



HAL
open science

Évolution et dynamique du système di-symbiotique chez les pucerons du genre *Cinara*

Jeff Rouil

► **To cite this version:**

Jeff Rouil. Évolution et dynamique du système di-symbiotique chez les pucerons du genre *Cinara*. Microbiologie et Parasitologie. Montpellier SupAgro, 2023. Français. NNT : 2023NSAM0004 . tel-04241506

HAL Id: tel-04241506

<https://theses.hal.science/tel-04241506v1>

Submitted on 13 Oct 2023

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.

**THÈSE POUR OBTENIR LE GRADE DE DOCTEUR
DE L'INSTITUT AGRO MONTPELLIER
ET DE L'UNIVERSITE DE MONTPELLIER**

En Évolution des systèmes infectieux

École doctorale GAIA

Unité de recherche CBGP (UMR1062)

**Évolution et dynamique du système di-symbiotique chez
les pucerons du genre *Cinara***

Présentée par Jeff ROUÏL

Le 8 février 2023

**Sous la direction de Emmanuelle JOUSSELIN
et le co-encadrement de Alejandro MANZANO-MARIN**

Devant le jury composé de

Yannick OUTREMAN, Professeur, Institut Agro Rennes-Angers

Franck DEDEINE, Maître de conférences, Université de Tours

Natacha KREMER, Chargée de recherche, CNRS, Lyon

Frédéric DELSUC, Directeur de recherche, CNRS, Montpellier

Mathieu SICARD, Professeur, Université de Montpellier

Emmanuelle JOUSSELIN, Directrice de recherche, INRAE, Montpellier

Rapporteur

Rapporteur

Examinatrice

Examineur

Examineur

Directrice



**UNIVERSITÉ
DE MONTPELLIER**

**L'INSTITUT
agro Montpellier**

Évolution et dynamique du système di-symbiotique chez les pucerons du genre *Cinara*

Résumé :

Des systèmes pluri-symbiotiques ont été mis en évidence chez plusieurs espèces de pucerons (Hemiptera : Aphididae) et plus particulièrement chez les espèces du genre *Cinara*. La caractérisation taxonomique des symbiotes présents chez une soixantaine d'espèces de *Cinara* a permis de déterminer que *Serratia symbiotica* est la bactérie la plus fréquemment retrouvée comme symbiote secondaire obligatoire. De plus, la reconstruction de l'histoire évolutive des associations symbiotiques chez *Cinara* semble indiquer une acquisition ancestrale de *Serratia symbiotica* comme co-symbiote obligatoire, même si la présence d'autres bactéries chez certaines espèces indique que des évènements de remplacement ont eu lieu. L'analyse des génomes de souches de *Serratia symbiotica* ont montré que ces symbiotes possèdent des caractéristiques génomiques différentes et des variations de leur forme et de leur localisation dans des bactériocytes. Ces données semblent montrer un gradient de modifications dans le passage d'une bactérie facultative à une bactérie endosymbiote fixée sur lequel se trouve l'ensemble des co-symbiotes obligatoires des *Cinara*. Au cours de cette thèse, à l'aide d'approches de phylogénomique, nous avons décrit l'histoire évolutive de *Serratia symbiotica* et mis en évidence de multiples acquisitions indépendantes de cette bactérie comme co-symbiote obligatoire des *Cinara*. Ces transitions d'un mode de vie facultatif à obligatoire sont associées à des réductions de génome et de taux de G-C plus ou moins importantes selon les lignées, révélant l'ancienneté relative de l'association. Parmi les évènements d'acquisitions, nous avons montré que l'un d'entre eux avait conduit à une co-spéciation durant environ 20Ma des trois partenaires de la symbiose nutritionnelle (*Cinara*, *Buchnera* et *Serratia*). Dans ce clade, nous avons observé des structures cellulaires et des caractéristiques génomiques stables pour les deux co-symbiotes. De plus, l'analyse des taux de substitutions de chacun des partenaires a permis de décrire pour la première fois une évolution parallèle entre les deux symbiotes et leur hôte. L'analyse détaillée de la composition en gènes des bactéries co-symbiotiques de ce clade a mis en évidence une complémentarité forte pour la production des nutriments nécessaires à l'hôte. De plus, l'analyse de l'ensemble de leurs gènes a révélé qu'une forte sélection purifiante s'exerce sur eux. Dans une autre partie, une approche expérimentale de quantification des symbiotes au cours de développement sur trois espèces de *Cinara* suggère une différence de capacité de régulation du co-symbiote *Serratia*. Cette différence est observée

entre deux espèces de *Cinara* pour lesquelles les *Serratia* sont issues d'acquisition indépendantes et dans lesquelles sa localisation est différente. Finalement, une analyse de comparaison génomique sur des phages APSE (bactériophage intégré) présents dans les co-symbiotes de *Cinara* et procurant une défense contre les parasitoïdes a permis de décrire la diversité des APSE et de l'évolution de leur génome chez les pucerons. Cette analyse a aussi montré que les éléments mobiles pourraient jouer un rôle dans l'acquisition de nouveaux gènes dans la cassette de la toxine défensive. L'ensemble de ces travaux révèle un système pluri-symbiotique dynamique dans lequel le nouveau partenaire bactérien est régulièrement remplacé au cours de l'évolution.

