with *Cardinium* bacteria. *Annals of the Entomological Society of America*, **98**, 629-635.

Purcell A.H. **1982**. Insect vector relationships with procaryotic plant pathogens. *Annual Review of Phytopathology*, **20**, 397-417.

## Q

Qin J., Li R., Raes J., Arumugam M., Burgdorf K.S., Manichanh C. et al. **2010**. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, **464**, 59-65.

## R

Rao Q., Luo C., Zhang H., Guo X. et Devine G.J. **2011**. Distribution and dynamics of *Bemisia tabaci* invasive biotypes in central China. *Bulletin of Entomological Research*, **101**, 81-88.

Rao Q., Wang S., Su Y.L., Bing X.L., Liu S.S. et Wang X.W. **2012a**. Draft genome sequence of "*Candidatus* Hamiltonella defensa," an endosymbiont of the whitefly *Bemisia tabaci*. *Journal of Bacteriology*, **194**, 3558-3558.

Rao Q., Xu Y.H., Luo C., Zhang H.Y., Jones C.M., Devine G.J. *et al.* **2012b**. Characterisation of neonicotinoid and pymetrozine resistance in strains of *Bemisia tabaci* (Hemiptera: Aleyrodidae) from China. *Journal of Integrative Agriculture*, **11**, 321-326.

Raz E. **2009**. Mucosal immunity: aliment and ailments. *Mucosal immunology*, **3**, 4-7.

Ren Q. et Paulsen I.T. **2005**. Comparative analyses of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. *PLoS Computational Biology*, **1**, e27.

Roossinck M.J. **2011**. The good viruses: viral mutualistic symbioses. *Nature Reviews Microbiology*, **9**, 99-108.

Rosenberg E., Koren O., Reshef L., Efrony R. et Zilber-Rosenberg I. **2007**. The role of microorganisms in coral health, disease and evolution. *Nature Reviews Microbiology*, **5**, 355-362.

Rubinstein G. et Czosnek H. **1997**. Long-term association of tomato yellow leaf curl virus with its whitefly vector *Bemisia tabaci*: effect on the insect transmission capacity, longevity and fecundity. *Journal of General Virology*, **78**, 2683-2689.

Rosas-Pérez T., Rosenblueth M., Rincón-Rosales R., Mora J. et Martínez-Romero E. **2014**. Genome sequence of “*Candidatus Walczuchella monophlebidarum*” the flavobacterial endosymbiont of *Llaveia axin axin* (Hemiptera: Coccoidea: Monophlebidae). *Genome Biology and Evolution*, **6**, 714-726.

Ross L., Shuker D.M., Normark B.B. et Pen I. **2012**. The role of endosymbionts in the evolution of haploid-male genetic systems in scale insects (Coccoidea). *Ecology and Evolution*, **2**, 1071-1081.

Rothman S.C. et Kirsch J.F. **2003**. How does an enzyme evolved in vitro compare to naturally occurring homologs possessing the targeted function? Tyrosine aminotransferase from aspartate aminotransferase. *Journal of Molecular Biology*, **327**, 593-608.

Russell J.A., Latorre A., Sabater-Muñoz B., Moya A. et Moran N.A. **2003**. Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. *Molecular Ecology*, **12**, 1061-1075.

Russell J.A. et Moran N.A. **2006**. Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 603-610.

Russell C.W., Bouvaine S., Newell P.D. et Douglas A.E. **2013**. Shared metabolic pathways in a coevolved insect-bacterial symbiosis. *Applied and Environmental Microbiology*, **79**, 6117-6123.

Ryan F.P. **2009**. An alternative approach to medical genetics based on modern evolutionary biology. Part 2: retroviral symbiosis. *Journal of the Royal Society of Medicine*, **102**, 324-331.



Sabree Z.L., Huang C.Y., Okusu A., Moran N.A. et Normark B.B. **2013**. The nutrient supplying capabilities of *Uzinura*, an endosymbiont of armoured scale insects. *Environmental Microbiology*, **15**, 1988-1999.

- Sachs J.L., Mueller U.G., Wilcox T.P. et Bull J.J. **2004**. The evolution of cooperation. *The Quarterly Review of Biology*, **79**, 135-160.
- Sachs J.L. et Simms E.L. **2006**. Pathways to mutualism breakdown. *Trends in Ecology & Evolution*, **21**, 585-592.
- Sachs J.L. et Wilcox T.P. **2006**. A shift to parasitism in the jellyfish symbiont *Symbiodinium microadriaticum*. *Proceedings of The Royal Society B: Biological Sciences*, **273**, 425-429.
- Sachs J.L., Skophammer R.G. et Regus J.U. **2011**. Evolutionary transitions in bacterial symbiosis. *Proceedings of the National Academy of Sciences*, **108**, 10800-10807.
- Sachs J.L., Skophammer R.G., Bansal N. et Stajich J.E. **2014**. Evolutionary origins and diversification of proteobacterial mutualists. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20132146.
- Saldías M.S., Ortega X. et Valvano M.A. **2009**. *Burkholderia cenocepacia* O antigen lipopolysaccharide prevents phagocytosis by macrophages and adhesion to epithelial cells. *Journal of Medical Microbiology*, **58**, 1542-1548.
- Sandkvist M. **2001**. Biology of type II secretion. *Molecular Microbiology*, **40**, 271-283.
- Sandström J.P., Russell J.A., White J.P. et Moran N.A. **2001**. Independent origins and horizontal transfer of bacterial symbionts of aphids. *Molecular Ecology*, **10**, 217-228.
- Santos-Garcia D., Farnier P.A., Beitia F., Zchori-Fein E., Vavre F., Mouton L. *et al.* **2012**. Complete genome sequence of “*Candidatus Portiera aleyrodidarum*” BT-QVLC, an obligate symbiont that supplies amino acids and carotenoids to *Bemisia tabaci*. *Journal of Bacteriology*, **194**, 6654-6655.
- Santos-Garcia D., Latorre A., Moya A., Gibbs G., Hartung V., Dettner K. *et al.* **2014**. Small but powerful, the primary endosymbiont of moss bugs, *Candidatus Evansia muelleri*, holds a reduced genome with large biosynthetic capabilities. *Genome Biology and Evolution*, **6**, 1875-1893.
- Sapp J. **2004**. The dynamics of symbiosis: an historical overview. *Canadian Journal of Botany*, **1056**, 1046-1056.
- Scarborough C.L., Ferrari J. et Godfray H.C.J. **2005**. Aphid protected from pathogen by endosymbiont. *Science*, **310**, 1781-1781.
- Schattner P., Brooks A.N., Lowe T.M. **2005**. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research*, **33**, W686-689.
- Schmid M., Sieber R., Zimmermann Y.S. et Vorburger C. **2012**. Development, specificity and sublethal effects of symbiont-conferred resistance to parasitoids in aphids. *Functional Ecology*, **26**, 207-215.
- Schmitz A., Anselme C., Ravallec M., Rebuf C., Simon J.C., Gatti J.L. *et al.* **2012**. The cellular immune response of the pea aphid to foreign intrusion and symbiotic challenge. *PLoS One*, **7**, e42114.
- Shoemaker D.D., Dyer K.A., Ahrens M., McAbee K. et Jaenike J. **2004**. Decreased diversity but increased substitution rate in host mtDNA as a consequence of *Wolbachia* endosymbiont infection. *Genetics*, **168**, 2049–2058.

- Schröder D., Deppisch H., Obermayer M., Krohne G., Stackebrandt E., Hölldobler B. *et al.* **1996**. Intracellular endosymbiotic bacteria of *Camponotus* species (carpenter ants): systematics, evolution and ultrastructural characterization. *Molecular Microbiology*, **21**, 479-489.
- Schulz P.D.A. **1998**. *Phloem. Structure related to function*. In Progress in Botany, 429-475. Springer Berlin Heidelberg.
- Schwendener S. **1869**. Die Algentypen der Flechtengonidien. *Schultze*, Basel.
- Selander R.K., Caugant D.A. et Whittam T.S. **1987**. *Genetic structure and variation in natural populations of Escherichia coli*. In Neidhardt F.C., Ingraham J.L., Low K.B., Magasanik B., Schaechter M. et Umberger H.E. (ed.), *Escherichia coli and Salmonella typhimurium: cellular and molecular biology*. American Society for Microbiology, Washington, D.C., 1625-1648.
- Shi X., Pan H., Zhang H., Jiao X., Xie W., Wu Q. *et al.* **2014**. *Bemisia tabaci* Q carrying tomato yellow leaf curl virus strongly suppresses host plant defenses. *Scientific reports*, **4**.
- Shigenobu S., Watanabe H., Hattori M., Sakaki Y. et Ishikawa H. **2000**. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature*, **407**, 81-86.
- Shigenobu S. et Wilson A.C. **2011**. Genomic revelations of a mutualism: the pea aphid and its obligate bacterial symbiont. *Cellular and Molecular Life Sciences*, **68**, 1297-1309.
- Silva F.J., Latorre A. et Moya A. **2001**. Genome size reduction through multiple events of gene disintegration in *Buchnera* APS. *Trends in Genetics*, **17**, 615-618.
- Silva F.J., Latorre A. et Moya A. **2003**. Why are the genomes of endosymbiotic bacteria so stable? *Trends in Genetics*, **19**, 176-180.
- Silva F.J., Belda E. et Talens S.E. **2006**. Differential annotation of tRNA genes with anticodon CAT in bacterial genomes. *Nucleic Acids Research*, **34**, 6015-6022.
- Simon J.-C., Boutin S., Tsuchida T., Koga R., Le Gallic J.-F., Frantz A. *et al.* **2011**. Facultative symbiont infections affect aphid reproduction. *PloS One*, **6**, e21831.
- Sintupachee S., Milne J.R., Poonchaisri S., Baimai V. et Kittayapong P. **2006**. Closely related *Wolbachia* strains within the pumpkin arthropod community and the potential for horizontal transmission via the plant. *Microbial Ecology*, **51**, 294-301.
- Skaljac M., Zanic K., Ban S.G., Kontsedalov S. et Ghanim M. **2010**. Co-infection and localization of secondary symbionts in two whitefly species. *BMC Microbiology*, **10**, 142.
- Skurnik M., Peippo A. et Erelä E. **2000**. Characterization of the O-antigen gene clusters of *Yersinia pseudotuberculosis* and the cryptic O-antigen gene cluster of *Yersinia pestis* shows that the plague bacillus is most closely related to and has evolved from *Y. pseudotuberculosis* serotype O: 1b. *Molecular Microbiology*, **37**, 316-330.
- Sloan D.B. et Moran N.A. **2012a**. Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. *Molecular Biology and Evolution*, **29**, 3781-3792.
- Sloan D.B. et Moran N.A. **2012b**. Endosymbiotic bacteria as a source of carotenoids in whiteflies. *Biology Letters*, **8**, 986-989.

Sloan D.B. et Moran N.A. **2013**. The evolution of genomic instability in the obligate endosymbionts of whiteflies. *Genome Biology and Evolution*, **5**, 783-793.

Sloan D.B., Nakabachi A., Richards S., Qu J., Murali S.C., Gibbs R.A. *et al.* **2014**. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Molecular Biology and Evolution*, **31**, 857-871.

Smith F.A. et Smith S.E. **1996**. Mutualism and parasitism: diversity in function and structure in the 'Arbuscular' (VA) mycorrhizal symbiosis. *Advances in Botanical Research*, **22**, 1-43.

Smoot M.E., Ono K., Ruscheinski J., Wang P.L. et Ideker T. **2011**. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431-432.

Spaulding A.W. et von Dohlen C.D. **2001**. Psyllid endosymbionts exhibit patterns of co-speciation with hosts and destabilizing substitutions in ribosomal RNA. *Insect Molecular Biology*, **10**, 57-67.

Steinert M., Hentschel U. et Hacker J. **2000**. Symbiosis and pathogenesis: evolution of the microbe-host interaction. *Naturwissenschaften*, **87**, 1-11.

Stouthamer R., Breeuwer J.A. et Hurst G.D. **1999**. *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. *Annual Reviews in Microbiology*, **53**, 71-102.

Su Q., Pan H., Liu B., Chu D., Xie W., Wu Q. *et al.* **2013**. Insect symbiont facilitates vector acquisition, retention, and transmission of plant virus. *Scientific Reports*, **3**.

Su Q., Xie W., Wang S., Wu Q., Liu B., Fang Y. *et al.* **2014**. The endosymbiont *Hamiltonella* increases the growth rate of its host *Bemisia tabaci* during periods of nutritional stress. *PLoS One*, **9**, e89002.

Subandiyah S., Nikoh N., Tsuyumu S., Somowiyarjo S. et Fukatsu T. **2000**. Complex endosymbiotic microbiota of the citrus psyllid *Diaphorina citri* (Homoptera: Psylloidea). *Zoological Science*, **17**, 983-989.

Swoboda J.G., Campbell J., Meredith T.C. et Walker S. **2010**. Wall teichoic acid function, biosynthesis, and inhibition. *Chembiochem*, **11**, 35-45.

*T*

Takeuchi Y. **1996**. *Global dynamical properties of Lotka-Volterra systems*. Singapore: World Scientific.

Tamas I., Klasson L., Canbäck B., Näslund A.K., Eriksson A.S., Wernegreen J.J. *et al.* **2002**. 50 million years of genomic stasis in endosymbiotic bacteria. *Science*, **296**, 2376-2379.

Tamm C.O. **1991**. Nitrogen in terrestrial ecosystems: questions of productivity, vegetational changes, and ecosystem stability. *Ecological Studies*, **81**.

Tatusov R.L., Koonin E.V. et Lipman D.J. **1997**. A genomic perspective on protein families. *Science*, **278**, 631-637.



- Taylor T.N. et Osborn J.M. **1996**. The importance of fungi in shaping the paleoecosystem. *Review of Paleobotany and Palynology*, **90**, 249-262.
- Taylor B.L. et Zhulin I.B. **1999**. PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiology and Molecular Biology Reviews*, **63**, 479-506.
- Taylor M.W., Radax R., Steger D. et Wagner M. **2007**. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiology and Molecular Biology Reviews*, **71**, 295-347.
- Telesnicki M.C., Ghersa C.M., Martínez-Ghersa M.A. et Arneodo J.D. **2012**. Identificación molecular del endosimbionte secundario *Hamiltonella defensa* en el pulgón amarillo de los cereales, *Metopolophium dirhodum*. *Revista Argentina de Microbiología*, **44**, 255-258.
- Tenaillon O., Le Nagard H., Godelle B. et Taddei F. **2000**. Mutators and sex in bacteria: conflict between adaptive strategies. *Proceedings of the National Academy of Sciences*, **97**, 10465–10470.
- Teixeira D.D.C., Ayres J., Kitajima E.W., Danet L., Jagoueix-Eveillard S., Saillard C. *et al.* **2005**. First report of a Huanglongbing-like disease of citrus in Sao Paulo State, Brazil and association of a new *Liberibacter* species, “*Candidatus Liberibacter americanus*”, with the disease. *Plant Disease*, **89**, 107-107.
- Teixeira L., Ferreira A. et Ashburner M. **2008**. The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLoS Biology*, **6**, e1000002.
- Thao M.L., Clark M.A., Baumann L., Brennan E.B., Moran N.A. et Baumann P. **2000**. Secondary endosymbionts of psyllids have been acquired multiple times. *Current Microbiology*, **41**, 300-304.
- Thao M.L. et Baumann P. **2004**. Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. *Applied and Environmental Microbiology*, **70**, 3401-3406.
- Thierry M., Becker N., Hajri A., Reynaud B., Lett J.-M. et Delatte H. **2011**. Symbiont diversity and non-random hybridization among indigenous (Ms) and invasive (B) biotypes of *Bemisia tabaci*. *Molecular Ecology*, **20**, 2172–87.
- Thomas G.H., Zucker J., Macdonald S.J., Sorokin A., Goryanin I. et Douglas A.E. **2009**. A fragile metabolic network adapted for cooperation in the symbiotic bacterium *Buchnera aphidicola*. *BMC Systems Biology*, **3**, 24.
- Thompson J.D., Higgins D.G. et Gibson T.J. **1994**. ClustalW. *Nucleic Acids Research*, **22**, 4673-4680.
- Thomson N.R., Howard S., Wren B.W., Holden M.T., Crossman L., Challis G.L. *et al.* **2006**. The complete genome sequence and comparative genome analysis of the high pathogenicity *Yersinia enterocolitica* strain 8081. *PLoS Genetics*, **2**, e206.
- von Tils D., Blädel I., Schmidt M.A. et Heusipp G. **2012**. Type II secretion in *Yersinia*—a secretion system for pathogenicity and environmental fitness. *Frontiers in Cellular and Infection Microbiology*, **2**.
- Timmis J.N., Ayliffe M.A., Huang C.Y. et Martin W. **2004**. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics*, **5**, 123-135.
- Toft C. et Andersson S.G. **2010**. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nature Reviews Genetics*, **11**, 465-475.

Toh H., Weiss B.L., Perkin S.A., Yamashita A., Oshima K., Hattori M. *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.

Tsagkarakou A., Mouton L., Kristoffersen J.B., Dokianakis E., Grispou M. et Bourtzis K. **2012**. Population genetic structure and secondary endosymbionts of Q *Bemisia tabaci* (Hemiptera: Aleyrodidae) from Greece. *Bulletin of Entomological Research*, **102**, 353-365.

Tsuchida T., Koga R., Meng X.Y., Matsumoto T. et Fukatsu T. **2004**. Characterization of a facultative endosymbiotic bacterium of the pea aphid *Acyrtosiphon pisum*. *Microbial Ecology*, **49**, 126-133.

Tsuchida T., Koga R., Meng X.Y., Matsumoto T. et Fukatsu T. **2005**. Characterization of a facultative endosymbiotic bacterium of the pea aphid *Acyrtosiphon pisum*. *Microbial Ecology*, **49**, 126-133.

Tsuchida T., Koga R., Horikawa M., Tsunoda T., Maoka T., Matsumoto S. *et al.* **2010**. Symbiotic bacterium modifies aphid body color. *Science*, **330**, 1102-1104.

U

UniProt Consortium. **2008**. The universal protein resource (UniProt). *Nucleic acids research*, **36**, D190-D195.

V

Vacelet J. **1975**. Étude en microscopie électronique de l'association entre bactéries et spongiaires du genre *Verongia* (Dictyoceratida). *Journal de Microscopie et de Biologie Cellulaire*, **23**, 271-288.

Vallenet D., Labarre L., Rouy Z., Barbe V., Bocs S. et Cruveiller S. **2006**. MaGe: a microbial genome annotation system supported by synteny results. *Nucleic acids research*, **34**, 53-65

Van Dijk T. 2008. *Microbial ecology research trends*. Nova Publishers.

Van Rhijn P. et Vanderleyden J. **1995**. The Rhizobium-plant symbiosis. *Microbiological Reviews*, **59**, 124-142.

Vashishtha A., Sharama K.K. et Lakhanpaul S. **2011**. Co-existence, phylogeny and putative role of *Wolbachia* and yeast-like symbiont (YLS) in *Kerria lacca* (Kerr). *Current Microbiology*, **63**, 206-212.

Vautrin E., Genieys S., Charles S. et Vavre F. **2008**. Do vertically transmitted symbionts co-existing in a single host compete or cooperate? A modelling approach. *Journal of Evolutionary Biology*, **21**, 145-161.

Vautrin E. et Vavre F. **2009**. Interactions between vertically transmitted symbionts: cooperation or conflict? *Trends in Microbiology*, **17**, 95-99.

Vavre F. **2002**. La bactérie *wolbachia*, parasite féministe. *Pour la Science*, **297**, 34-39.



Vogel K.J. et Moran N.A. **2013**. Functional and evolutionary analysis of the genome of an obligate fungal symbiont. *Genome Biology and Evolution*, **5**, 891-904.

Vorburger C. et Gousskov A. **2011**. Only helpful when required: a longevity cost of harbouring defensive symbionts. *Journal of Evolutionary Biology*, **24**, 1611-1617.

Vorwerk S., Martinez-Torres D. et Forneck A. **2007**. *Pantoea agglomerans*-associated bacteria in grape phylloxera (*Daktulosphaira vitifoliae*, Fitch). *Agricultural and Forest Entomology*, **9**, 57-64.

## W

Wallace B., Yang Y.J., Hong J.S. et Lum D. **1990**. Cloning and sequencing of a gene encoding a glutamate and aspartate carrier of *Escherichia coli* K-12. *Journal of Bacteriology*, **172**, 3214-3220.

Walter J. et Ley R. **2011**. The human gut microbiome: ecology and recent evolutionary changes. *Annual Review of Microbiology*, **65**, 411-429.

Wang X.W., Luan J.B., Li J.M., Bao Y.Y., Zhang C.X. et Liu S.S. **2010**. De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics*, **11**, 400.

Wang J., Weiss B.L. et Aksoy S. **2013**. Tsetse fly microbiota: form and function. *Frontiers in Cellular and Infection Microbiology*, **3**, 69.

Weeks A.R., Velten R. et Stouthamer R. **2003**. Incidence of a new sex-ratio-distorting endosymbiotic bacterium among arthropods. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 1857-1865.

Wei S.L. et Young R.E. **1989**. Development of symbiotic bacterial bioluminescence in a nearshore cephalopod, *Euprymna scolopes*. *Marine Biology*, **103**, 541-546.

Weldon S.R., Strand M.R. et Oliver K.M. **2013**. Phage loss and the breakdown of a defensive symbiosis in aphids. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20122103.

Wernegreen J.J. **2005**. For better or worse: genomic consequences of intracellular mutualism and parasitism. *Current Opinion in Genetics & Development*, **15**, 572-583.

Wernegreen J.J. **2012**. Endosymbiosis. *Current Biology*, **22**, R555-R561.

Werren J.H. **1997**. Biology of *Wolbachia*. *Annual Review of Entomology*, **42**, 587-609.

Werren J.H., Baldo L. et Clark M.E. **2008**. *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology*, **6**, 741-751.

West S.A., Kiers E.T., Simms E.L. et Denison R.F. **2002**. Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proceedings of the Royal Society B: Biological Sciences*, **269**, 685-694.

Wicken A.J. et Knox K.W. **1975**. Lipoteichoic acids: a new class of bacterial antigen. *Science*, **187**, 1161-1167.

Wilkinson T.L. et Douglas A.E. **1995**. Why pea aphids (*Acyrtosiphon pisum*) lacking symbiotic bacteria have elevated levels of the amino acid glutamine. *Journal of Insect Physiology*, **41**, 921-927.

Wilks J., Beilinson H. et Golovkina T.V. **2013**. Dual role of commensal bacteria in viral infections. *Immunological Reviews*, **255**, 222-229.

Wilson A.C., Ashton P.D., Calevro F., Charles H., Colella S., Febvay G. **2010**. Genomic insight into the amino acid relations of the pea aphid, *Acyrtosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. *Insect Molecular Biology*, **19**, 249-258.

Woyke T., Tighe D., Mavromatis K., Clum A., Copeland A., Schackwitz W. *et al.* **2010**. One bacterial cell, one complete genome. *PLoS One*, **5**, e10314.

Wrede C., Dreier A., Kokoschka S. et Hoppert M. **2012**. Archaea in symbioses. *Archaea*, **2012**.

Wu M., Sun L.V, Vamathevan J., Riegler M., Deboy R., Brownlie J.C. *et al.* **2004**. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biology*, **2**, 0327-0341.

Wu D., Daugherty S.C., Van Aken S.E., Pai G.H., Watkins K.L., Khouri H. *et al.* **2006**. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biology*, **4**, e188.

4

Young F.E. **1966**. Fractionation and partial characterization of the products of autolysis of cell walls of *Bacillus subtilis*. *Journal of Bacteriology*, **92**, 839-846.

Z

Zabalou S., Riegler M., Theodorakopoulou M., Stauffer C., Savakis C. et Bourtzis K. **2004**. *Wolbachia*-induced cytoplasmic incompatibility as a means for insect pest population control. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 15042-15045.

Zchori-Fein E. et Brown J.K. **2002**. Diversity of prokaryotes associated with *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae). *Annals of the Entomological Society of America*, **95**, 711-718.

Zchori-Fein, E. et Perlman S.J. **2004**. Distribution of the bacterial symbiont *Cardinium* in arthropods. *Molecular Ecology*, **13**, 2009-2016.

Zchori-Fein E. et Bourtzis K. **2011**. *Manipulative tenants: bacteria associated with arthropods*. CRC Press.

Zeh J.A. **2004**. Sexy sons: a dead end for cytoplasmic genes. *Proceedings of the Royal Society B: Biological Sciences*, **271**, S306-S309.

Zientz E., Dandekar T. et Gross R. **2004**. Metabolic interdependence of obligate intracellular bacteria and their insect hosts. *Microbiology and Molecular Biology Reviews*, **68**, 745-770.

Zug R. et Hammerstein P. **2012**. Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PloS One*, **7**, e38544.

# ANNEXES

---

**ANNEXE 1.** Complete genome sequence of “*Candidatus Portiera aleyrodidarum*” BT-QVLC, an obligate symbiont that supplies amino acids and carotenoids to *Bemisia tabaci*.

**ANNEXE 2.** The genome of *Cardinium* cBtQ1 provides insights into genome reduction, symbiont motility and its settlement in *Bemisia tabaci*.

# Complete Genome Sequence of “*Candidatus Portiera aleyrodidarum*” BT-QVLC, an Obligate Symbiont That Supplies Amino Acids and Carotenoids to *Bemisia tabaci*

Diego Santos-García,<sup>a</sup> Pierre-Antoine Farnier,<sup>b</sup> Francisco Beitia,<sup>c</sup> Einat Zchori-Fein,<sup>d</sup> Fabrice Vavre,<sup>b</sup> Laurence Mouton,<sup>b</sup> Andrés Moya,<sup>a,e</sup> Amparo Latorre,<sup>a,e</sup> and Francisco J. Silva<sup>a,e</sup>

Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Valencia, Spain<sup>a</sup>; Université Claude Bernard Lyon 1, Laboratoire de Biométrie et Biologie Evolutive, UMR CNRS 5558, Villeurbanne, France<sup>b</sup>; Instituto Valenciano de Investigaciones Agrarias, Unidad Asociada de Entomología IVIA/CIB-CSIC, Valencia, Spain<sup>c</sup>; Agricultural Research Organization, Department of Entomology, Newe Ya'ar Research Center, Ramat Yishay, Israel<sup>d</sup>; and Unidad Mixta de Investigación en Genómica y Salud (Centro Superior de Investigación en Salud Pública [CSISP]), Valencia, Spain<sup>e</sup>

**The genome of “*Candidatus Portiera aleyrodidarum*,” the primary endosymbiont of the whitefly *Bemisia tabaci* (Mediterranean species), is reported. It presents a reduced genome (357 kb) encoding the capability to synthesize, or participate in the synthesis of, several amino acids and carotenoids, being the first insect endosymbiont capable of supplying carotenoids.**

Whiteflies are small sap-sucking insects belonging to the family Aleyrodidae (2). *Bemisia tabaci*, which is able to feed on many plant species, being one of the worst agricultural pests, has been described as a species complex composed of 24 indistinguishable species (formerly called biotypes) (3). The most important biotypes (based on economic losses) are B (or Middle East-Asia Minor 1 species), which is the most widespread, and Q (or Mediterranean species), which shows more resistance to insecticides (9).

*B. tabaci* harbors a primary (obligate) endosymbiont, “*Candidatus Portiera aleyrodidarum*,” which is located in specialized cells called bacteriocytes (1), and also may contain different secondary (facultative) symbionts (7).

We report here the complete genome of the primary symbiont “*Ca. Portiera aleyrodidarum*” BT-QVLC from *B. tabaci* (Mediterranean species) strain QHC-VLC, which also harbors “*Candidatus Hamiltonella defensa*” (13) and “*Candidatus Cardinium hertigii*” (20).

Enriched bacterial samples were obtained by following a modified Harrison protocol (8) and used for sequencing single-end and 3-kb pair-end libraries using 454 GS-FLX Titanium and a 5-kb mate-pair using HiSeq2000. Filtered reads were used for MIRA *de novo* hybrid assembly, and the software program Gap4 was used for editing and closing the genome (41× combined coverage). The software program Prodigal and the BASYS and RAST servers were used for the initial gene calling, followed by hand curation.

“*Ca. Portiera aleyrodidarum*” has a small chromosome (357,472 bp) with 26.1% G+C content and 68% coding density, values derived from a strong genome reduction process (5, 12, 14), being in the frontier between reduced and extremely reduced symbiotic genomes (10, 11, 16). The genome contains 246 coding genes, 8 pseudogenes, and 38 noncoding RNA genes (33 tRNAs, 3 rRNAs, *rnpB*, and tmRNA). It includes an almost complete information transfer gene repertoire (4) but lacks the genes encoding the aminoacyl-tRNA synthetases for arginine, methionine, threonine, and tryptophan. “*Ca. Portiera aleyrodidarum*” is able to synthesize the essential amino acids threonine and tryptophan and the nonessential serine and lacks the last gene of the phenylalanine, valine, leucine, and isoleucine pathways. The lysine pathway

is also incomplete, lacking the *dapF* and *lysA* genes, with *dapB* pseudogenized. Moreover, the histidine pathway is disrupted, lacking the two last enzymes of the pathway. In addition, for the methionine pathways, it encodes only the last enzyme. These results suggest potential complementations with the insect, or the secondary endosymbionts, to complete the missing steps (6, 18, 19). Interestingly, “*Ca. Portiera aleyrodidarum*” has the ability to produce carotenoids, compounds that insects usually obtain from the diet. They are a source for insect pigments and have a canonical antioxidant action. Recently, carotenoids have been proposed, using light as a source to generate electrons, to reduce NAD<sup>+</sup> in the pea aphid (17).

The genome of this species from the B biotype has been recently described (15). Gene content in the two genomes is the same, although there are some discrepancies due to differences in annotation (pseudogenes, hypothetical proteins, and noncoding RNA genes). The average nucleotide identity between genes of the two strains is 99.3%.

In summary, we propose “*Ca. Portiera aleyrodidarum*” as a symbiont able not only to supplement the diet of its host but to use carotenoids as a way of protecting against oxidative stresses by supplying it with reducing power.

**Nucleotide sequence accession number.** The genome was deposited in GenBank under accession number CP003835.

## ACKNOWLEDGMENTS

This work was supported by grants BFU2009-12895-CO2-01 from the Spanish Ministry of Science and Innovation and Prometeo 92/2009 from Generalitat Valenciana (Spain) and by EU COST Action FA0701. These results have been achieved within the framework of the 1st call on Mediterranean agriculture carried out by ARIMNet, with funding from MOARD (IL), ANR (FR), INIA (ES), NAGREF-DEMETER (GR), and GDAR (TR). D.S.-G. is a recipient of a contract from Prometeo 92/2009.

Received 20 September 2012 Accepted 25 September 2012

Address correspondence to Francisco J. Silva, francisco.silva@uv.es.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.01793-12

## REFERENCES

- Baumann L, et al. 2004. Sequence analysis of DNA fragments from the genome of the primary endosymbiont of the whitefly *Bemisia tabaci*. *Curr. Microbiol.* **48**:77–81.
- Byrne DN, Bellows TS. 1991. Whitefly biology. *Annu. Rev. Entomol.* **36**:431–457.
- De Barro PJ, Liu SS, Boykin LM, Dinsdale AB. 2011. *Bemisia tabaci*: a statement of species status. *Annu. Rev. Entomol.* **56**:1–19.
- Gil R, Silva FJ, Pereto J, Moya A. 2004. Determination of the core of a minimal bacterial gene set. *Microbiol. Mol. Biol. Rev.* **68**:518–537.
- Gomez-Valero L, Latorre A, Silva FJ. 2004. The evolutionary fate of nonfunctional DNA in the bacterial endosymbiont *Buchnera aphidicola*. *Mol. Biol. Evol.* **21**:2172–2181.
- Gosalbes MJ, Lamelas A, Moya A, Latorre A. 2008. The striking case of tryptophan provision in the cedar aphid *Cinara cedri*. *J. Bacteriol.* **190**:6026–6029.
- Gottlieb Y, et al. 2008. Inherited intracellular ecosystem: symbiotic bacteria share bacteriocytes in whiteflies. *FASEB J.* **22**:2591–2599.
- Harrison CP, Douglas AE, Dixon AFG. 1989. A rapid method to isolate symbiotic bacteria from aphids. *J. Invertebr. Pathol.* **53**:427–428.
- Khasdan V, et al. 2005. DNA markers for identifying biotypes B and Q of *Bemisia tabaci* (Hemiptera: Aleyrodidae) and studying population dynamics. *Bull. Entomol. Res.* **95**:605–613.
- McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**:13–26.
- Moya A, Pereto J, Gil R, Latorre A. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat. Rev. Genet.* **9**:218–229.
- Perez-Brocal V, et al. 2006. A small microbial genome: the end of a long symbiotic relationship? *Science* **314**:312–313.
- Rao Q, et al. 2012. Draft genome sequence of “Candidatus Hamiltonella defensa,” an endosymbiont of the whitefly *Bemisia tabaci*. *J. Bacteriol.* **194**:3558.
- Silva FJ, Latorre A, Moya A. 2001. Genome size reduction through multiple events of gene disintegration in *Buchnera* APS. *Trends Genet.* **17**:615–618.
- Sloan DB, Moran NA. Endosymbiotic bacteria as a source of carotenoids in whiteflies. *Biol. Lett.*, in press.
- Tamames J, et al. 2007. The frontier between cell and organelle: genome analysis of *Candidatus Carsonella ruddii*. *BMC Evol. Biol.* **7**:181. doi:10.1186/1471-2148-7-181.
- Valmalette JC, et al. 2012. Light-induced electron transfer and ATP synthesis in a carotene synthesizing insect. *Sci. Rep.* **2**:579. doi:10.1038/srep00579.
- Wilson AC, et al. 2010. Genomic insight into the amino acid relations of the pea aphid, *Acyrtosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. *Insect Mol. Biol.* **19**(Suppl. 2):249–258.
- Wu D, et al. 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol.* **4**:1079–1092.
- Zchori-Fein E, Perlman SJ, Kelly SE, Katzir N, Hunter MS. 2004. Characterization of a ‘*Bacteroidetes*’ symbiont in *Encarsia* wasps (Hymenoptera: Aphelinidae): proposal of ‘*Candidatus Cardinium hertigii*’. *Int. J. Syst. Evol. Microbiol.* **54**:961–968.