Mots clefs : Endosymbiote – interaction interspécifique – bactérie – phylogénomique – co-évolution – pucerons

Evolution and dynamic of di-symbiotic system in *Cinara* Aphids

Abstract :

The existence of multi-symbiotic systems has been demonstrated in several aphid species (Hemiptera: Aphididae) and more particularly in species of the genus *Cinara*. The taxonomic characterisation of the symbionts present in about sixty species of *Cinara* has made it possible to determine that *Serratia symbiotica* is the most frequently found bacterium as an obligatory secondary symbiont. Furthermore, the reconstruction of the evolutionary history of symbiotic associations in *Cinara* species suggests an ancestral acquisition of *Serratia symbiotica* as an obligate co-symbiont, although the presence of other bacteria in some species indicates that replacement events have occurred. More recently, genome analysis of *Serratia symbiotica* strains associated have shown that these symbionts exhibit different genomic characteristics ranging and variations in their shape and localisation in bacteriocytes. These data seem to show a gradient of changes in the transition from a facultative life-style to obligate endosymbiont in which we observe all of the (now obligate) co-symbionts of *Cinara*. During this phd, using phylogenomics, we inferred the evolutionary history of *Serratia symbiotica* within *Cinara* and highlighted multiple independent acquisitions of this bacterium as a co-obligate symbiont of *Cinara*. These transitions from facultative to obligate life-style are associated with genome reductions and low G-C content that are more or less important depending on the lineage, revealing the relative age of the symbiotic association with *Serratia*. Among the inferred acquisition events, one in particular led to co-speciation over about 20 Ma of the three partners of the nutritional symbiosis (*Cinara*, *Buchnera* and *Serratia*). In this clade, we observed stable cellular and genomic structures with highly reduced genomes and low G-C content for both co-symbionts. Furthermore, the analysis of the substitution rates of each partner allowed us to describe for the first time a parallel evolution between the two symbionts and their host. Further analysis of the gene composition of the co-symbiotic bacteria in this clade revealed strong complementation for the production of nutrients required by the host. Furthermore, the analysis of their set of genes revealed that they are under strong purifying selection. In addition an experimental approach aimed at quantifying the symbionts during development on three *Cinara* species suggest that aphid species differ in their capacity to regulate the co-symbiont *Serratia*. This difference is observed between two *Cinara* species for which the *Serratia* originated from independent acquisitions and in which its localisation is different. Finally, a genomic comparison analysis on APSE phages (integrated bacteriophage) present in *Cinara* co-

symbionts and providing defence against parasitoids allowed a description the diversity of APSEs and their genome evolution in aphids. This analysis also showed that mobile elements could play a role in the acquisition of new genes in the defensive toxin cassette. Altogether, this study reveals a dynamic multi-partner endosymbiosis where the new partner is recurrently replace.

Key words : Endosymbiont – interspecific interaction – bacteria – phylogenomic – co-evolution – Aphids

Remerciements

Je voudrais tout d'abord remercier les membres de mon jury, Yannick Outreman, Franck Dedeine, Natacha Kremer, Frédéric Delsuc et Mathieu Sicard, d'avoir accepté d'évaluer mon travail. Ce faisant, vous me permettez de présenter mes travaux effectués pendant ces trois années de doctorats, et pour cela, je vous en remercie.

Merci à tous ceux et celles qui ont pu m'aider dans mon travail ou, pour d'autres, de me l'avoir fait oublier ! Tout d'abord un grand merci à mon encadrement Emmanuelle et Alejandro de m'avoir fait confiance pour ce projet, moi qui ne connaissais quasiment pas le vaste monde de la phylogénie. De m'avoir accompagné de la reconstruction de mon premier arbre, à la rédaction de ce manuscrit en passant par le repotage de nombreux arbres qui ne seront pas mort en vain. Merci Emmanuelle pour ta patience et le temps que tu as passé à m'aider à façonner cette thèse, même avec un emploi du temps (sur)chargé, tu as toujours su trouver du temps. Merci à Anne-Laure pour m'avoir formée et accompagnée durant les milliers d'extractions et PCR et surtout de les avoir finies pour moi. On a réussi à mettre en place du clonage ensemble (au vu des poubelles de séquençage) On s'est cloné un peu nous-même et plein d'autre truc au passage ;). Merci à Armelle pour son expertise entomologie, moi qui ignorais que les pucerons n'étaient pas tous les mêmes. J'ai pu découvrir l'existence de plein de petites bêtes sur des pins ! Qui l'aurait cru !

Un grand merci à tous les membres du CBGP de m'avoir accueilli et d'avoir su faire de ce labo perdu dans la campagne un lieu convivial et chaleureux où on se sent bien comme à la maison. Pour ces bons moments passés à parler de sciences, mais aussi et surtout de toute autres choses pendant les très nombreuses pauses-café et bière party. Sabine, Nath, Guigui, Nicolas, Frago, Laure, Anne, Max, Anne-laure... et à tous les autres qui font partie du noyau dur de la bonne ambiance, je vous laisse vous partager ma note du bière-party ne vous battez pas ! J'espère passer encore de bons moments en votre compagnie.