# The Genome of *Cardinium* cBtQ1 Provides Insights into Genome Reduction, Symbiont Motility, and Its Settlement in *Bemisia tabaci*

Diego Santos-Garcia<sup>1</sup>, Pierre-Antoine Rollat-Farnier<sup>2,3</sup>, Francisco Beitia<sup>4</sup>, Einat Zchori-Fein<sup>5</sup>, Fabrice Vavre<sup>2,3</sup>, Laurence Mouton<sup>2</sup>, Andrés Moya<sup>1,6</sup>, Amparo Latorre<sup>1,6,\*</sup>, and Francisco J. Silva<sup>1,6,\*</sup>

<sup>1</sup>Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Spain

<sup>2</sup>Université de Lyon, Université Lyon1, Laboratoire de Biométrie et Biologie Evolutive, UMR CNRS 5558, Villeurbanne, France

<sup>3</sup>BAMBOO Research team, INRIA Grenoble, Rhône-Alpes, France

<sup>4</sup>Instituto Valenciano de Investigaciones Agrarias, Unidad Asociada de Entomología IVIA/CIB-CSIC, Valencia, Spain

<sup>5</sup>Department of Entomology, Newe Ya'ar Research Center, Agricultural Research Organization, Ramat Yishay, Israel

<sup>6</sup>Unidad Mixta de Investigación en Genómica y Salud (FISABIO-Salud Pública and Universitat de València), Valencia, Spain

\*Corresponding author: E-mail: amparo.latorre@uv.es; francisco.silva@uv.es.

Accepted: April 4, 2014

**Data deposition:** *Cardinium* cBtQ1 genome has been deposited at GenBank/EMBL/DDBJ databases under the accession CBQZ010000001–CBQZ010000011 for the chromosomal contigs and HG422566 for the plasmid. Cytochrome oxidase I (COI) sequences from *Bemisia tabaci* and 16S rRNA sequences from *Cardinium* sp. derived from this work have been deposited at GenBank/EMBL/DDBJ under accession HG421085–HG421096 and HG421077–HG421084, respectively.

## Abstract

Many insects harbor inherited bacterial endosymbionts. Although some of them are not strictly essential and are considered facultative, they can be a key to host survival under specific environmental conditions, such as parasitoid attacks, climate changes, or insecticide pressures. The whitefly *Bemisia tabaci* is at the top of the list of organisms inflicting agricultural damage and outbreaks, and changes in its distribution may be associated to global warming. In this work, we have sequenced and analyzed the genome of *Cardinium* cBtQ1, a facultative bacterial endosymbiont of *B. tabaci* and propose that it belongs to a new taxonomic family, which also includes *Candidatus Amoebophilus asiaticus* and *Cardinium* cEper1, endosymbionts of amoeba and wasps, respectively. Reconstruction of their last common ancestors' gene contents revealed an initial massive gene loss from the free-living ancestor. This was followed in *Cardinium* by smaller losses, associated with settlement in arthropods. Some of these losses, affecting cofactor and amino acid biosynthetic encoding genes, took place in *Cardinium* cBtQ1 after its divergence from the *Cardinium* cEper1 lineage and were related to its settlement in the whitefly and its endosymbionts. Furthermore, the *Cardinium* cBtQ1 genome displays a large proportion of transposable elements, which have recently inactivated genes and produced chromosomal rearrangements. The genome also contains a chromosomal duplication and a multicopy plasmid, which harbors several genes putatively associated with gliding motility, as well as two other genes encoding proteins with potential insecticidal activity. As gene amplification is very rare in endosymbionts, an important function of these genes cannot be ruled out.

**Key words:** Amoebophilaceae, IS elements, gliding motility, *Candidatus Cardinium hertigii*, host–symbiont interaction.

## Introduction

The sweet potato whitefly *Bemisia tabaci* (Aleyrodidae) is an important polyphagous agricultural pest (Byrne and Bellows 1991). It feeds on plant phloem, causing many economic losses, affecting the plants directly and/or indirectly by transmitting several viruses. *Bemisia tabaci* was originally described

as a complex species comprised by many biotypes (Brown et al. 1995), but recently it has been suggested that it is actually a complex of at least 24 morphologically indistinguishable species (De Barro et al. 2011), although classification in biotypes still persists. Among the most important biotypes are B (Middle East-Asia Minor 1 species), which arose in the



Middle East but now has a cosmopolitan distribution, and Q (Mediterranean species), which originated in the Mediterranean basin and is also widely distributed at present. A phylogenetic analysis of mitochondrial biotype Q individuals worldwide has revealed the existence of three subclades (Q1–Q3) (Gueguen et al. 2010).

Similar to other phloem-feeding insects, which need essential amino acids and other nutrients due to their unbalanced diet, whiteflies have established mutualistic relationships with bacterial endosymbionts that provide essential compounds (Baumann 2005; Moran et al. 2008; Moya et al. 2008). All whiteflies species harbor the obligate, or primary endosymbiont, “*Candidatus Portiera aleyrodidarum*” (Gammaproteobacteria; hereafter *Portiera*) (Thao and Baumann 2004), which is restricted to specialized insect cells called bacteriocytes. In *B. tabaci*, several other bacterial species (secondary or facultative symbionts) may coexist with *Portiera* (Gottlieb et al. 2008). The most frequently observed are *Arsenophonus* sp. and “*Candidatus Hamiltonella defensa*” (hereafter *Hamiltonella*) (both Gammaproteobacteria), which share bacteriocytes with *Portiera*. Less frequently present are *Rickettsia* sp., *Wolbachia* sp. (both Alphaproteobacteria) and “*Candidatus Cardinium* sp.” (Bacteroidetes), detected in bacteriocytes but also in other insect tissues.

Symbiosis between whiteflies and *Portiera* has been shaped through a long-term relationship, mainly characterized by *Portiera* providing essential nutrients (amino acids and carotenoids) to the host, thus compensating for its deficient diet, as revealed by genome sequencing (Santos-Garcia et al. 2012; Sloan and Moran 2012). However, the effects of facultative endosymbionts need to be determined. So far, it has been noted that secondary symbionts in other insects can have different effects, such as reproductive manipulation, nutritional contribution, temperature tolerance, and defence against pathogens and parasitoids (Feldhaar 2011; White et al. 2011).

“*Ca. Cardinium hertigii*” (hereafter *C. hertigii*) was first characterized in *Encarsia* wasps, which are parasitoids of *B. tabaci*, and it was proposed as the species type (Zchori-Fein et al. 2004). However, in recent years, infections with bacteria belonging to the genus *Cardinium* have been detected not only in whiteflies but also in other insects (armored scale, sharpshooters, and *Culicoides* spp.) and other arthropods (mites, ticks, spiders, and copepods). Nowadays, the infection rate in arthropods has been estimated as close to 7% (Zchori-Fein and Perlman 2004; Gruwell et al. 2009; Nakamura et al. 2009). Based on molecular data (16S rRNA and *gyrB* genes) and the presence of microtubule-like complexes, a morphological feature shared by all known *Cardinium*, the genus has been divided into supergroups and strains, following a nomenclature similar to *Wolbachia* endosymbionts, with four described supergroups (A, B, C, and D) (Lo et al. 2002; Nakamura et al. 2009; Edlund et al. 2012). Similarly, in several arthropod taxa, *C. hertigii* has been described as a reproductive manipulator through diverse effects such as feminization,

cytoplasmic incompatibility, and induction of parthenogenesis (White et al. 2011). However, these effects have not been found in other species (e.g., *B. tabaci*), suggesting that *C. hertigii* might also be a mutualistic endosymbiont. This aforesaid claim has been supported by the recently released genome of *Cardinium* cEper1 (endosymbiont of the wasp *Encarsia pergandiella*), which encodes a complete biotin biosynthetic pathway, suggesting a potential role in wasp nutrition (Penz et al. 2012). The 16S rRNA gene sequences from several *Cardinium* endosymbionts of *B. tabaci* reported to date are very similar (>99%) to those of the species type *C. hertigii*, suggesting that they are strains of the same species. Several analyses of secondary endosymbiont coinfections revealed that *B. tabaci* individuals infected only with *C. hertigii* are very unusual, especially for the C1 strain (supergroup A), the most widespread *Cardinium* among whiteflies (Gueguen et al. 2010). Generally, the coexistence with *Hamiltonella* is the most frequently observed, although combinations with other secondary symbionts, such as *Wolbachia* sp. or *Rickettsia* sp., are also possible (Gueguen et al. 2010; Skaljác et al. 2010; Park et al. 2012).

The laboratory strain *B. tabaci* QHC-VLC, belonging to the Q1 subclade, harbors *Cardinium* cBtQ1, which belongs to the C1 strain according to its 16S rRNA gene. This strain coexists within bacteriocytes harboring *Portiera* and *Hamiltonella* and can also be found scattered in different tissues of the whitefly (Gottlieb et al. 2008), similar to other hosts infected by *C. hertigii* (Bigliardi et al. 2006; Kitajima et al. 2007; Nakamura et al. 2009). In contrast, *C. hertigii* strains from different *Encarsia* species have only been detected in the ovaries and accessory cells (Zchori-Fein et al. 2004; Penz et al. 2012).

In this work, the genome of *Cardinium* cBtQ1 endosymbiont of *B. tabaci* QHC-VLC was sequenced and compared with that of the cEper1 strain and with several other Bacteroidetes, including the parasitic amoeba endosymbiont “*Candidatus Amoebophilus asiaticus*” (hereafter referred to as *A. asiaticus*) (Horn et al. 2001; Schmitz-Esser et al. 2010). Overall, all the analysis indicated that the *Cardinium* cBtQ1 genome has undergone changes to facilitate its recent establishment in *B. tabaci*.

## Materials and Methods

### Genome Assembly and Annotation

Enriched bacterial samples (Harrison et al. 1989) were collected from approximately 40,000 *B. tabaci* strain QHC-VLC adult whiteflies. DNA was extracted with the JetFlex Genomic DNA purification kit following the manufacturer’s instructions (Genomed). DNA was pyrosequenced using Roche 454 GS FLX Titanium single-end (shotgun) and paired-end (3 kb) libraries, and an Illumina HiSeq2000 mate-pair library (5 kb). The genome assembly was complex due to the high number of



repetitive sequences, around 14% of the genome. This includes the presence of long duplications, almost 100% identical, and a large number of IS (Insertion Sequence elements). Duplicated regions were detected by abnormal coverage peaks, the differential genomic context (different presence of IS elements, genes, nonrepetitive intergenic regions), and paired-ends/mate-paired connections. The ends of most contigs were incomplete IS elements of different types. Because the contigs ending in the same IS type usually did not overlap, we expected a slight underestimation of the percentage of repetitive elements of *Cardinium* cBtQ1 genome. See [supplementary materials and methods, Supplementary Material](#) online ([supplementary file S1, Supplementary Material](#) online) for a detailed description of the assembling and annotation procedure, as well as the software used.

TBLASTX was used to compare the gene content between plasmids of *Cardinium* cBtQ1 and cEper1, and BLAST hits and gene order were plotted with the genoPlotR package (Guy et al. 2010) from R software (R Development Core Team 2011).

### Phylogenetic and Phylogenomic Analyses

High-quality sequenced 16S rRNAs (including fragments of more than 900 bp) were downloaded from SILVA (Quast et al. 2013) and National Center for Biotechnology Information (NCBI) databases. ssu-aligner was employed for the alignment, because it takes into account the secondary structure (based on covariance models) of the 16S rRNA genes (Nawrocki 2009). Predefined masking was selected to ensure reproducibility in future alignments, and finally, the alignment was refined with Gblocks (Castresana 2000) allowing 50% in coverage gaps. General time reversible, with estimates of invariant sites and gamma-distribution among-site rate variation (GTR+I+G), was selected as the best model using jModeltest2 (Darriba et al. 2012). Additionally, amino acid sequences for the *gyrB* gene were downloaded from the NCBI database, aligned with MAFFT using the L-INS-i algorithm (Katoh et al. 2002) and refined with Gblocks. ProtTest3 (Darriba et al. 2011) gave the improved general amino-acid replacement matrix, with gamma distributed rates across sites (LG+G) as the best evolutionary model. All accession numbers are supplied in the [supplementary table S1, Supplementary Material](#) online.

For phylogenomic reconstructions, 37 orthologous single copy genes related to the translation/transduction machinery and protein folding functions that were present in 61 Bacteroidetes genomes (with the exception of *Candidatus* Sulcia mulleri CARI that lacked 9 of these genes) and a non-Bacteroidetes species, used as outgroup, were selected using the homology search tool from the Microbial Genome Database (Uchiyama 2003) ([supplementary table S1, Supplementary Material](#) online). Encoding proteins were downloaded and aligned with MAFFT v6.717b (L-INS-i) and

concatenated. The selected best model was LG+G+F based on ProtTest3 results.

The first 15 BLASTP hits using the encoded proteins of *Cardinium* cBtQ1 *cgl* gene (*CHV\_c0068*) and the *Leishmania major* genes *cgl* (cystathionine gamma-lyase, LmjF35.3230), *cbl* (cystathionine beta-lyase, LmjF32.2640), and *cgs* (cystathionine gamma-synthase, LmjF14.0460) were selected, and their amino acid sequences downloaded. In addition, the first 15 BLASTP hits against Bacteroidetes for the cystathionine beta-lyase protein (*metC*) from *Escherichia coli* str. K-12 (NP\_417481.1) were also selected. Amino acid sequences were aligned with MAFFT v6.717b (L-INS-i), and ProtTest3 was used to select the appropriate model for the alignment, in this case, the LG+G model. For plasmid TraG protein, the first 50 BLASTP best hits were downloaded (including *Cardinium* cEper1 and *A. asiaticus*) and aligned with MAFFT v6.717b (L-INS-i). For TraG, the best model selected with ProtTest3 was LG+G+I.

RaxML (Stamatakis 2006) was used to calculate maximum likelihood phylogenetic trees for all the alignments, using optimizations for branch lengths and model parameters, and 1,000 rapid bootstrap replicates. Models were adjusted for each case. In addition, PhyloBayes3 (Lartillot et al. 2009) was used to infer Bayesian posterior distributions for each phylogenetic tree. In each case, evolutionary model was adjusted to the model selected (described above), and three independent chains were run for each alignment. Following Lartillot et al (2009) recommendations, all effective sizes were greater than 200 and maximum discrepancy between chains was less than 0.1. Finally, a majority rule consensus tree was calculated for each alignment.

### Genomic Redundancy and Mobile Elements

NUCmer from MUMmer 3 (Kurtz et al. 2004) was used to plot repetitive regions of *A. asiaticus* (AmAs), *Cardinium* cBtQ1, and *Cardinium* cEper1, using each genome as query and subject. Results were filtered and only sequences with at least 95% identity and 500 bp length were used. To estimate the level of redundancy in the genomes, a BLASTN approach using each genome (chromosome plus plasmid concatenated) as query and subject was performed with e-value cutoff of  $1e^{-20}$ . The BLASTN results were transferred to a spreadsheet, where any alignment with an identity smaller than 95% was removed. Lines were sorted through several steps to identify the single copy segments of the genome. The repetitive fraction was estimated subtracting the summation of the single copy segments from the total genome length. IS elements were detected as described previously (Gil et al. 2008), refined using the web server ISSaga (Varani et al. 2011) and deposited in ISfinder database. Reference copies for each mobile element were used to search with BLASTX against the nonredundant NCBI database ( $1e^{-3}$  e-value cutoff). The 25 best hits for each mobile element were used as the input for MEGAN4

(Huson et al. 2011), and taxonomical assignments were done with the default LCA parameters. The genomes of *A. asiaticus*, *Cardinium* cEper1, and *Cardinium* cBtQ1 (with contigs concatenated in decreasing order, excluding the plasmid) were used to compute nucleotide synteny blocks with progressive Mauve aligner (Darling et al. 2010). *Cardinium* cBtQ1 was set as the reference for gene order and the alignment was plotted with the genoPlotR package.

### Orthologous Gene Identification

After phylogenomic reconstruction, all genomes belonging to the order Cytophagales, including both *Cardinium* strains (cBtQ1 and cEper1) and *A. asiaticus* (AmAs), were used for orthologous gene identification. *Flavobacterium johnsoniae* (Bacteroidetes: Flavobacteriales), whose gliding motility has been broadly studied, was selected as the outgroup for the order Cytophagales. All proteins, including those from plasmids, were used as the input for OrthoMCL (Li et al. 2003). OrthoMCL and COG (Clusters of Orthologous Groups of proteins) (Tatusov et al. 2000) profile assignment pipelines were run as described previously (1.5 as inflation value, 70% of match cutoff, and an e-value cutoff of  $1e^{-5}$ ) (Manzano-Marín et al. 2012). Gene clusters may contain zero, one, two, or more genes in each genome. Some gene clusters were manually refined because OrthoMCL failed to recognize some orthologous genes in endosymbionts.

Orthologous genes for *A. asiaticus* and both *Cardinium* strains were classified as core genome (genes shared by the three genomes), genes shared by two of the three organisms and strain-specific genes (supplementary table S2, Supplementary Material online). Euler diagrams were obtained with gplots package (Warnes et al. 2013) from R.

### Last Common Ancestor Reconstruction

All genomes used in the OrthoMCL clustering method were used to reconstruct the putative last common ancestor (LCA) gene contents for each node in the phylogeny (fig. 1 and supplementary table S3, Supplementary Material online). The MPR (most parsimonious reconstruction) function in ape package (Paradis et al. 2004) from R was used to infer the ancestral state for each character (gene clusters) in each node (LCA). Pseudogenes for all genomes were manually revised. For orthologous cluster assignments of the pseudogenes, TBLASTX was used (e value of  $1e^{-5}$  and an overlap of 80% query subject) against the proteins present in the orthologous clusters. Pseudogenes that did not modify the LCA reconstruction (strain-specific genes) were not considered. Pseudogenes that were mobile elements were also excluded. Parsimony reconstruction for orthologous groups that included the previously selected pseudogenes were checked using parsimony reconstruction of discrete characters in Mesquite (Maddison WP and Maddison DR 2011).

For each reconstructed LCA and genome, COG categories were assigned for each orthologous cluster based on the initial OrthoMCL results and the COG assignment described above (Manzano-Marín et al. 2012). For each orthologous cluster, COG categories with less than a 10% of a cluster, as well as the unassigned category, were removed. The LCA, indeterminations (the presence/absence of the gene in the LCA node could not be determined) were counted as half (0.5), instead of presence (1) or absence (0). Relative percentages of each COG type using LCA4 or LCA2 as reference were plotted using the gplots heatmap2 function without dendrograms and reordering functions. Euler diagram was plotted using gplots. COG profiles, stated as the absolute number of COG categories divided by the total number of COG for each genome or LCA, were plotted as a heatmap with gplots allowing hierarchical clustering (dendrograms are grouping the most similar rows or columns together). Cyclobacteraceae family habitats were obtained from GOLD database (supplementary table S3, Supplementary Material online).

### Analyses of *B. tabaci* and *Encarsia* spp. Samples

See supplementary materials and methods and tables S4–S6, Supplementary Material online (supplementary file S1, Supplementary Material online).

### Fluorescent In Situ Hybridization

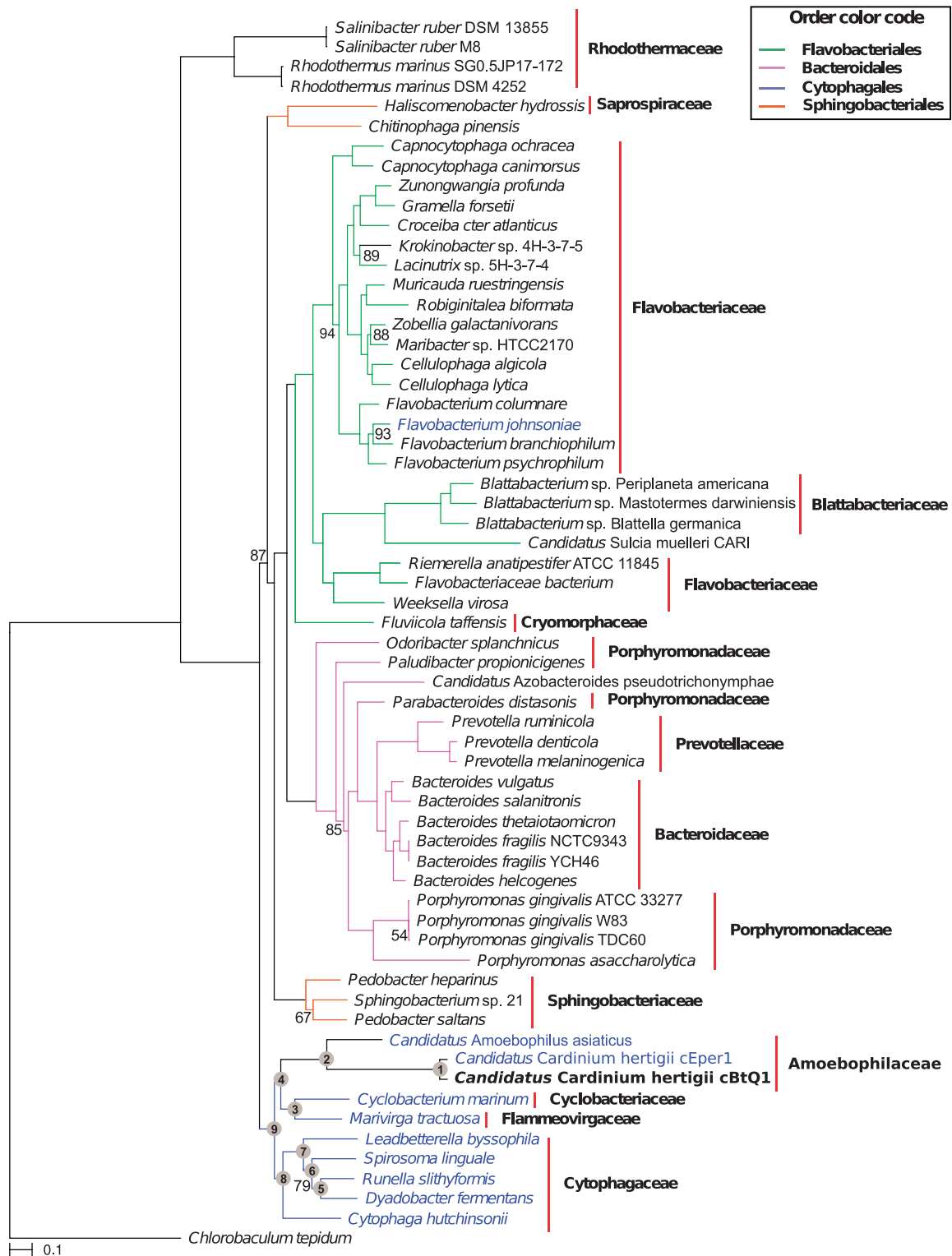
See supplementary materials and methods, Supplementary Material online (supplementary file S1, Supplementary Material online).

## Results

### General Features of *Cardinium* cBtQ1 Genome

*Cardinium* cBtQ1 has a relatively small genome (1.065 Mb) composed of a chromosome (1.013 Mb) and a circular plasmid (52 kb) named pcBtQ1 (table 1). The chromosomal sequence is composed of 11 contigs (612, 80.4, 77.6, 73.8, 66.9, 41.4, 30.3, 13.5, 7.4, 5.1, and 4.1 kb, respectively) with average coverages of 90 $\times$  and 547 $\times$  for 454 and Illumina platforms, respectively, whereas the plasmid is in a single contig with coverages of 595 $\times$  (454) and 4,046 $\times$  (Illumina). Paired-end and mate-pair information confirmed that the plasmid was a single closed circular contig. Also, the higher coverage of the plasmid compared with the chromosomal contigs is indicative of a multicopy plasmid, probably between six and seven copies relative to the chromosome.

We found 709 and 30 coding genes on the chromosome and plasmid, respectively. Many of them were annotated as encoding conserved or hypothetical proteins. We also annotated 156 pseudogenes in the chromosome, 132 derived from transposases and 24 from nontransposase genes (supplementary table S7, Supplementary Material online), and 4 in the plasmid (3 transposases and 1 resolvase). The genome



**FIG. 1.**—Phylogenomic maximum likelihood tree of 61 Bacteroidetes. Phylogenomic reconstruction was done under the LG+G+F model on a concatenated alignment of 37 proteins. *Cardinium* genomes fall in the Cytophagales clade, with *Marivirga tractuosa* and *Cy. marinum* as the closest free-living relatives. *Cardinium* cBtQ1 is displayed in bold. Family names are displayed on the right delimited by a horizontal red line. The genomes used for the LCA reconstruction are shown in blue. Numbers inside gray dots show the LCAs reconstructed in each node. Only maximum likelihood bootstrap values below 95% are displayed. Bayesian posterior probabilities for each node were above 0.95 and are also not displayed. *Chlorobaculum tepidum* was used as outgroup.

**Table 1**General Genomic Features of *Cardinium* Strains and *Amoebophilus asiaticus*

Bacterial Genome	<i>Cardinium</i> cBtQ1 <sup>a</sup>		<i>Cardinium</i> cEper1 <sup>b</sup>		<i>A. asiaticus</i> 5a2
Host	<i>Bemisia tabaci</i>		<i>Encarsia pergandiella</i>		<i>Acanthamoeba</i> spp.
	Chromosome	Plasmid	Chromosome	Plasmid	Chromosome
Contigs	11	1	1	1	1
Size (kb)	1,013	52	887	58	1,884
GC (%)	35	32	36	31	35
CDS	709	30	841	65	1,557
Average CDS length (bp)	1,033	1,389	911	733	990
Coding density (%)	79.7	80.1	85.5	82.1	81.8
rRNAs	3	—	3	—	3
tRNAs	35	—	37	—	35
Other RNA genes	2	—	—	—	—
Pseudogenes (total)	156	4	3	—	222
Pseudogenes (transposase)	132	3	3	—	—
Pseudogene (other CDS)	24	1	—	—	—
Reference	This study		Penz et al. (2012)		Schmitz-Esser et al. (2010)

<sup>a</sup>The chromosome of *Cardinium* cBtQ1 is not closed, but it is considered a high-quality draft genome.<sup>b</sup>The chromosome of *Cardinium* cEper1 contains a single gap not closed due to repetitive elements.

contains one set of rRNA genes distributed in two segments, one including the 16S rRNA and the other the 23S and the 5S rRNA genes. It also contains a set of 35 tRNA genes, which are able to completely decode the mRNA sequences, and two other noncoding RNA genes (*mpB* and *tmRNA*) (table 1).

#### Taxonomic Status of *Cardinium* cBtQ1 Endosymbiont of *B. tabaci* Biotype Q

We performed a phylogenomic reconstruction of a Bacteroidetes phylogeny (see [supplementary table S1, Supplementary Material](#) online, for locus tags in each genome), which showed that both *Cardinium* and *A. asiaticus* formed a well-differentiated clade, related to the families Cyclobacteriaceae and Flammeovirgaceae, with the family Cytophagaceae slightly more distant (fig. 1). This reconstruction was used to select the genomes for subsequent analyses.

Because this phylogenomic reconstruction is consistent with other reported studies (Gupta and Lorenzini 2007; Karlsson et al. 2011), and due to the high bootstrap values obtained, we propose that the *Cardinium/Amoebophilus* clade should be assigned to the order Cytophagales, instead of remaining in the nonclassified Bacteroidetes. Furthermore, the analysis suggests that within the Cytophagales, *Cardinium* forms a new family phylogenetically allied with the Cyclobacteriaceae and Flammeovirgaceae, to which we propose the name Amoebophilaceae (fig. 1).

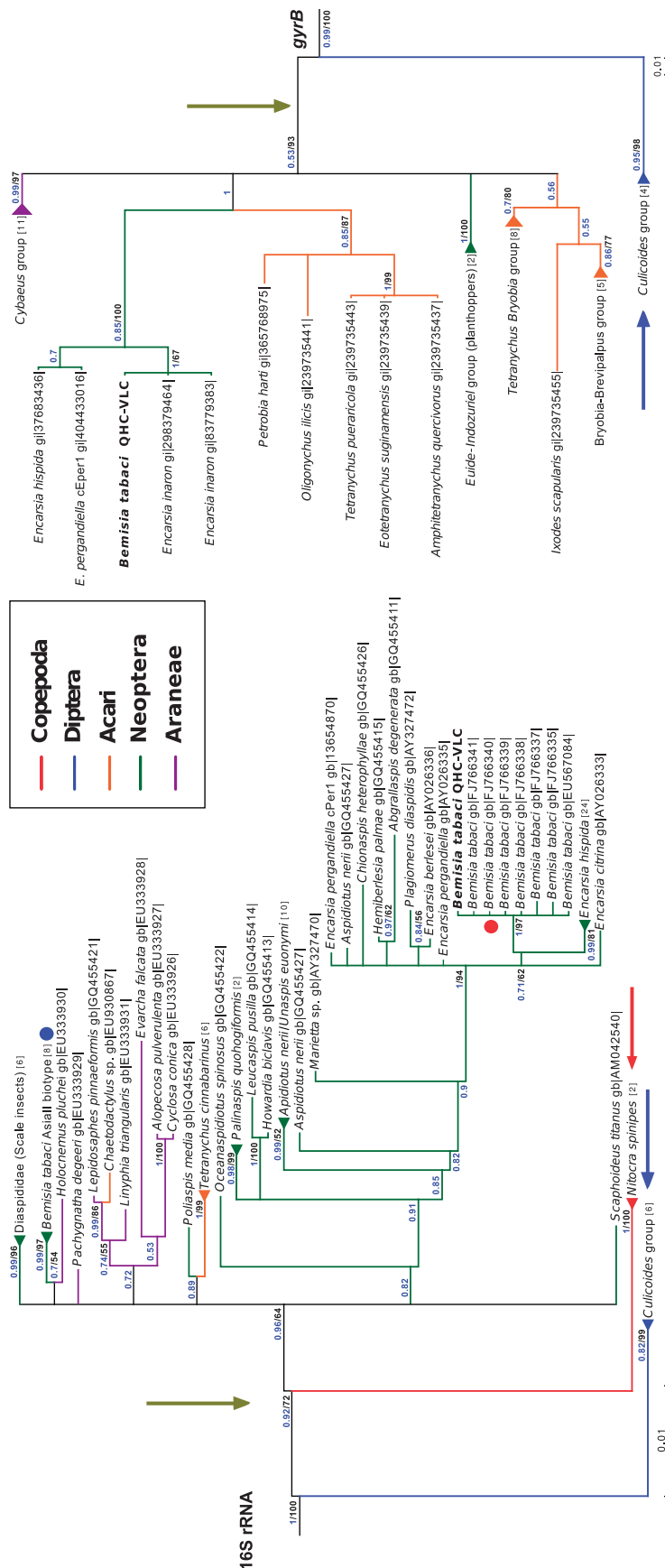
To establish the relationship of *Cardinium* cBtQ1 to other *Cardinium* endosymbionts based on 16S rRNA sequences, a covariance model aligner was employed. The 16S rRNA alignment was used to infer a phylogeny, showing that almost all *Cardinium* endosymbionts of *B. tabaci* (including cBtQ1) were present in a clade with other *Cardinium* endosymbionts of

several *Encarsia* species (99.14% identity with the whole 16SrRNA of *Cardinium* from *E. pergandiella*) (fig. 2, left). A phylogeny with *gyrB* was also performed, which corroborated the close phylogenetic relation with *Cardinium* from *Encarsia* spp. and also showed that *Cardinium* cBtQ1 was embedded in the *Cardinium–Encarsia* clade (fig. 2, right). Because 16S rRNA sequences of *Cardinium* cBtQ1 and the species type *C. hertigii* (symbiont of *E. hispida*) show only 1.2% of differences, we propose that *Cardinium* cBtQ1 is a strain of the latter, in agreement with previous reports (Zchori-Fein and Perlman 2004). The *Bemisia/Encarsia* clade belongs to the *Cardinium* group A, which is well differentiated from the other two groups included in the analysis: Group C, which is specific to the genus *Culicoides* (Nakamura et al. 2009) and group D, which is present in some *Copepoda* spp. (Edlund et al. 2012) (fig. 2).