Je tiens en particulier à remercier mes co-bureaux, Isis alias petit poulain pour ta bizarrerie attachante qui va grandement me manquer. Je me souviendrai toujours de ton premier jour de

thèse aussi connu comme le premier jour de confinement, tu es rentrée dans le bureau, tu as dit bonjour, tu as pris ton ordi et tu es partie ! Depuis ce jour, bien des choses se sont passé, on s'est battue... T'as fini dans la mer, on s'est entraîné et on a surtout bien rigolé ! Je garde précieusement quelques dossiers pour ta soutenance à toi sur qui l'alcool n'a pas d'effet « petit poisson !!! » et je te souhaite du courage pour la suite et à bientôt ! Je garde le souvenir de ton bisou encré dans ma mémoire ! Merci aussi à mon petit ours qui m'a suivi dans mon bureau pendant plus de 4 ans pour le meilleur et pour le pire, comme elle a si bien dit : « Je remercie principalement Marie Bouilloud, une co-bureau merveilleusement géniale, la personne la plus formidable du monde, la plus chiante aussi. Tu as su égayer mes moments de thèses par des beaux chants et danses insupportables, tes râles intenses, tes pleurs et tes rires pas très discret, tes crises de chocolats, tes expressions qui veulent rien dire et tes lectures de l'avenir flippantes... Tu as cherché chaque problème de thèse et non thèse de manière déterminé jusqu'à me faire craquer et tu as su me soutenir à ta manière chelou par des coups de pieds au cul parfois et des conseils en noix de coco. Les pauses cafés du matin ne seront plus jamais les mêmes sans toi. Tu as aussi été présente pendant ces 3 ans comme un membre à part entière de ma famille, après tout tu as été ma fille aux eux de la caf... mais je me n'étalerai pas sur cette partie car ça rentrera pas dans mes 2 pages de remerciement de thèses. Et j'espère secrètement que tu seras encore ma co-bureau dans un futur, parce qu'après tout jamais deux sans trois... ». Je te dis donc à dans un an quand tu viendras de nouveau t'incruster dans mon bureau ! Je t'accueillerai avec plaisir, tu resteras la meilleure collègue de tous les temps et je ne te remercierai jamais assez pour tout ce que tu as fait pour moi. On a découvert Montpellier et ses alentours ensemble et je t'attends pour découvrir notre future destination de post-doc. Hâte de te voir docteure ! Je serais la ! Et puis merci aussi à mes deux bébés qui ont bien grandi Axo et Lotl (merci Isis de les avoir nommées au passage prends soin d'eux).

Merci à tous les autres qui ont œuvré sans forcément le savoir au bon déroulement de ma thèse par leur compagnie, les bons moments passés ensemble au labo, mais aussi à l'extérieur que ce soit pour des randos des soirées plage ou des beuveries. Un merci particulier à la team de coureuse, Bérénice, Marie, Cécile, Marjolaine j'espère ne pas trop vous avoir forcé à courir et que vous continuerez sans moi. Je suis fière d'avoir pu me défouler en compagnie d'athlètes telles que vous ! Merci à mon choupisson garde-fou qui m'a supporté soutenu et bien plus pendant les derniers mois, mais aussi corrigé sûrement plusieurs centaines de fautes d'orthographe ! Je ne t'en remercierai jamais assez. Tu étais toujours partante pour toute activité

et surtout pour faire du sport que ce soit pour marcher, grimper, se baigner en plein hiver ou courir... Les Zinzolines gagnent les courses et j'espère continuerons de les gagner ! On se voit bientôt pour le Gr20, le trail du mont blanc et le défi finale la diagonale des fous, je n'oublie pas. Merci à ma petite Chacha, Cynthia et bien d'autres pour nos nombreuses danses endiablées ! Hâte de pouvoir vous montrer mes progrès. Un grand merci à la konfrerie boule de neige et aux dieux du médocs qui m'ont permis de m'évader de temps en temps aux quatre coins de la France et m'avoir supporté et soutenu promis, je reviens bientôt vous voir !

Merci à l'entreprise Poulain pour la fabrication des 8 kilos de chocolat chaud consommé et bien sûr aux nombreuses vaches qui ont participé à fournir le lait. Finalement merci à toi qui lis mon manuscrit que tu sois stagiaire, doctorant en galère ou autres ; que tu lises ce manuscrit par nécessité ou par passion pour les pucerons ;)

Liste des publications

Jeff Rouïl, Emmanuelle Jousselin, Armelle Coeur d'acier, Corinne Cruaud, Alejandro Manzano-Marín, The Protector within: Comparative Genomics of APSE Phages across Aphids Reveals Rampant Recombination and Diverse Toxin Arsenals, *Genome Biology and Evolution*, Volume 12, Issue 6, June 2020, Pages 878–889, <https://doi.org/10.1093/gbe/evaa089>

Table des matières

Introduction	1
I. Les symbioses mutualistes au sein des arthropodes	8
Aperçu de la diversité des symbioses mutualistes au sein des arthropodes	8
Caractérisation des symbioses mutualistes au sein des arthropodes	9
II. Evolution des génomes endosymbiotiques bactériens	13
III. Les systèmes di-symbiotiques dans lesquels les symbiotes assurent la nutrition de leurs hôtes	17
IV. Les pucerons (Hemiptera : Aphididae)	19
V. Les pucerons du genre <i>Cinara</i>	22
VI. Objectifs	27
Chapitre 1 : Histoire évolutive de l'acquisition de <i>Serratia symbiotica</i> comme co-symbiote obligatoire dans le genre <i>Cinara</i>.	31
I. Introduction	33
II. Matériel et méthode	40
Collecte des pucerons, extraction de l'ADN, séquençage et assemblage	40
Phylogénie	41
Phylogénie mitochondriale de <i>Cinara</i>	42
Phylogénies des symbiotes : <i>Buchnera</i> et <i>Serratia</i>	43
Réconciliation	44
Localisation et morphologie des co-symbiotes obligatoires	46
III. Résultats	48
Caractéristiques génomiques	48
Phylogénies	49
Réconciliation	50
FISH	53
IV. Discussion	54
Y'a-t-il eu cospéciation entre les trois partenaires de cette symbiose tri-partite	54
Une histoire de co-diversification complexe : quels processus expliquent cette dynamique ?	57
Conclusion	58
Chapitre 2 : Dual nutritional symbionts in conifer feeding aphids: cospeciation and parallel evolution in symbiont genomes	60
I. Introduction	64
II. Materials and methods	67
Aphid collection, DNA extraction and sequencing	67
Fluorescence in situ hybridisation microscopy	68
Genome assembly and annotation	68
Intraspecific variation in symbiont genomes in two <i>Cinara</i> species	70
Phylogenetic analyses	70
<i>Cinara</i> mitochondrial phylogeny	70

<i>Buchnera and Serratia Symbiont phylogenies</i>	70
Examining patterns of genome evolution in co-evolving symbionts	71
<i>Estimating evolutionary rates</i>	71
<i>Evolutionary rate correlation</i>	72
<i>Test of positive and relaxed selection</i>	72
<i>Intraspecific variations in symbiont genomes</i>	73
III. Results	74
Endosymbiont Gene content	74
Endosymbionts nutritional complementation	78
Phylogenetic history of <i>Cinara</i> di-symbiotic system	80
Endosymbionts genome evolution	80
<i>Rate correlation</i>	81
<i>Signature of selection across endosymbiont genomes</i>	82
<i>Intraspecific variations in symbiont genomes</i>	82
Localisation of <i>Serratia symbiotica</i> in <i>Cinara</i> species	83
IV. Discussion	83
<i>Serratia symbiotica has co-specified with a clade of Cinara</i>	84
<i>Serratia and Buchnera evolve at the same pace</i>	85
<i>Patterns of gene losses, signature of selection across genomes in aphid di-symbiotic systems</i>	87
<i>Conclusions</i>	89
Chapitre 3 : Dynamique des endosymbiotes nutritionnel et contrôle de l'hôte.	93
Abstract	95
Résumé	97
I. Introduction	98
II. Matériels et méthodes	101
Matériel biologique	101
Echantillonnage des stades de développement	103
Protocole de quantification	105
<i>Extraction</i>	105
<i>Conception des amorces</i>	105
<i>Courbe standard : clonage</i>	106
<i>PCR quantitative</i>	108
Analyses	109
Microscopie à hybridation in situ par fluorescence	110
III. Résultats	111
Quantification de la densité en symbiote dans les espèces de <i>Cinara</i>	111
Localisation des co-symbiotes dans les espèces de <i>Cinara</i>	114
IV. Discussion	115
Chapitre 4 : The protector within: Comparative genomics of APSE phages across aphids reveals rampant recombination and diverse toxin arsenals	120
Contexte de l'étude	122
Discussion générale	136
Une histoire évolutive complexe	138
<i>De multiples acquisitions</i>	138

<i>Et de la co-spéciation</i>	139
Pourquoi une telle dynamique et pourquoi <i>Serratia symbiotica</i> ?	140
<i>Un avantage évolutif pour l'hôte</i>	140
<i>Ou une compétition entre endosymbiotes bactériens</i>	142
Régulations de l'interaction hôte co-symbiote	143
Réduction des génomes symbiotiques	144
Conclusion	145
<i>Références bibliographiques</i>	148
<i>Annexes 1 : Matériel supplémentaire du chapitre 1</i>	174
<i>Annexes 2 : Matériel supplémentaire du chapitre 2</i>	182
<i>Annexes 3 : Matériel supplémentaire du chapitre 3</i>	196
<i>Annexes 4 : Matériel supplémentaire du chapitre 4</i>	201

Introduction

Depuis l'émergence de l'écologie en tant que discipline scientifique, les organismes tendent à être définis non seulement par leurs caractéristiques individuelles mais aussi par l'ensemble des interactions qu'ils ont avec leur environnement et avec les autres organismes (Boucher et al., 1982; Gilbert et al., 2012). En effet, les connaissances acquises durant les deux derniers siècles nous permettent aujourd'hui d'affirmer que chaque être vivant est en constante interaction avec une multitude d'autres organismes partageant le même environnement. Ainsi les notions **d'interaction positive**¹ ou négative entre organismes sont apparues. L'ensemble de ces interactions, qu'elles soient néfastes, neutres ou bénéfiques pour les organismes qui y sont impliqués, ont un rôle important dans la diversification écologique et évolutive au sein du règne animal et végétal. En effet, celles-ci contraignent les organismes à se développer et évoluer conjointement. De cela découlent de nouveaux processus biologiques pouvant jouer un rôle important dans l'évolution des traits et la diversification des êtres vivants (Boucher et al., 1982; Christian et al., 2015; Gilbert et al., 2012; McFall-Ngai et al., 2013).