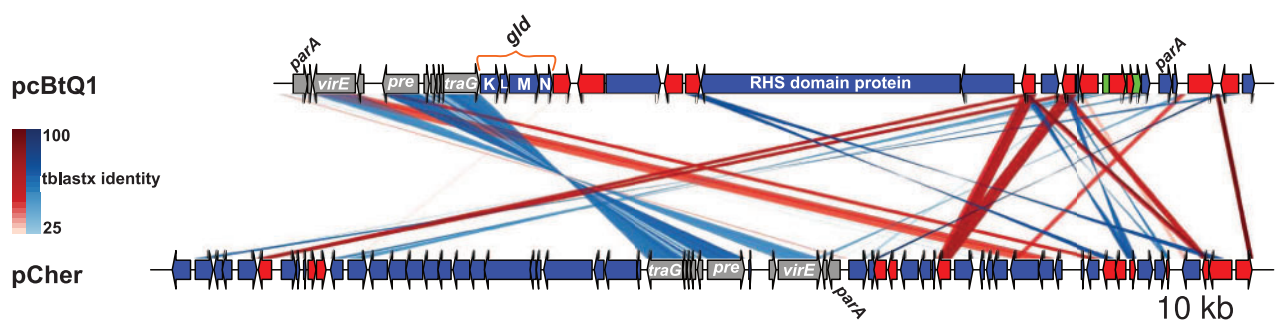
#### Genome Comparison of *Cardinium* Strains and *A. asiaticus*

Compared with *A. asiaticus* (Schmitz-Esser et al. 2010), the genomes of *Cardinium* cEper1 (Penz et al. 2012) and *Cardinium* cBtQ1 (this work) contain a smaller number of coding genes (table 1). However, *Cardinium* cEper1 has almost no annotated pseudogenes, a difference that might reflect gene annotation criteria because open reading frames that belong to transposase fragments are annotated as coding sequences (CDSs). The average gene identity between *Cardinium* cEper1 and cBtQ1 was 92.9%, whereas the average protein identity was 91.8%. The genome fraction assigned to coding genes (labeled as coding density in table 1) was approximately 6% smaller in *Cardinium* cBtQ1 than in *Cardinium* cEper1.





**Fig. 2.**—Bayesian phylogenetic tree for *Cardinium* from different arthropod species. Phylogenetic relationships among different *Cardinium* supergroups are shown. The *Cardinium* group A strains (green vertical arrow) seem the most widespread group, infecting different arthropod orders including the subclass *Neoptera* (which includes whiteflies). Groups C (blue arrow) and D (red arrow) showed a more restricted pattern of hosts. Group C is restricted to the order *Diptera*, whereas Group D is detected on subclass *Copepoda*. On the 16S rRNA phylogeny, *Cardinium* endosymbionts of *Bemisia tabaci* are grouped in two different strains, the C1 (red dot) that belongs to the Mediterranean species (Q biotype) and the C11 (blue dot) that belongs to the Asia II species. The GTR+I+G model was selected for 16S rRNA and the LG+G model for *gyrB*. *Cardinium* strain cBtQ1, endosymbiont of *B. tabaci* QHC-VLC, is displayed in bold type. In both cases, *Amoebophilus asiaticus* was used as the outgroup but was excluded from the figure for plotting reasons. Triangles represent collapsed branches of the same species or genus with the number of collapsed sequences between square brackets. Arthropoda taxonomic names and their respective colors are shown in the upper box. Accession numbers for noncollapsed branches are displayed. Only Bayesian posterior values above 0.5 are displayed (blue), and branches under this score were condensed. Bootstrap values from maximum likelihood phylogenetic reconstruction above 50% are also displayed (black).

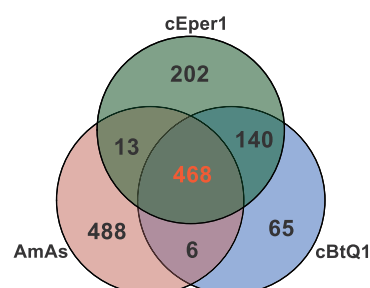


**Fig. 3.**—Comparison of *Cardinium* plasmids. TBLASTX comparison of the plasmids from *Cardinium* cBtQ1 (pcBtQ1) and cEper1 (pCher). Gray arrows are genes included in the syntenic block, blue arrows nontransposase genes, red arrows transposase genes, and the green arrow is a resolvase pseudogene. Red lines show genes in the same orientation and blue lines in reverse orientation. Some gene names are shown in the plot.

*Cardinium* cEper1 also contains a plasmid (pCher) of similar size to pcBtQ1. However, the gene content of both plasmids is different, with only a few shared genes in a syntenic segment showing a high level of nucleotide identity (fig. 3). For example, *CHV\_p004* (*virE*, virulence-associated E family protein), *CHV\_p006* (*pre*, plasmid recombination enzyme), and *CHV\_p011* (*traG*, putative conjugal transfer protein TraG) all show nucleotide identities that range from 76% to 83% between both plasmids. Phylogenetic reconstructions for each of these genes support the close relation between both plasmids (see TraG phylogeny in [supplementary file S1: supplementary fig. S1, Supplementary Material](#) online). Nucleotide, gene order conservation, and phylogeny suggest that both plasmids derive from a common ancestral plasmid present before the split of both *Cardinium* lineages. A putative *traG* gene (*Aasi\_0886*), closely related to *Cardinium traG* phylogenetically ([supplementary fig. S1, Supplementary Material](#) online), is harbored in the *A. asiaticus* chromosome, suggesting that the origin of the plasmid can be traced to the family Amoebophilaceae, with its subsequent chromosomal insertion in *A. asiaticus*.

All three genomes share a core of 468 gene clusters ([supplementary table S2, Supplementary Material](#) online, and [fig. 4](#)), of which six encode putative host interacting proteins (Penz et al. 2012). There are 140 unique gene clusters that were present in both *Cardinium* but not in *A. asiaticus*, 46 of which encode hypothetical proteins, 15 are membrane transport related proteins, and 13 are putative host interacting proteins. Among the remaining *Cardinium* shared genes, there are six coding for transposases, two for phage-derived proteins (or *afp*-like proteins), and five for vitamin biosynthetic proteins ([supplementary table S2, Supplementary Material](#) online).

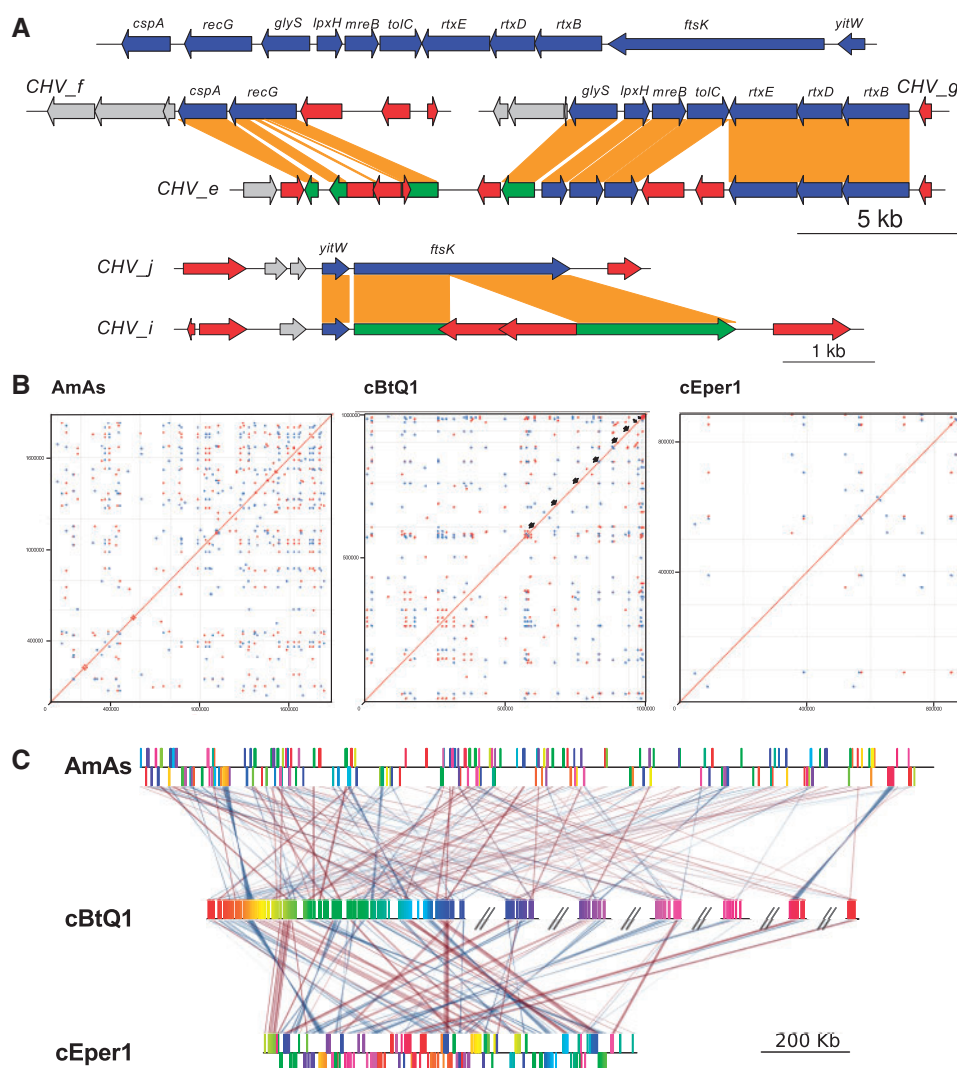
*Cardinium* cEper1 has 202 strain-specific gene clusters, which include, among others, those encoding hypothetical proteins (145), transposases (30), host-interacting proteins (6), and biotin (2) and pyridoxal (1) biosynthetic enzymes. *Cardinium* cEper1 and *A. asiaticus* share 13 gene clusters



**Fig. 4.**—Euler diagram of orthologous clusters. Euler diagram representing the core genome, the strain-specific orthologous clusters, and the orthologous clusters shared by only two organisms. Numbers inside each subspace represent the number of orthologous gene clusters assigned to the subspace. Core genome set is displayed in orange. cBtQ1, *Cardinium* cBtQ1; cEper1, *Cardinium* cEper1; AmAs, *Amoebophilus asiaticus*.

with most of them defined as hypothetical proteins (6), mobile elements (3), a cell-wall-related protein, a membrane protein, and a host-manipulation protein ([supplementary table S2, Supplementary Material](#) online).

*Cardinium* cBtQ1 contains 71 gene clusters, 65 strain-specific, and 6 shared with *A. asiaticus*, which are not present in *Cardinium* cEper1. These gene clusters include ankyrin-domain-containing proteins (14), hypothetical proteins (35), transposases, and other mobile elements (4). A set of very interesting genes is located in the multicopy plasmid of *Cardinium* cBtQ1. They include four gliding genes (*gldK*, *gldL*, *gldM*, and *gldN*, see [fig. 3](#)) that are related to motility in members of the phylum Bacteroidetes (shared with *A. asiaticus*). The fact that the chromosome of *Cardinium* cBtQ1 contains four duplicated genes *rtxBDE* (*A. asiaticus* contained a single copy of the paralogous genes *rtxBE*) and *tolC* related to type I secretion system (T1SS) is also remarkable because only a few sequenced Bacteroidetes harbor secretion system types I, III, IV, or VI (McBride and Zhu 2013). The *rtxBDE* genes, related to the hemolysin secretion proteins (Hly), seem to be



**Fig. 5.**—Redundancy and synteny between *Cardinium* and *Amoebophilus asiaticus* genomes. (A) Putative linear representation of the ancestral genomic region before duplication (on top) and the present state of the two duplications, which are distributed in five contigs (on bottom). Note that *lpxH*, *mreB*, *tolC*, *rtxBDE*, and *yitW* conserve two intact copies, whereas the other duplicated genes only maintain one intact copy. Red arrows are mobile elements, blue arrows genes in the duplicated region, green arrows pseudogenized genes, and gray arrows adjacent genes outside the duplication. Orange bars connect the two duplicated copies of each gene. Contig names are plotted at the beginning or the end of the contig (*CHV\_*), and only regions that contain the duplications are shown. The right ends of contigs *CHV\_g* and *CHV\_e* are connected through paired-end information with the right ends of either contig *CHV\_j* or *CHV\_i*. In both cases, a complete ISCa1 copy, whose fragments are detected at the end of the contigs, is required for joining. (B) Mummer plot showing direct (red) and inverted (blue) genomic repeats with at least 500 bp lengths and 95% similarity. For *A. asiaticus* (AmAs) and *Cardinium* cEper1 (cEper1), inner plot lines denote the division of the chromosome in base pairs sections. Black arrows point contig ends for the largest contigs in *Cardinium* cBtQ1. These contigs were placed in order of decreasing length. Because plots are not scaled to genome size due to limitations of the software, it is noteworthy that the *A. asiaticus* genome is less repetitive than *Cardinium* cBtQ1 although the more compact plot in the former may alter that impression. (C) Common pairwise syntenic blocks of more than 1 kb for *A. asiaticus* (AmAs), *Cardinium* cBtQ1 (cBtQ1), and cEper1 (cEper1). The chromosome of cBtQ1 was taken as reference. Contigs in cBtQ1 are ordered in order of decreasing length and denoted by double slashes. For plotting reasons, only the seven longest cBtQ1 contigs are shown. Red and blue lines show blocks in direct and inverted orientation. The stronger the line, the more nucleotide identity between syntenic blocks.

an event of horizontal gene transfer (HGT), with RTX toxin transport system of *Vibrio* as best BLAST hits. The chromosomal segment involving these genes is duplicated in *Cardinium* cBtQ1 (fig. 5A).

Other interesting *Cardinium* cBtQ1-specific genes, also located in the plasmid, are the putative toxin-related genes *CHV\_p018* and *CHV\_p021*. The latter encodes a long (4,603 amino acids) RHS-repeat-associated core-domain protein with

C-terminus ankyrin repeats. Large proteins with RHS domains have been related with bacterial insecticidal toxins and intercellular signaling proteins (TIGR03696). Although no clear signal peptide was detected in *CHV\_p021*, the presence of ankyrins in the C-terminus domain in combination with a signal peptide has been attributed to protein secretion by T1SS (Kaur et al. 2012). Because the best BLASTX hits, with 63% query coverage and 36% identity (on average), belong to *Daphnia*, *Wolbachia* (HGT event), and mosquitoes, it suggests that the target of this protein can be a conserved protein in arthropods.

The level of redundancy in the genome of *Cardinium* cBtQ1 (~14%) was twice as high as the level found in *Cardinium* cEper1 and *A. asiaticus* (~7% in both cases), most of which is associated with mobile elements (fig. 5B). The mobile elements of *Cardinium* cBtQ1, and their inactive derivatives, account for approximately 166 kb of the chromosome (196 copies) and 12.5 kb of the plasmid (12 copies) (supplementary table S8, Supplementary Material online). From this number of mobile element copies, only 48 contained a functional transposase gene, whereas 135 were transposase pseudogenes. These transposase proteins were classified in 20 different IS families, with only eight being complete IS elements (containing intact transposase genes and inverted repeats at their ends) and could be named according to the ISfinder recommendations and deposited under the names ISCca1-8 (supplementary table S8, Supplementary Material online). Only three mobile element types were specific of the *Cardinium* cBtQ1 (ISCca6, nv\_IS3, and the Retron type one), whereas the rest of transposases were shared with *A. asiaticus*, *Cardinium* cEper1 or both.

It would appear that at least some IS in *Cardinium* are still active, in contrast with *A. asiaticus* (Schmitz-Esser et al. 2011), given we can observe very recent gene duplication events (based on > 99.9% nucleotide identity), with one of the copies interrupted by the insertion of an IS (e.g., pseudogenes *recG* or *ftsK*) (fig. 5A). Another important feature is the presence of a repetitive element composed of a copy of ISCca4 and a copy of nv\_IS2, resulting in an IS that is apparently active. The inactivation of *ftsK* was produced by the insertion of this chimeric IS. Recombination associated to IS may be the cause of the high number of rearrangements in the *Cardinium* cBtQ1, with only some microsyntenic blocks maintained (fig. 5C).

Finally, *Cardinium* cBtQ1 contains a recent chromosome segmental duplication (almost 100% identical) involving at least 11 genes and around 17 kb, distributed in five contigs (fig. 5A). The duplicated region contains the genes *cpsA* (a protease), *recG* (recombination and DNA repair), *glyS* (aminoacyl-tRNA-ligase), *lpxH* (lipid A biosynthesis), *mreB* (actin-like bacterial cytoskeleton), *tolC* (T1SS transmembrane transporter), *rtxB*, *rtxD*, and *rtxE* (T1SS ABC transporters HlyB and HlyD), *ftsK* (cell cycle and chromosome partitioning), and *yitW* (putative chromosome partitioning related function). Although one of the two copies of *cpsA*, *recG*, *glyS*, and

*ftsK* is pseudogenized by IS insertions, the rest of the genes conserve the two functional copies (fig. 5A), indicating that their retention could be advantageous for the organism fitness. However, the fact that the two copies are still active due to their recent duplication cannot be ruled out.

### Evolution of Gene Repertoires in Lineages of *Amoebophilus* and *Cardinium*

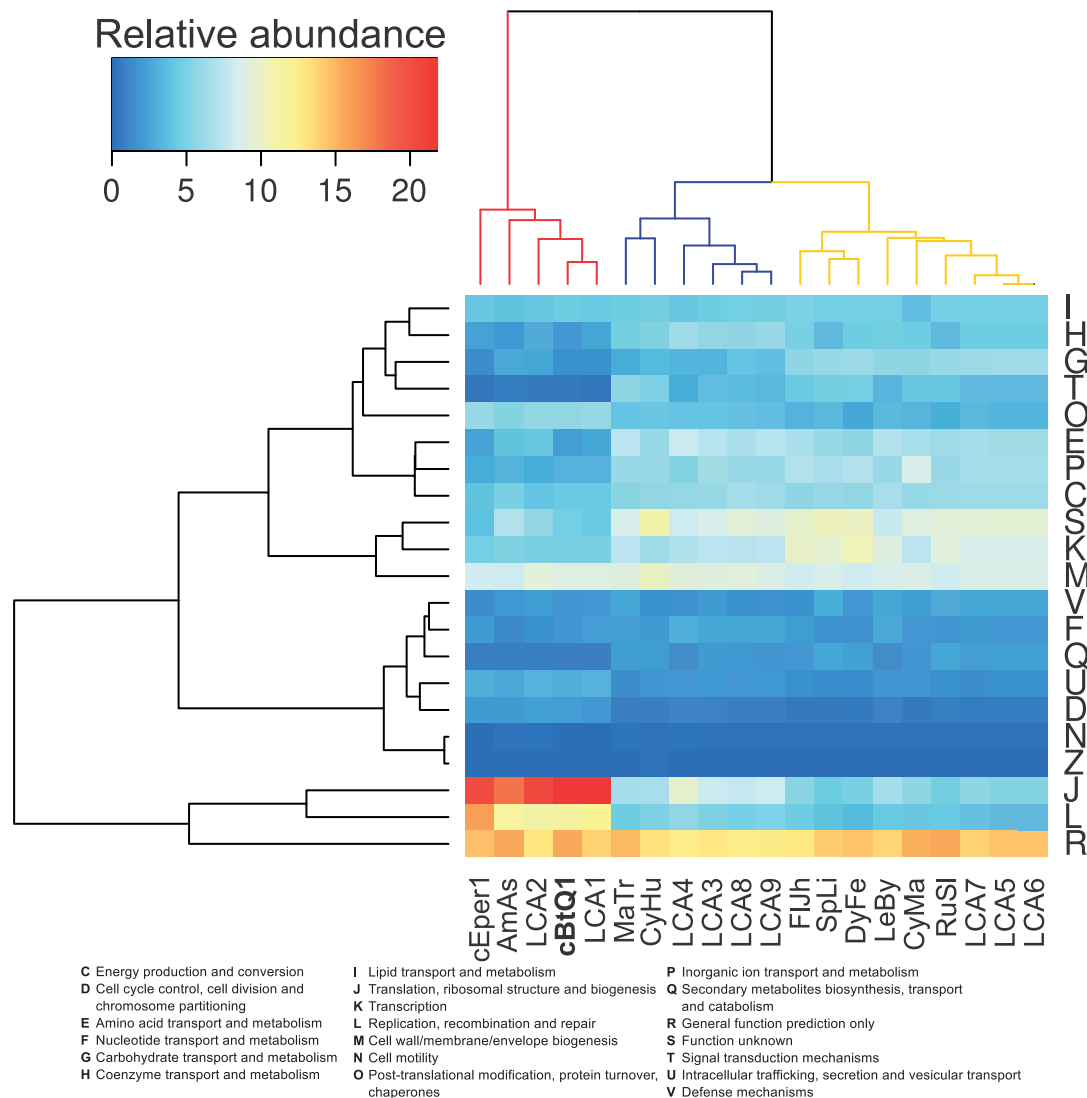
Gene clusters obtained with OrthoMCL were classified according to COG categories and used to reconstruct by maximum parsimony the gene cluster content in the nine LCA corresponding to each node denoted with gray circles (fig. 1). Several analyses were performed to compare the gene content of the present and reconstructed ancestral genomes.

First, hierarchical clustering based on the relative abundance (percentage) of each COG category in each genome was performed (fig. 6). Three main clusters were observed: One that contained the endosymbiotic genomes and the LCA1 and 2; a second that grouped *Marivirgia tractuosa* and *Cyclobacterium marinum* with the LCA3, 4, 8, and 9; and a third that contained the rest of the genomes and LCAs. The second cluster (fig. 6, blue) showed a clear reduction in some COG groups as G and K but an enrichment in the H and J groups when it was compared with the third cluster (fig. 6, yellow). Also, hierarchical clustering grouped both *Cardinium* strains with *A. asiaticus* and LCA1 and LCA2. They showed a stronger conservation of genes in some COG categories such as J and O, when they were compared with the free-living Bacteroidetes, a signal also observed in other symbiotic reduced genomes. LCA4, the ancestor of the *Cardinium/A. asiaticus* lineage and family Cyclobacteriaceae, was close to the free-living Cyclobacteriaceae (see habitats in supplementary table S3, Supplementary Material online), suggesting a similar free-living style. Parsimony reconstruction assigned 1,301 gene clusters to LCA4 and the equally parsimonious presence/absence of other 684 gene clusters.

Second, the transition from LCA4 to LCA2 was examined comparing the percentages of gene clusters in each COG category regarding LCA4 (100%) (fig. 7A). It had a strong impact in the number of gene clusters with more than half of them being lost (LCA2, 655 gene clusters plus 36 present/absent) (supplementary table S3, Supplementary Material online). Relatively to LCA4 (100%), there was a high reduction in all categories except for some housekeeping categories such as J, L, and D with more than 80% of gene clusters conserved. Biosynthetic capabilities of LCA2 were clearly reduced (e.g., COG categories C, E, F, and H).

Third, the transition from LCA2 to LCA1 and to both *Cardinium* and *A. asiaticus* was examined comparing the percentages of gene clusters in each COG category regarding LCA2 (100%) (fig. 7B). The transition to LCA1 (649 gene





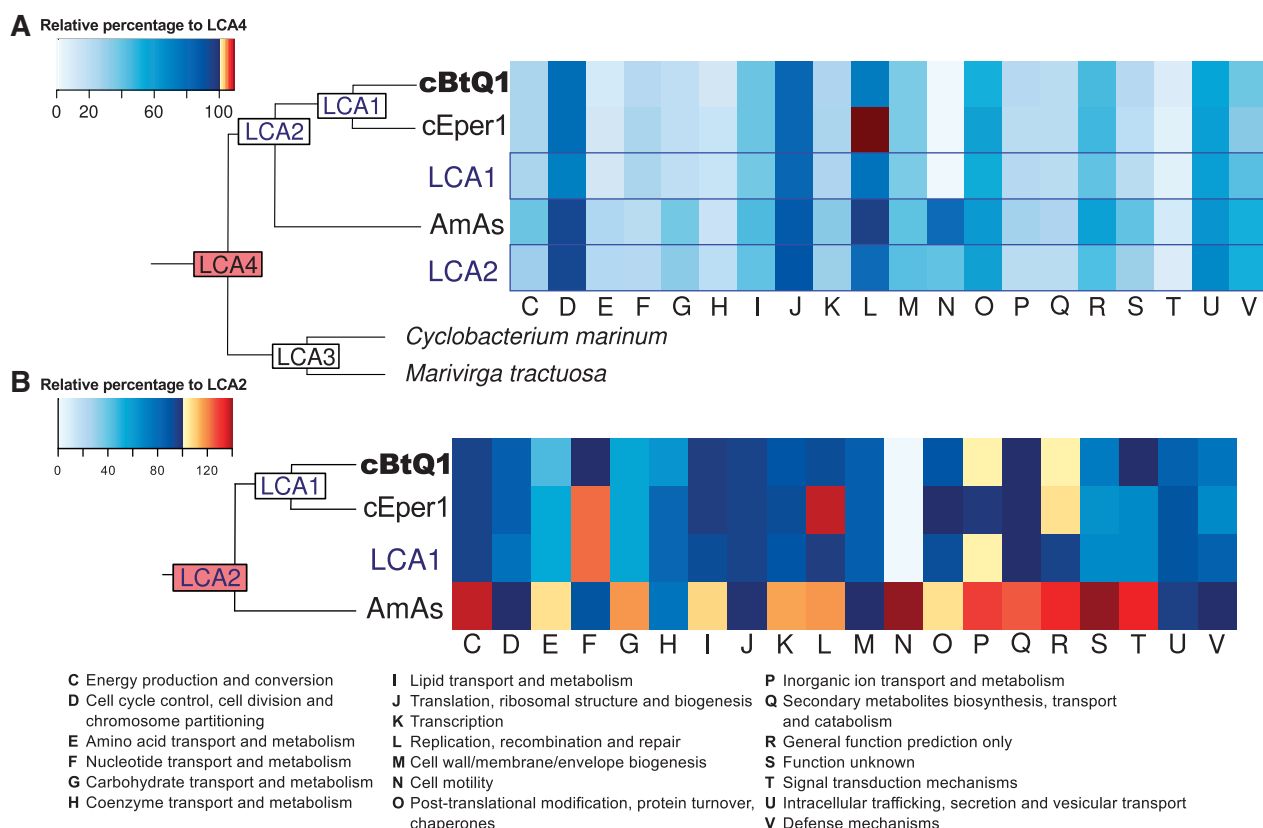
**Fig. 6.**—Relative abundance of gene clusters in Bacteroidetes and LCAs. Hierarchical clustering heatmap representing the relative abundance (percentage) of each COG category in relation to the total number of gene clusters in each genome. Three main COG clusters (left) are observed: Highly retained categories (J, L, and R), medium retained categories (I, H, G, T, O, E, P, C, S, K, and M), and low retained categories (V, F, Q, U, D, N, and Z). Three main species/LCA cluster (up) are cEper1, AmAs, cBtQ1, LCA1, and LCA2 (only symbionts, left cluster); MaTr, CyHu, LCA3, LCA4, LCA8, and LCA9 (middle cluster), and FlJh, SpLi, DyFe, LeBy, CyMa, RuSl, LCA5, LCA6, and LCA7. Species clustering together by COG categories could have similar metabolic features and consequently, a similar ecological niche. *Cardinium* cEper1 (cEper1), *Ca. Amoebophilus* (AmAs), *Cardinium* cBtQ1 (cBtQ1), *Marivirga tractuosa* (MaTr), *Cytophaga hutchinsonii* (CyHu), *Flavobacterium johnsoniae* (FlJh), *Spirosoma linguale* (SpLi), *Dyadobacter fermentans* (DyFe), *Leadbetterella byssophila* (LeBy), *Cyclobacterium marinum* (CyMa), and *Runella slithyformis* (RuSl).

clusters) produced the loss of 160 gene clusters, although 118 new gene clusters were acquired. Comparing the number of gene clusters of LCA2 to LCA1, and to both *Cardinium* and *A. asiaticus*, we observed several differences among COG categories (fig. 7B, supplementary table S3, Supplementary Material online). First, *A. asiaticus* showed 331 strain-specific gene clusters, distributed in several categories, not present in LCA2. Second, the reductive evolution of the *Cardinium* lineage was more clearly observed in several COG categories, such as E, G, H, S, T, and V. The absence of gene clusters in

*Cardinium* for the N category was probably due to the fact that some genes related with motility have not been yet annotated in the COG database, especially those involved in gliding motility (see later) that are, in fact, present in *Cardinium* cBtQ1, *A. asiaticus*, LCA1, LCA2, and LCA4.

#### Biosynthetic Capabilities in *Cardinium* cBtQ1

*Cardinium* cBtQ1, according to KEGG classification pathways, presents low biosynthetic capabilities, similar to those



**Fig. 7.**—Last free-living common ancestor comparison. (A) Heatmap showing the percentage of genes in each COG category, compared with the number of the same category in LCA4 (100%). In left, reduced phylogenomic reconstruction with the name of each LCA reconstructed. (B) The same heatmap type comparing to LCA2 (100%). L category in *Cardinium* cEper1 is an artifact produced by an incorrect annotation of inactivated transposases as CDS instead of pseudogenes. COG definitions are below. cBtQ1, *Cardinium* cBtQ1; cEper1, *Cardinium* cEper1; AmAs, *Amoebophilus asiaticus*.

observed in *Cardinium* cEper1 and *A. asiaticus* (Karlsson et al. 2011; Penz et al. 2012). The main differences between the biosynthetic capabilities of both *Cardinium* strains are in the biosynthesis of vitamins and cofactors. Both bacteria are able to produce lipoate, a key cofactor for intermediate metabolism and an important antioxidant molecule (Spalding and Prigge 2010). *Cardinium* cEper1 has the genes *pdxS* and *pdxT* and can synthesize pyridoxal 5-phosphate (precursor of vitamin B6). However, *pdxT* was pseudogenized by an IS transposition in cBtQ1. This event seems to have happened recently, because the *pdxT* pseudogene is 93.5% identical to the cEper1 gene, a percentage higher than the average gene identity between these two strains.

In addition, it is noteworthy that in *Cardinium* cEper1, the presence of a complete biotin operon, a coenzyme belonging to vitamin B class, is an event of HGT from Alphaproteobacteria to the genus *Cardinium* (Penz et al. 2012). The loss of the ability to synthesize biotin in *Cardinium* cBtQ1 seems to have taken place by the combined effect of the insertion of an IS and a later deletion event, removing the complete *bioB* gene and almost the complete

sequence of the adjacent *bioF* gene (92.5% identical to cEper1 in the remnant segment). Another recent signal of the loss of a nutritional contribution is the pyridoxal-dependent enzyme cystathionine gamma-lyase (involved in the synthesis of cysteine) whose CDS contains an internal stop codon mutation that produces the pseudogene CHV\_c0068 in cBtQ1 (94.9% identical to cEper1 gene). A phylogenetic analysis showed that the functional gene, present in this state in cEper1, was acquired by an ancestor through HGT from a eukaryote, perhaps an amoeba (supplementary file S1: supplementary fig. S2, Supplementary Material online). The phylogenetic analysis, including the three in silico identified cystathionine metabolizing enzymes of *L. major* (Williams et al. 2009), showed its closer relation with *L. major* cystathionine gamma-lyase rather than with *L. major* cystathionine beta-lyase, as previously annotated in *Cardinium* cEper1 (Penz et al. 2012).

In *Cardinium* cBtQ1, the inability to synthesize pyridoxal and biotin may be complemented by other facultative endosymbionts of the host. In the case of *B. tabaci* strain QHC-VLC, such enzyme synthesis could be achieved by *Hamiltonella*,

which is located inside the bacteriocytes (supplementary fig. S3, Supplementary Material online). The complete sets of genes required for the synthesis of both cofactors was described for the *Hamiltonella* secondary endosymbiont of the aphid *Acyrtosiphon pisum* (Degnan et al. 2009), and all of them were also found by BLAST similarity in the draft genome of *Hamiltonella* from *B. tabaci* strain QHC-VLC (unpublished data).

### Gliding Genes in *Cardinium*

As stated earlier, the genome of *Cardinium* cBtQ1 harbors four gliding genes organized in an operon (*gldK*, *gldL*, *gldM*, and *gldN*) in the pcBtQ1 plasmid (fig. 3). To state that the differential tissue locations of *Cardinium* endosymbionts of *B. tabaci* (scattered) and *Encarsia* spp. (restricted to ovaries) were correlated with the presence of the four gliding genes, two populations of *E. pergandiella* (USA and Brazil), one of *E. hispida* (Italy) and one from *E. inaron* (USA) were screened for their presence. None of these species gave a positive result for any of the gliding genes, whereas the presence of *Cardinium* was stated in all of them through the amplification of the control gene *gyrB* (supplementary file S1: supplementary materials and methods and tables S4 and S6, Supplementary Material online).

For *B. tabaci*, adult whiteflies were sampled in 12 points from four localities of the province of Valencia (Spain) (supplementary file S1: supplementary materials and methods and tables S5 and S6, Supplementary Material online). Polymerase chain reaction (PCR) amplification revealed the presence of the four gliding genes (as well as of the large plasmid gene *CHV\_p021*) in all the samples (except one that failed in all the PCR amplifications). To determine insect biotypes, a mitochondrial cytochrome oxidase I (COI) fragment was sequenced in four females per sampling point. Nine of the samples belonged to the biotype Q (Mediterranean species), whereas two were from the biotype S (Sub-Saharan Africa species), an uncommon biotype in Spain (Moya et al. 2001) (supplementary file S1: supplementary fig. S4, Supplementary Material online). All the samples, including those of biotype S, harbored *Cardinium*, which was determined by PCR amplification of a fragment of the 16S rRNA gene. PCR products from biotype S (sample F) and biotype Q (sample B) (1,100 bp) were sequenced and resulted 100% identical to *Cardinium* cBtQ1.