Au cours du siècle dernier l'avancée des technologies a largement facilité la détection et l'observation des micro-organismes et a profondément changé notre façon d'étudier les organismes vivants. En effet, le séquençage haut débit couplé au développement de la bioinformatique et des capacités de calcul ont par exemple permis le développement de la **métagénomique microbienne** et la découverte de l'existence d'une grande diversité de micro-organismes. L'amélioration des techniques de microscopie ont quant à elles facilité la description des bactéries ainsi que permis de décrire leur localisation et leur structuration dans différents environnements. Ces technologies ont donc permis de mettre en évidence l'omniprésence des micro-organismes dans l'environnement et leur grande diversité. Ces avancées ont également permis d'établir que les micro-organismes sont impliqués dans des relations multiples et complexes, non seulement entre eux, mais aussi avec l'ensemble des macro-organismes appartenant au règne animal et végétal.

Ces associations peuvent être impliquées dans diverses fonctions biologiques essentielles telles que le développement (comme chez les souris (Hooper et al., 2001) ; les salamandres (Olivier and Moon, 2010) ; les calamars (McFall-Ngai et al., 2012), la digestion (chez l'humain (Le Chatelier et al., 2013) ; les termites (Brune, 2014)), la nutrition (chez les pucerons (Baumann, 2005a) ; les cochenilles (McCutcheon and von Dohlen, 2011) ; les plantes (Philippot et al., 2013)), la reproduction (chez les guêpes parasitoïdes (Dedeine et al., 2001;

¹ Les termes en gras sont définis dans un encadré page 7

Giorgini et al., 2010)), l'immunité (chez la souris (Mazmanian et al., 2005) ; et divers autres mammifères (Lanning et al., 2005); mais aussi le corail (Rosenberg et al., 2007); les plantes (Herre et al., 2007)... Parmi toutes ces interactions, certaines sont aujourd'hui bien documentées et étudiées et peuvent être considérées comme des exemples assez emblématiques (ou systèmes d'études modèles) d'interactions durables entre macro-organismes et micro-organismes (Christian et al., 2015; Gilbert et al., 2012; McFall-Ngai et al., 2013).

Ainsi, chez les plantes, l'association de la famille des Fabacées avec les rhizobiums est particulièrement étudiée pour ses intérêts agronomiques, puisque celle-ci favorise la croissance de plantes via un apport d'azote par les bactéries en échange de substrat carboné apporté par la plante (Van rhijn and Vanderleyden, 1995). Chez l'humain le microbiote intestinal a fait l'objet de nombreuses études au cours de la dernière décennie et a amené à la découverte d'un écosystème complexe avec une grande diversité de micro-organismes aux rôles très divers et essentiels en santé humaine. En effet, en plus de son rôle majeur dans la digestion, le microbiote intestinal a un rôle important dans le métabolisme et les fonctions immunitaires de l'humain (Doré et al., 2013; Le Chatelier et al., 2013). Ces relations entre micro-organismes et macro-organismes peuvent être transitoires (c'est-à-dire que les partenaires de ces interactions ne sont en contact que pendant une courte durée de leur cycle de vie). Mais il existe également de nombreux exemples **d'interactions durables** entre des micro-organismes et des macro-organismes, on parle alors **de symbiose**.

Par ailleurs, l'association entre micro-organismes est aujourd'hui considérée comme un facteur important dans l'innovation macro-évolutive (Margulis L. and Fester R., 1991). En effet, celle-ci est à l'origine de l'apparition de la cellule eucaryote dont les mitochondries et les chloroplastes sont respectivement issus d'association avec une alphaprotéobactérie et une cyanobactérie autrefois libre (Margulis, 1981; Moreira and Lopez-Garcia, 1998; Muñoz-Gómez et al., 2022, 2017). L'interaction entre les bactéries et les protistes semble également pouvoir initier la multicellularité dans certains cas particuliers et pourrait être à l'origine de l'apparition d'organismes multicellulaires (comme observé chez des choanoflagellés) (Alegado et al., 2012; Dayel et al., 2011).

Ces découvertes d'interactions intimes et omniprésentes ont profondément remis en question la vision généralement admise des "individus" et toutes les conceptions classiques de l'individualité rendant les limites de l'organisme plus floues et parfois difficiles à poser (Gilbert et al., 2012). Pour rendre compte de cette complexité, le terme "**holobionte**" a été introduit pour

décrire le complexe multi-espèces composé à la fois des éléments hôtes et des populations persistantes de symbiotes (Gilbert et al., 2012; Rosenberg et al., 2007). Ainsi, l'holobionte, avec sa communauté intégrée d'espèces, devient une unité de sélection naturelle. Face à cette nouvelle échelle de complexité du vivant, de nouveaux axes de recherches se sont mis en place dans lesquelles la diversité des associations avec des microorganismes devient une caractéristique centrale dans la description des espèces dans tous leurs aspects : développemental, physiologique, génétique et même évolutif (Christian et al., 2015; Gilbert et al., 2012).

Dans ce contexte, l'étude de l'évolution de ces associations devient essentielle pour comprendre la diversité et le fonctionnement du vivant. Malgré les débats sur l'holobionte (Douglas and Werren, 2016; Moran and Sloan, 2015), les études sur les systèmes symbiotiques ne sont aujourd'hui plus ciblées uniquement sur quelques systèmes emblématiques, mais étendues à de nombreux systèmes biologiques. Cet essor a un fort impact sur les recherches fondamentales menées dans le domaine de l'écologie évolutive mais il permet également d'envisager des applications dans différents domaines de la biologie, en médecine ou en agronomie.