The presence of additional secondary endosymbionts was also checked finding that biotype Q individuals also contained *Hamiltonella*, whereas biotype S contained *Arsenophonus* and *Wolbachia* (supplementary file S1: supplementary materials and methods and tables S5 and S6, Supplementary Material online).

## Discussion

Although the cost of maintaining an obligate mutualistic endosymbiont may be compensated in many cases by its supplementation of the host diet, the maintenance of stable associations with facultative symbionts may require either reproductive manipulation or compensatory benefits from the endosymbiont (Oliver et al. 2010; White et al. 2011; Wernegreen 2012). Both facultative and obligate endosymbionts are transmitted vertically, but facultative symbionts may retain the ability to be horizontally transmitted as revealed by the incongruence of host and symbiont phylogenetic reconstructions (Russell et al. 2003).

Symbionts belonging to the genus *Cardinium* are present in many types of arthropods including arachnids (Nakamura et al. 2009), crustaceans (Edlund et al. 2012), and insects (Zchori-Fein et al. 2004; Zchori-Fein and Perlman 2004; Gruwell et al. 2009; Nakamura et al. 2009), indicating that members of this genus can colonize new niches. Phylogenies and nucleotide conservation of two genes (16S rRNA and *gyrB*) show that *Cardinium* from *B. tabaci* QHC-VLC was closely related to *Cardinium* from armored scale insects and parasitoid wasps, including the species type *C. hertigii* from the wasp *E. hispida* (Zchori-Fein et al. 2004). Researchers have also proposed that *Cardinium* cBtQ1 was a strain of *C. hertigii*, closely related to *Cardinium* cEper1 (symbiont of *E. pergandiella*) whose genome has recently been reported (Penz et al. 2012). In this work, we have sequenced the genome of *Cardinium* cBtQ1, and all the analyses carried out confirm that this strain is closely related to *Cardinium* cEper1, despite being endosymbionts of whiteflies and parasitoids, respectively. There are, however, some differences that could be related to the massive presence of IS in *Cardinium* cBtQ1, as well as to the adaptation to the specific hosts of each strain (see later).

The high number of available Bacteroidetes genomes provided a robust phylogeny (fig. 1), which showed that both *Cardinium* (cEper1 and cBtQ1) and *A. asiaticus* formed a well-defined clade, distant from other family members of Cyclobacteriaceae, Cytophagaceae, and Flammeovirgaceae in the order Cytophagales. Thus, we propose that they form a new family, to be named Amoebophilaceae. This name is proposed because *A. asiaticus* has a larger genome than *Cardinium* spp. and shares more genes with LCA2 than either *Cardinium* strain. Also on the basis of this phylogeny, we were able to infer the coding gene contents (clusters of genes) of the most recent common ancestor in the *Cardinium*/*Amoebophilus* clade (LCA4, LCA2, and LCA1) and to analyze the evolution of the gene repertoires (figs. 6 and 7).

Despite it is not the objective of this work, we can extract general ideas from the cluster analysis (fig. 6). The distribution of genes in COG categories is a well predictor of the way of life of the organisms. Hierarchical clustering indicates that LCA3 to 9 were, similar to free-living Bacteroidetes, able to

occupy different niches (fig. 6). For example, the differences between the abundance of G category in the middle and right clusters could be related to a more restricted source of carbohydrates (niche specialization). It also seems that the increase of the coenzyme metabolism (H) in the middle cluster could be advantageous for the establishment of symbiotic relationships (*Cyclobacterium* was found in the celomic fluid of a sand dollar, [supplementary table S3, Supplementary Material online](#)).

The number of gene clusters estimated in LCA4 was between 1,301 and 1,985, showing a COG profile similar to those of *M. tractuosa* (Flammeovirgaceae) and *Cy. marinum* (Cyclobacteriaceae) (fig. 6), and differing from the COG profiles of LCA2 and LCA1, which clustered with those of the symbiotic bacteria *A. asiaticus* and *Cardinium* (fig. 6). Because species of the family Cyclobacteriaceae are mostly marine free-living bacteria, and associations with animal hosts have been described ([supplementary table S3, Supplementary Material online](#)), we propose that LCA4 was probably a marine free-living bacterium, with a wide range of functional capabilities and the ability to establish symbiotic relationships. Also, it is likely that LCA4 was able to glide because it contained the whole set of gliding genes essential for gliding, including the *sprATE* genes (McBride and Zhu 2013).

The transition from LCA4 to LCA2 was clearly a reductive process that affected almost all COG categories (fig. 7A) producing an ancestral endosymbiont with few biosynthetic capabilities. Considering that the species derived from LCA2 were endosymbionts of amoebas (Horn et al. 2001; Schmitz-Esser et al. 2010) or insects (Zchori-Fein and Perlman 2004; Penz et al. 2012), the most probable reason for this reduction was the transition from a free living to intracellular life style, to start a symbiotic (either mutualistic or parasitic) relationship with a eukaryotic host. During this transition, the number of gene clusters and associated functions was reduced, although LCA2 maintained the ability to acquire new genes by HGT. The higher number of gene clusters in *A. asiaticus* versus LCA2 (298) could be due to this fact, although other reasons, such as the possibility of a biased sample of genomes, or different annotation problems, would also explain its high number of specific gene clusters.

Genome reduction was an ongoing process in the LCA1/*Cardinium* clade, and it was notorious for some categories such as E and G (fig. 7B). Moreover, the comparative analysis of the gene contents of the two *Cardinium* strains revealed that, despite being very similar at nucleotide level (92.9% nucleotide identity), revealing a recent evolutionary divergence, there are some relevant differences between the genomes of both strains, indicating differences in the evolution of endosymbiosis in *Encarsia* spp. and *B. tabaci*.

First, both *Cardinium* contain a plasmid of 50–60 kb with many differences in gene content (fig. 3). However, both plasmids contain a short syntenic block of genes, whose nucleotide content and gene order conservation, as well as the gene

phylogenies carried out, suggests that both derived from an ancestral plasmid present in their LCA (LCA1), and the differences must have been accumulated after the split of both lineages. These differences are due to the insertion of mobile elements, sometimes carrying accessory genes, and to the transfer of genes from the chromosome.

Second, there are also differences in the chromosome of both *Cardinium* strains. The *Cardinium* cBtQ1 chromosome is 126 kb larger (probably a bit larger because, as a draft genome, gaps are not taken into account for this calculation) than that of *Cardinium* cEper1; nevertheless, this difference is not associated with a greater number of genes but to the presence of a large number of pseudogenes, most of them due to defective transposase encoding genes (table 1). The number of repeated sequences in the genome of *Cardinium* cBtQ1 is twice that of *Cardinium* cEper1, and most of these differences are due to a large number of IS in the former genome (fig. 5B). Moreover, some IS types seem to have their origins in Alphaproteobacteria, probably related to the genera *Rickettsia* or *Wolbachia*, supporting the idea of HGT events between secondary endosymbionts harbored by the same host (Toft and Andersson 2010; Schmitz-Esser et al. 2011; Duron 2013). A large number of mobile elements is a typical feature of endosymbionts that have established a recent relationship with their hosts, such as *Candidatus Sodalis pierantonius* str. SOPE (Gil et al. 2008, Oakeson et al. 2014) or *Sodalis glossinidius* (Belda et al. 2010). Also, enrichment in mobile elements seems to be linked to genome plasticity in some facultative symbionts (Gillespie et al. 2012). Several lines of evidence indicate that IS elements ([supplementary table S8, Supplementary Material online](#)) are already active in *Cardinium* cBtQ1. This activity combined with both the high number of copies throughout the genome, and a complete replication and repair machinery that can produce recombination, probably underlies the massive number of rearrangements in the genome of *Cardinium* cBtQ1, compared with *Cardinium* cEper1 and *A. asiaticus* (fig. 5C).

Finally, among the genes with an annotated function and differential presence in the two *Cardinium* strains, we consider some may give clues about the relationship between *Cardinium* cBtQ1 and *B. tabaci*. First, the presence of pseudogenes for *pdxT* (pyridoxal biosynthesis) and *cgl* (cystathionine gamma-lyase) and the loss of two genes of the biotin operon indicate that these genes were present in LCA1. This also indicates that the loss of these functions in *Cardinium* cBtQ1 is associated with settlement in a new environment composed by *B. tabaci* and its facultative symbiont *Hamiltonella defensa*. This symbiont seems to have become established in the populations of *B. tabaci* Q1 from Western Mediterranean based on the detection of frequencies of 100% (or almost 100%) in populations from North Africa and south-western Europe (Gueguen et al. 2010). Second, among 71 gene clusters of *Cardinium* cBtQ1, which are absent in *Cardinium* cEper1, there are up to 14 specific genes encoding ankyrin-domain



proteins, which can interact with the host's machinery, but further studies are needed to understand their functions. Finally, the most interesting genes are those whose expression has been amplified by either gene duplication (*mreB*, *lpxH*, *yitW*, and *rtxBDE/toIC*) or by their presence in a multicopy plasmid (e.g., *CHV\_p018*, a putative toxin-related gene, *CHV\_p021*, a RHS-domain protein, and the gliding genes *gldKLMN*) (supplementary table S2, Supplementary Material online).

On the basis of the ancestor reconstruction analysis, we can predict that the four gliding genes were present in LCA4, LCA2, and LCA1 and were lost in *Cardinium* cEper1. Although LCA4 conserved full gliding machinery, *sprATE* was lost in LCA2 possibly due to its accommodation to an intracellular environment. Because LCA1 conserved the *gldKLMN* operon, this suggests that *gldKLMN* was lost in the *Cardinium* cEper1 lineage. Also, as in the closest Bacteroidetes genomes, such as *A. asiaticus*, *M. tractuosa*, or *Cy. marinum*, these genes are located in the chromosome, and we can postulate that in *Cardinium* cBtQ1, they have been translocated to a multicopy plasmid conserving the operon order. This supports the importance of these genes for *Cardinium* cBtQ1 and suggests that they may explain why the strain is not confined to a single tissue in *B. tabaci* (Gottlieb et al. 2008) (supplementary fig. S3, Supplementary Material online), in opposition to *Cardinium* cEper1 that is restricted to the ovaries of *Encarsia* (Zchori-Fein et al. 2004; Penz et al. 2012). Moreover, the gene amplification in *Cardinium* cBtQ1 not only of the four gliding genes but also of *mreB* and of the cluster *rtxBDE/toIC* suggests that they may play an important role in this organism, as genome reduction is an ongoing process in this strain. There are two possible hypotheses: 1) those genes are involved in gliding as in other genomes; 2) they are involved in the novel type 9 secretion system (PorSS), which is also associated with the secretion of proteins involved in motility and toxins (Sato et al. 2010; McBride and Zhu 2013).

Different Bacteroidetes possess the ability to move by a gliding mechanism, which is related to the ability to degrade some components present in the environment such as chitin and cellulose (Spormann 1999; McBride 2004; Braun et al. 2005). Several examples of gliding have been reported in species of the class Cytophagia where *C. hertigii* was included (Xie et al. 2007; McBride and Zhu 2013), and the scattered pattern detected for the tissue distribution of *Cardinium* cBtQ1 in *B. tabaci* might be caused by a similar mechanism.

The motor model (or focal adhesion) proposed for gliding in myxobacteria (Spormann 1999; Mignot et al. 2007; Jarrell and McBride 2008; Nakane et al. 2013; Nan et al. 2013) could be extended to Bacteroidetes, including *Cardinium* cBtQ1, although in a more simplified manner. This model considers molecular motors that are associated with cytoskeletal filaments and use proton motive force to transmit force through the cell wall to attached dynamic focal adhesion complexes (adhesins) to the substrate, causing the cell to move forward

(Mignot et al. 2007; Sun et al. 2011; Nan et al. 2013). MreB, the homolog of the eukaryotic actin, has been proposed as the cytoskeletal part of the gliding machinery (Kearns 2007; Mauriello et al. 2010). Also, there may be an association with FtsZ, a protein that is part of the bacterium cytoskeleton and can produce force by itself (Erickson et al. 2010). Linkage between the cytoskeleton and gliding is supported by experimental data with the use of compound A22, which is able to affect the MreB structure and inhibits the gliding motility in *Mycobacteria* (Nan et al. 2011). Eleven genes, detected in most Bacteroidetes, have been defined as the core of the gliding machinery. Four of these genes (*gldB*, *gldD*, *gldH*, and *gldJ*) have unknown function, whereas the remaining seven genes (*gldK*, *gldL*, *gldM*, *gldN*, *sprA*, *sprE*, and *sprT*) also encode the PorSS system (McBride and Zhu 2013).

*Cardinium* cBtQ1 does not show the complete gliding machinery as it only contains four gliding core genes (*gldKLMN*). Neither homologous nor potential analogous genes of *gldBDHJ* have been detected. The *sprAET* genes are also absent, but their function would potentially be substituted by the cluster *rtxBDE/toIC*, which is also duplicated. RTX secretion system belongs to the T1SS and is able to transport proteins from the cytosol to the extracellular space in a SecYEG independent manner. Also, T1SS is able to secrete many different RTX family proteins and proteins without the C-terminal RTX nonapeptide (Linhartová et al. 2010; Kaur et al. 2012). The RTX system would secrete the adhesins (or other proteins that could interact with the host) across the bacterial membrane. However, we did not detect any orthologs to known adhesin proteins, but proteins with eukaryotic domains, such as ankyrins, TPR or WH2 (found in this strain), may function as adhesins in a multicellular eukaryotic organism. Moreover, *Cardinium* may be able to manipulate the host cytoskeleton to form a scaffold, which could be used by the gliding machinery (Haglund et al. 2010). Gliding seems to be a widespread direct invasion mechanism for different kinds of cells (Sibley et al. 1998; Furusawa et al. 2003; Sibley et al. 2004), thus *Cardinium* cBtQ1 and other Bacteroidetes might use this motor model system to invade new hosts or host tissues. In fact, *Cardinium* endosymbiont of *Ixodes scapularis* has been cultivated on insect cell lines and is capable of infecting new cells, even cell lines from different insect species, when they are added to the culture (Morimoto et al. 2006).

The second hypothesis would consider that the *gldKLMN* operon is not involved in gliding, but it is just required for secretion forming the PorSS system, which was initially described for *Porphyromonas gingivalis* as a novel secretion system with eight proteins involved (PorK, PorL, PorM, PorN, PorT, PorW, Sov, and PorP) (Sato et al. 2010). Putatively orthologous genes in the gliding system for the first seven are *gldK*, *gldL*, *gldM*, *gldN*, *sprT*, *sprE*, and *sprA*. The proposed orthologous gene for *porP* in *F. johnsoniae* was *Fjoh\_3477*. A similar gene was not detected in either *A. asiaticus* or *Cardinium*. Proteins secreted by the PorSS systems are adhesins, as well as some enzymes

such as chitinases, and gingipains in *F. johnsoniae* and *P. gingivalis*, respectively. Also, proteins secreted by the PorSS secretion system may contain a conserved C-terminal domain (TIGR4131 and 4183) (Sato et al. 2010; McBride and Zhu 2013). However, there were not proteins of *Cardinium* cBtQ1 with this domain. In the PorSS system, the presence of the protein complex GldKLMN is associated with the generation of the energy required for protein secretion by SprTEA. However, these proteins are not encoded in the genome of *Cardinium* cBtQ1 (they are also absent in *A. asiaticus*), and their substitution by the T1SS (RTX system) seems unlikely because T1SS has its own ATP-binding cassette, making the energy production function of the *gldKLMN* unnecessary. This suggests that the PorSS system does not work in *Cardinium* cBtQ1. Thus, we hypothesized that the retention of the necessary genes for the gliding movement (*gldKLMN*) and further amplification by the translocation to a multicopy plasmid, together with the acquisition of the RTX TISS is associated to the scattered phenotype of this strain and the ability to glide.

The frequent presence of *Cardinium* cBtQ1 in *B. tabaci* biotypes Q and S may be due to its ability to benefit its host (Oliver et al. 2008; Feldhaar 2011; Ferrari and Vavre 2011), possibly related to its motility feature, and/or to the presence of some putative toxin-related genes such as *CHV\_p018* (low *e*-value BLAST hit against RTX toxins of *H. defensa* from *A. pisum*) and *CHV\_p021*, which could play a role in intercellular competition, intercellular signaling, and insecticidal activity based on the presence of the RHS domain (Koskiniemi et al. 2013). As a mobile endosymbiont, *Cardinium* cBtQ1 may contact a parasitoid directly and secrete insecticidal toxins near it, could invade the parasitoid tissue, and kill it by secreting unknown toxins or by the cytotoxic effect of lipid A in a nonacclimated host (Furusawa et al. 2003; Caspi-Fluger et al. 2011; Rader et al. 2012). Furthermore, other effects increasing host fitness cannot be excluded, like heat-stress resistance or maybe some advantages conferred by the lipoate supplementation (Moran et al. 2008; Moya et al. 2008). However, fitness may only be improved in some environments or climate conditions.

In conclusion, we have reported the genome of *Cardinium* cBtQ1 endosymbiont of *B. tabaci*. Comparative genomics and ancestors' reconstruction of gene content have shed light on the drastic reduction in many functional categories that have taken place since its free-living ancestor up to the present. The loss of several cofactors and amino acid biosynthetic capabilities, retained in its close relative *Cardinium* cEper1, rules out an important role in host nutrition and suggests a relationship with its establishment in *B. tabaci* and its endosymbionts. The genome is still very dynamic, with many active transposable elements and rearrangements. On the basis of genomic data, we propose that *Cardinium* cBtQ1 has retained the minimal gliding core machinery present in its ancestors, which in combination with the acquired RTX secretion system might be used to move inside its host or invade new hosts.

## Supplementary Material

Supplementary file S1 is available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

F.J.S., A.L., A.M., D.S.-G., E.Z.F., F.V., and L.M. conceived the study. F.J.S., A.L., A.M., and D.S.-G. designed the experiments. D.S.-G. and F.B. performed the experiments. D.S.-G. and F.J.S. generated the data. D.S.-G., F.J.S., and A.L. analyzed the data. D.S.-G., F.J.S., and A.L. wrote the paper. F.J.S., A.L., A.M., D.S.-G., F.B., P.-A.R.-F., E.Z.F., F.V., and L.M. critically revised the paper. This work was supported by grants BFU2012-39816-C02-01 (cofinanced by FEDER funds and Ministerio de Economía y Competitividad, Spain) to A.L., Prometeo/2009/092 (Conselleria d'Educació, Generalitat Valenciana, Spain) to A.M., and EU COST Action FA0701. These results have been achieved within the framework of the 1st call on Mediterranean agriculture carried out by ARIMNet, with funding from MOARD (IL), ANR (FR), INIA (ES), NAGREF-DEMETER (GR), and GDAR (TR). D.S.-G. is a recipient of a contract from Prometeo 92/2009. P.-A.R.-F. is a recipient of a grant from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. [247073]10 SISYPHE. The authors gratefully acknowledge Martha S. Hunter and Massimo Giorgini for the *Encarsia* samples, and Mariana Reyes-Prieto and Alejandro Manzano-Marín for bioinformatic advice. They also acknowledge the SCSIE from the *Universitat de València* for sequencing and microscopy support. Part of this work was performed by F.J.S. during a sabbatical stay at the *Uppsala Universitet*.

## Literature Cited

- Baumann P. 2005. Biology bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annu Rev Microbiol.* 59:155–189.
- Belda E, Moya A, Bentley S, Silva FJ. 2010. Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of *Sodalis glossinidius*, the secondary endosymbiont of tsetse flies. *BMC Genomics* 11:449.
- Bigliardi E, et al. 2006. Ultrastructure of a novel *Cardinium* sp. symbiont in *Scaphoideus titanus* (Hemiptera: Cicadellidae). *Tissue Cell.* 38: 257–261.
- Braun TF, Khubbar MK, Saffarini DA, McBride MJ. 2005. *Flavobacterium johnsoniae* gliding motility genes identified by mariner mutagenesis. *J Bacteriol.* 187:6943–6952.
- Brown JK, Frohlich DR, Rosell RC. 1995. The sweetpotato or silverleaf whiteflies: biotypes of *Bemisia tabaci* or a species complex? *Annu Rev Entomol.* 40:511–534.
- Byrne DN, Bellows TS. 1991. Whitefly biology. *Annu Rev Entomol.* 36: 431–457.
- Caspi-Fluger A, et al. 2011. *Rickettsia* “in” and “out”: two different localization patterns of a bacterial symbiont in the same insect species. *PLoS One* 6:e21096.
- Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 17: 540–552.

- Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5: e11147.
- Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27:1164–1165.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*. 9:772.
- De Barro PJ, Liu S-S, Boykin LM, Dinsdale AB. 2011. *Bemisia tabaci*: a statement of species status. *Annu Rev Entomol*. 56:1–19.
- Degnan PH, Yu Y, Sisneros N, Wing RA, Moran NA. 2009. *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proc Natl Acad Sci U S A*. 106:9063–9068.
- Duron O. 2013. Lateral transfers of insertion sequences between *Wolbachia*, *Cardinium* and *Rickettsia* bacterial endosymbionts. *Heredity (Edinb)* 2:1–8.
- Edlund A, Ek K, Breitholtz M, Gorokhova E. 2012. Antibiotic-induced change of bacterial communities associated with the copepod *Nitocra spinipes*. *PLoS One* 7:e33107.
- Erickson HP, Anderson DE, Osawa M. 2010. FtsZ in bacterial cytokinesis: cytoskeleton and force generator all in one. *Microbiol Mol Biol Rev*. 74: 504–528.
- Feldhaar H. 2011. Bacterial symbionts as mediators of ecologically important traits of insect hosts. *Ecol Entomol*. 36:533–543.
- Ferrari J, Vavre F. 2011. Bacterial symbionts in insects or the story of communities affecting communities. *Philos Trans R Soc Lond B Biol Sci*. 366:1389–1400.
- Furusawa G, Yoshikawa T, Yasuda A, Sakata T. 2003. Algicidal activity and gliding motility of *Saprospira* sp. SS98-5. *Can J Microbiol*. 49:92–100.
- Gil R, et al. 2008. Massive presence of insertion sequences in the genome of SOPE, the primary endosymbiont of the rice weevil *Sitophilus oryzae*. *Int Microbiol*. 11:41–48.
- Gillespie JJ, et al. 2012. A *Rickettsia* genome overrun by mobile genetic elements provides insight into the acquisition of genes characteristic of an obligate intracellular lifestyle. *J Bacteriol*. 194:376–394.
- Gottlieb Y, et al. 2008. Inherited intracellular ecosystem: symbiotic bacteria share bacteriocytes in whiteflies. *FASEB J*. 22:2591–2599.
- Gruwell ME, Wu J, Normark BB. 2009. Diversity and phylogeny of *Cardinium* (Bacteroidetes) in armored scale insects (Hemiptera: Diaspididae). *Ann Entomol Soc Am*. 102:1050–1061.
- Gueguen G, et al. 2010. Endosymbiont metacommunities, mtDNA diversity and the evolution of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. *Mol Ecol*. 19:4365–4378.
- Gupta RS, Lorenzini E. 2007. Phylogeny and molecular signatures (conserved proteins and indels) that are specific for the Bacteroidetes and Chlorobi species. *BMC Evol Biol*. 7:71.
- Guy L, Kultima JR, Andersson SGE. 2010. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* 26:2334–2335.
- Haglund CM, Choe JE, Skau CT, Kovar DR, Welch MD. 2010. *Rickettsia* Sca2 is a bacterial formin-like mediator of actin-based motility. *Nat Cell Biol*. 12:1057–1063.
- Harrison CP, Douglas AE, Dixon AFG. 1989. A rapid method to isolate symbiotic bacteria from aphids. *J Invertebr Pathol*. 53:427–428.
- Horn M, et al. 2001. Members of the Cytophaga-Flavobacterium-Bacteroides phylum as intracellular bacteria of acanthamoebae: proposal of “*Candidatus* Amoebophilus asiaticus.”. *Environ Microbiol*. 3: 440–449.
- Huson DH, Mitra S, Ruscheweyh H-J, Weber N, Schuster SC. 2011. Integrative analysis of environmental sequences using MEGAN4. *Genome Res*. 21:1552–1560.
- Jarrell KF, McBride MJ. 2008. The surprisingly diverse ways that prokaryotes move. *Nat Rev Microbiol*. 6:466–476.
- Karlsson FH, Ussery DW, Nielsen J, Nookaew I. 2011. A closer look at bacteroides: phylogenetic relationship and genomic implications of a life in the human gut. *Microb Ecol*. 61:473–485.
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 30:3059–3066.
- Kaur SJ, et al. 2012. TolC-Dependent secretion of an Ankyrin repeat-containing protein of *Rickettsia typhi*. *J Bacteriol*. 194:4920–4932.
- Kearns DB. 2007. Bright insight into bacterial gliding. *Science* 315: 773–774.
- Kitajima EW, et al. 2007. In situ observation of the *Cardinium* symbionts of *Brevipalpus* (Acari: Tenuipalpidae) by electron microscopy. *Exp Appl Acarol* 42:263–271.
- Koskiniemi S, et al. 2013. Rhs proteins from diverse bacteria mediate intercellular competition. *Proc Natl Acad Sci U S A*. 110:7032–7037.
- Kurtz S, et al. 2004. Versatile and open software for comparing large genomes. *Genome Biol*. 5: R12.
- Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25:2286–2288.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 13:2178–2189.
- Linhartová I, et al. 2010. RTX proteins: a highly diverse family secreted by a common mechanism. *FEMS Microbiol Rev*. 34:1076–1112.
- Lo N, Casiraghi M, Salati E, Bazzocchi C, Bandi C. 2002. How many *Wolbachia* supergroups exist? *Mol Biol Evol*. 19:341–346.
- Maddison WP, Maddison DR. 2011. Mesquite: a modular system for evolutionary analysis, Version 2.75 [cited 2014 Apr 21]. Available from: <http://mesquiteproject.org>.
- Manzano-Marín A, Lamelas A, Moya A, Latorre A. 2012. Comparative genomics of *Serratia* spp.: two paths towards endosymbiotic life. *PLoS One* 7:e47274.
- Mauriello EMF, et al. 2010. Bacterial motility complexes require the actin-like protein, MreB and the Ras homologue, MglA. *EMBO J*. 29:315–326.
- McBride MJ. 2004. Cytophaga-flavobacterium gliding motility. *J Mol Microbiol Biotechnol*. 7:63–71.
- McBride MJ, Zhu Y. 2013. Gliding motility and Por secretion system genes are widespread among members of the phylum bacteroidetes. *J Bacteriol*. 195:270–278.
- Mignot T, Shaevitz JW, Hartzell PL, Zusman DR. 2007. Evidence that focal adhesion complexes power bacterial gliding motility. *Science* 315: 853–856.
- Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet*. 42:165–190.
- Morimoto S, Kurtti TJ, Noda H. 2006. In vitro cultivation and antibiotic susceptibility of a Cytophaga-like intracellular symbiont isolated from the tick *Ixodes scapularis*. *Curr Microbiol*. 52:324–329.
- Moya A, Guirao P, Cifuentes D, Beitia F, Cenis JL. 2001. Genetic diversity of Iberian populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae) based on random amplified polymorphic DNA-polymerase chain reaction. *Mol Ecol*. 10:891–897.
- Moya A, Peretó J, Gil R, Latorre A. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat Rev Genet*. 9: 218–229.
- Nakamura Y, et al. 2009. Prevalence of *Cardinium* bacteria in planthoppers and spider mites and taxonomic revision of “*Candidatus* *Cardinium hertigii*” based on detection of a new *Cardinium* group from biting midges. *Appl Environ Microbiol*. 75:6757–6763.
- Nakane D, Sato K, Wada H, McBride MJ, Nakayama K. 2013. Helical flow of surface protein required for bacterial gliding motility. *Proc Natl Acad Sci U S A*. 110:11145–11150.
- Nan B, et al. 2011. Myxobacteria gliding motility requires cytoskeleton rotation powered by proton motive force. *Proc Natl Acad Sci U S A*. 108:2498–2503.
- Nan B, et al. 2013. Flagella stator homologs function as motors for myxobacterial gliding motility by moving in helical trajectories. *Proc Natl Acad Sci U S A*. 110:E1508–E1513.

- Nawrocki EP. 2009. Structural RNA homology search and alignment using covariance models [Ph.D. thesis]. [Saint Louis]: Washington University, School of Medicine.
- Oakeson KF, et al. 2014. Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome Biol Evol.* 6:76–93.
- Oliver KM, Campos J, Moran NA, Hunter MS. 2008. Population dynamics of defensive symbionts in aphids. *Proc Biol Sci.* 275:293–299.
- Oliver KM, Degnan PH, Burke GR, Moran NA. 2010. Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annu Rev Entomol.* 55:247–266.
- Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Park J, et al. 2012. Identification of biotypes and secondary endosymbionts of *Bemisia tabaci* in Korea and relationships with the occurrence of TYLCV disease. *J Asia Pac Entomol.* 15:186–191.
- Penz T, et al. 2012. Comparative genomics suggests an independent origin of cytoplasmic incompatibility in *Cardinium hertigii*. *PLoS Genet.* 8: e1003012.
- Quast C, et al. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41: D590–D596.
- R Development Core Team. 2011. R: a language and environment for statistical computing. Vienna (Austria): the R Foundation for Statistical Computing.
- Rader BA, Kremer N, Apicella MA, Goldman WE, McFall-Ngai MJ. 2012. Modulation of symbiont lipid A signaling by host alkaline phosphatases in the squid-vibrio symbiosis. *MBio.* 3:e00093–12.
- Russell JA, Latorre A, Sabater-Muñoz B, Moya A, Moran NA. 2003. Sidestepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. *Mol Ecol.* 12:1061–1075.
- Santos-García D, et al. 2012. Complete genome sequence of “*Candidatus Portiera aleyrodidarum*” BT-QVLC, an obligate symbiont that supplies amino acids and carotenoids to *Bemisia tabaci*. *J Bacteriol.* 194: 6654–6655.
- Sato K, et al. 2010. A protein secretion system linked to bacteroidetes gliding motility and pathogenesis. *Proc Natl Acad Sci U S A.* 107: 276–281.
- Schmitz-Esser S, Penz T, Spang A, Horn M. 2011. A bacterial genome in transition—an exceptional enrichment of IS elements but lack of evidence for recent transposition in the symbiont *Amoebophilus asiaticus*. *BMC Evol Biol.* 11:270.
- Schmitz-Esser S, et al. 2010. The genome of the amoeba symbiont “*Candidatus Amoebophilus asiaticus*” reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J Bacteriol.* 192:1045–1057.
- Sibley LD, Håkansson S, Carruthers VB. 1998. Gliding motility: an efficient mechanism for cell penetration. *Curr Biol.* 8:R12–R14.
- Sibley LD, et al. 2004. Intracellular parasite invasion strategies. *Science* 304: 248–253.
- Skaljic M, Zanik K, Ban SG, Kontsedalov S, Ghanim M. 2010. Co-infection and localization of secondary symbionts in two whitefly species. *BMC Microbiol.* 10:142.
- Sloan DB, Moran NA. 2012. Endosymbiotic bacteria as a source of carotenoids in whiteflies. *Biol Lett.* 8:986–989.
- Spalding MD, Prigge ST. 2010. Lipoid acid metabolism in microbial pathogens. *Microbiol Mol Biol Rev.* 74:200–228.
- Spormann AM. 1999. Gliding motility in bacteria: insights from studies of *Myxococcus xanthus*. *Microbiol Mol Biol Rev.* 63:621–641.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Sun M, Wartel M, Cascales E, Shaevitz JW, Mignot T. 2011. Motor-driven intracellular transport powers bacterial gliding motility. *Proc Natl Acad Sci U S A.* 108:7559–7564.
- Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28:33–36.
- Thao MLL, Baumann P. 2004. Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. *Appl Environ Microbiol.* 70:3401–3406.
- Toft C, Andersson SGE. 2010. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat Rev Genet.* 11:465–475.
- Uchiyama I. 2003. MGD: microbial genome database for comparative analysis. *Nucleic Acids Res.* 31:58–62.
- Varani AM, Siguier P, Goubeyre E, Charneau V, Chandler M. 2011. ISSaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biol.* 12:R30.
- Warnes GR, Bolker B, Lumley T. 2013. gplots: various R programming tools for plotting data.
- Wernegreen JJ. 2012. Strategies of genomic integration within insect-bacterial mutualisms. *Biol Bull.* 223:112–122.
- White JA, Kelly SE, Cockburn SN, Perlman SJ, Hunter MS. 2011. Endosymbiont costs and benefits in a parasitoid infected with both *Wolbachia* and *Cardinium*. *Heredity (Edinb)* 106:585–591.
- Williams AM, Westrop GD, Coombs GH. 2009. Two pathways for cysteine biosynthesis in *Leishmania major*. *Biochem J.* 420:451–462.
- Xie G, et al. 2007. Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*. *Appl Environ Microbiol.* 73: 3536–3546.
- Zchori-Fein E, Perlman SJ. 2004. Distribution of the bacterial symbiont *Cardinium* in arthropods. *Mol Ecol.* 13:2009–2016.
- Zchori-Fein E, et al. 2004. Characterization of a “Bacteroidetes” symbiont in *Encarsia* wasps (Hymenoptera: Aphelinidae): proposal of “*Candidatus Cardinium hertigii*.”. *Int J Syst Evol Microbiol.* 54: 961–968.

Associate editor: Richard Cordaux