La transplantation de microbiote intestinal est un exemple d'utilisation des découvertes récentes sur la diversité et le rôle prépondérant des micro-organismes en médecine. Elle est utilisée pour le traitement de l'infection récurrente à *Clostridium difficile*. Celle-ci permettrait de rétablir un microbiote intestinal normal qui agit comme barrière à la prolifération de *C. difficile*. Cette technique s'est avérée nettement plus efficace que l'utilisation d'antibiotiques (Gough et al., 2011). Un autre exemple de manipulation du microbiote est celui de la suppression de symbiotes essentiels à la survie d'insectes comme moyen écologique de lutter contre les vecteurs de maladies ou des ravageurs des cultures (Elston et al., 2022; Mendiola et al., 2020; Zindel et al., 2011). Un exemple emblématique en santé humaine concerne la manipulation de souches de *Wolbachia* (un endosymbiote répandu des arthropodes) qui a permis de bloquer la transmission de différents virus véhiculés par les moustiques (Dengue ou Zika) (Flores and O'Neill, 2018). De telles applications sont encore rares en agriculture, mais on a cependant pu montrer que la transmission de *Xylella fastidiosa*, une bactérie phytopathogène causant des épidémies importantes, est réduite lorsque son insecte vecteur est infecté avec une souche de *Pantoea agglomerans* (un symbiote intestinal des insectes) génétiquement modifiée pour exprimer des peptides antimicrobiens (Mendiola et al., 2020). Certains symbiotes intestinaux, du genre *Asaia*, semblent également capables de réduire la

transmission de phytoplasmes transportés par les cicadelles (Mendiola et al., 2020). Ces exemples montrent que la recherche sur le microbiote, les systèmes symbiotiques et l'étude des conséquences de ces associations sur le phénotype des hôtes peut avoir des enjeux importants dans de nombreux domaines d'application.

Définitions :

Interaction positive : Les interactions positives sont toutes les interactions non prédatrices entre deux espèces ou plus qui ont un effet positif sur au moins une des espèces impliquées par une augmentation relative de leur fitness. L'autre espèce n'est quant à elle pas altérée par l'interaction ou bénéficie aussi d'un effet positif. Cela inclut donc les interactions de facilitation, de mutualisme (ou coopération entre espèces), de commensalisme et ce quelle que soit la nature facultative ou obligatoire de l'interaction (Bertness et Callaway 1994; Boucher, James, et Keeler 1982).

Interaction négative : Les interactions négatives sont toutes les interactions entre deux espèces ou plus qui ont un effet négatif sur au moins une des espèces impliquées par une diminution relative de leur fitness. L'autre espèce quant à elle peut bénéficier de cette interaction ou au contraire être impactée négativement par celle-ci. Cela inclut donc les interactions de compétition, prédation, le parasitisme et ce quelle que soit la nature facultative ou obligatoire de l'interaction.

Interaction durable : Les interactions durables sont toutes les associations prolongées (c'est-à-dire que l'on retrouve sur plusieurs générations) et intimes entre deux espèces ou plus et impliquent alors une adaptation réciproque (coévolution) des partenaires visant à rencontrer, échanger, communiquer, éviter, tuer, survivre ou exploiter le ou les individus des autres espèces. Elles englobent ainsi toutes les interactions négatives et positives.

Symbiose : Plusieurs définitions de la symbiose existent aujourd'hui, dans le cadre de cette thèse, la définition énoncée par de Bary 1879 sera utilisée. Cette définition décrit la symbiose comme étant une interaction intime et durable pendant au moins une partie du cycle de vie entre des organismes d'espèces différentes, on appelle alors « hôte » l'organisme le plus grand et « symbiote » le plus petit. (de, Bary 1879 ; Moran, McCutcheon, et Nakabachi 2008).

Holobionte : Dans cette thèse, le terme holobionte est utilisé pour décrire le complexe multi-espèces composé à la fois des éléments hôtes et des populations persistantes de symbiotes (Margulis and Fester 1991). Une deuxième définition plus large décrit l'holobionte comme le complexe formé par l'hôte et l'ensemble de ses microorganismes (aussi appelé microbiome) (Simon et al. 2019). L'holobionte, avec sa communauté intégrée d'espèces, serait alors l'unité principale sur laquelle agit la sélection naturelle (Gilbert, Sapp, et Tauber 2012; Rosenberg et al. 2007; Rosenberg et Zilber-Rosenberg 2016).

Métagénomique microbienne : La métagénomique est une technique de séquençage et d'analyse de l'ADN contenu dans un milieu. Elle consiste à séquencer les génomes (shotgun) ou un gène particulier (metabarcoding) de plusieurs individus d'espèces différentes présents dans un environnement spécifique. Ainsi l'analyse nous permet de déterminer la composition du microbiome. C'est à dire quelles espèces sont présentes, parfois leurs abondances et leur diversités dans un environnement donné (Combes, 1995).

Coévolution : La coévolution est le processus d'adaptation réciproque des partenaires d'une interaction interspécifique : il est utilisé pour décrire les cas où l'évolution d'une espèce affecte l'évolution d'une autre espèce (ou de plusieurs) et réciproquement. Il a été défini clairement par Janzen (1980)

Cospéciation : La cospéciation correspond à une spéciation conjointe des partenaires d'une symbiose. Les événements de spéciation des symbiotes suivent alors les événements de spéciation de leurs hôtes.

I. Les symbioses mutualistes au sein des arthropodes

Aperçu de la diversité des symbioses mutualistes au sein des arthropodes

Dans la nature, les associations entre insectes et bactéries sont courantes et comportent une grande diversité, à la fois des espèces impliquées, du rôle et de la durabilité des associations (Moran et al., 2008a). Un des exemples les plus emblématiques d'une association entre un insecte et une bactérie est peut-être celui des pucerons et de leur bactérie nutritionnelle obligatoire *Buchnera aphidicola*. Cette association fut découverte par Paul Buchner. Il avait décrit chez plusieurs groupes d'hémiptères la présence d'endosymbiotes et avait fait le lien entre un régime alimentaire déficient en nutriments essentiels (les hémiptères se nourrissent de la sève des plantes, riche en sucre mais pauvre en vitamines et en acides aminées) chez ces insectes et l'établissement de symbioses bactériennes (Buchner, 1953a). Il a alors suggéré que la fonction des endosymbiotes était la synthèse des nutriments manquants pour l'hôte (Buchner, 1953a).

Aujourd'hui, cette symbiose est bien décrite, vérifiant ainsi les hypothèses de Buchner quant au rôle nutritionnel de *Buchnera* (Baumann, 2005b; Douglas, 1998a). Celle-ci sera décrite plus en détail dans la partie suivante de cette introduction. Parallèlement à ce système symbiotique emblématique, de nombreux autres symbioses nutritionnelles ont été décrites telles que celles entre les mouches tsétsé (*Glossina sp.*) et la bactérie *Wigglesworthia glossinidia* qui permet aussi la synthèse des nutriments manquants dans l'alimentation spécialisée en sang de son hôte (Aksoy, 1995; Chen et al., 1999). Chez le charançon *Sitophilus oryzae*, une association symbiotique impliquée dans la production de vitamines est aussi décrite avec une enterobactériaceae SOPE ((Heddi et al., 1999, 1998).

Il existe également des lignées de symbiotes particulières qui sont retrouvées chez plusieurs hôtes parfois éloignés phylogénétiquement. C'est le cas de *Wolbachia*, qui est probablement le symbiote bactérien le plus commun et le plus répandu dans l'ensemble parmi les arthropodes (Hilgenboecker et al., 2008; Zug et al., 2012). Il est en effet associé à de nombreux ordres tels que les Hémiptères, Isopodes, Lépidoptères, Acariens, Hyménoptères, Thysanoptères, Coléoptères, Diptères, Lépidoptères and Pseudoscorpions, Orthoptères. Cette bactérie a également étendu sa niche au-delà des arthropodes puisqu'on la trouve associée à des nématodes (Fenn and Blaxter, 2007). Au sein de ces systèmes, *Wobachia* peut être associée à différents types de manipulation de la reproduction (Werren et al., 2008). D'autres exemples

d'endosymbiotes répandus existent comme *Cardinium* que l'on retrouve chez des Hémiptères, des Hyménoptères ainsi que chez différents Arachnides (Bressan et al., 2012; Duron et al., 2008; Zchori-Fein and Perlman, 2004). On peut également citer *Arsenophonus* qui est lui aussi associé à plusieurs ordres d'insectes (Diptères (Nováková et al., 2009), Hémiptères (Jousselin et al., 2013), Hyménoptères (Taylor et al., 2011)).

Face à l'augmentation des études sur les systèmes symbiotiques et les nombreuses symbioses nouvellement décrites qui en découlent, il est difficile de faire une synthèse nouvelle (et à jour) sur ce sujet. Il existe déjà de nombreux articles de revues, concernant ces symbioses. Certaines portent sur l'importance de ces symbioses dans des groupes écologiques particuliers : par exemple chez les insectes phytophages (Frago et al., 2020, 2012; Giron et al., 2017; Hansen and Moran, 2014), chez les arthropodes hématophages (Duron et Gottlieb 2020). De nombreux articles portent sur une lignée de symbiotes en particulier et ses effets divers sur leurs hôtes, il existe ainsi plusieurs articles de synthèse sur différents aspects de la biologie et l'écologie des *Wolbachia* (e.g. (Serbus et al., 2008; Werren, 1997). Enfin, certaines synthèses sont centrées sur l'écologie et l'évolution de ces associations, c'est-à-dire comment ces systèmes se mettent en place et sont évolutivement stables et leur rôle dans l'écologie de leurs hôtes. Le but ici n'est pas de refaire une revue sur les symbiotes des arthropodes, la liste d'exemples présentés à travers cette introduction n'est pas exhaustive. Mon objectif ici sera plutôt d'exposer les caractéristiques de ces symbiotes qu'il est important de définir et de bien caractériser afin de comprendre le fonctionnement et l'évolution de ces associations.

Caractérisation des symbioses mutualistes au sein des arthropodes

Les symbioses peuvent tout d'abord être caractérisées selon **la localisation du symbiote** (Martin and Schwab, 2012). On parle ainsi d'ectosymbiose lorsque le symbiote vit hors de l'hôte. Celui-ci peut être hébergé dans des structures spécialisées, c'est le cas par exemples des champignons symbiotiques associés à certains scolytes xylophages qui participent à la décomposition du bois et facilitent la nutrition de leurs insectes hôtes, ils vivent dans des invaginations cuticulaires appelées Mycangia (Six and Klepzig, 2004). Par opposition, lorsque le symbiote vit à l'intérieur de l'hôte on parle d'endosymbiose. Les symbiotes sont alors essentiellement des bactéries. Leur localisation peut être intracellulaire : c'est-à-dire qu'ils sont trouvés dans des cellules de l'hôte, c'est notamment le cas de *Buchnera aphidicola* chez les pucerons (Braendle et al., 2003; Buchner, 1953a). Ils peuvent aussi être extracellulaires et dans

ce cas on trouve ces symbiotes dans les tissus ou les cavités de l'hôte. C'est le cas des symbiotes intestinaux que l'on trouve dans le système digestif des termites (Brune, 2014).

On peut également classer ces interactions **selon les fonctions et les bénéfices (ou les couts) apportés par les symbiotes**. Chez les insectes, les symbiotes participent à des processus physiologiques très divers tels que la reproduction (Dedeine et al., 2001; Kremer et al., 2009; Sicard et al., 2014a, 2014b; Wang et al., 2020), c'est le cas de la symbiose impliquant *Wolbachia*, que l'on retrouve chez de nombreux groupes d'insectes (Werren et al., 2008). Une des particularités de *Wolbachia* est la variabilité des bénéfices de son interaction avec l'hôte, en effet, dans certains cas un effet délétère de sa présence a été décrit. C'est par exemple le cas chez le scarabée noir de la farine (*Tribolium madens*) où la présence de *Wolbachia* induit la mort de tous les embryons mâles (Fialho and Stevens, 2000). Alors que dans d'autres cas celle-ci est bénéfique pour l'hôte, c'est le cas pour *Asobara tabida* une guêpe parasitoïde pour qui elle est devenue indispensable pour l'oogénèse (Dedeine et al., 2001). Certains symbiotes participent également à la protection de l'hôte (Degnan and Moran, 2008; Haine, 2008; Parker, 2021; Polin et al., 2015), notamment contre les ennemis naturels. C'est le cas du symbiote *Hamiltonella defensa* qui engendre une protection contre les parasitoïdes chez les pucerons (Oliver et al., 2003; Polin et al., 2015, 2014). Les symbiotes peuvent aussi apporter une protection contre des variations de l'environnements comme montré pour certaines souches de *Serratia symbiotica* dont la présence permet de protéger leurs pucerons hôtes des chocs thermiques (Chen et al., 2000; Montllor et al., 2002; Russell and Moran, 2006). Certains symbiotes ont la capacité de modifier le comportement de leur hôte (Hosokawa and Fukatsu, 2020; Sochard et al., 2021), il a été constaté que les symbiotes secondaires portés par *Acyrtosiphon pisum* pouvaient avoir une influence sur la production de la progéniture ailée avec des conséquences potentielles sur la dispersion et la survie (Sochard et al., 2021). Finalement, le cas des symbiotes associés à la nutrition est courant, le répertoire métabolique des bactéries étant très vaste, l'association des insectes avec des bactéries permet une alimentation sur des substrats pauvres ou difficiles à digérer (Akman Gündüz and Douglas, 2009; Douglas, 1998b; Duron and Gottlieb, 2020; Garber et al., 2021). C'est par exemple le cas chez les termites associées à toute une communauté de bactéries qui permettent la digestion du bois (Brune, 2014) ou encore des tiques pour compléter leur alimentation pauvre en nutriments (Binetruy et al., 2020). L'ensemble de ces associations symbiotiques de par leur participation à des processus physiologiques, favorisent la colonisation de nouvelles niches écologiques par leurs hôtes (Moran et al., 2008a; Weber and Agrawal, 2014; Wernegreen, 2012)

et jouent probablement un rôle important dans la diversification des insectes (Chomicki et al., 2019; Janson et al., 2008).

Une autre caractéristique très importante à prendre en compte dans l'étude des symbiotes est le **degré de dépendance** des partenaires de l'interaction. En effet, parmi les endosymbioses, on peut discriminer deux catégories selon le niveau d'interdépendance hôte-symbiote, on parle ainsi de symbiose facultative quand les partenaires peuvent survivre et se reproduire l'un sans l'autre et de symbiose obligatoire lorsqu'ils ne le peuvent pas. Dans le cas d'une symbiose facultative, le symbiote n'est pas retrouvé chez tous les individus de l'espèce hôte et sa prévalence peut varier selon les populations. L'association avec ces symbiotes peut augmenter la valeur sélective des individus dans un environnement donné (Ferrari and Vavre, 2011; Moran, 2007; Oliver et al., 2010). Parmi ces symbiotes facultatifs, on peut retrouver fréquemment les symbiotes jouant un rôle dans la protection face aux ennemis naturels mais aussi les symbiotes manipulateurs de la reproduction comme *Wolbachia*. Ces derniers, a priori apportent rarement un bénéfice à l'hôte mais on suppose qu'ils peuvent parfois jouer un rôle dans la défense de leurs hôte (Zug and Hammerstein, 2015) et peuvent également devenir des symbiotes obligatoires si ils manipulent suffisamment les défenses immunitaires de l'hôte pour que leur suppression empêche la reproduction de leur hôte (Dedeine et al., 2001; Pannebakker et al., 2007). Ainsi on voit que les bénéfices apportés par les symbiotes sont variables, parfois divers pour un même symbiote et dépendent fortement de l'environnement. La balance coût-bénéfice pour ces symbiotes est donc parfois très difficile à mesurer et très souvent contexte dépendante.

Les symbioses obligatoires quant à elles, impliquent que les partenaires sont dépendants l'un de l'autre pour leur survivre et/ou leur reproduction. Cette interdépendance implique alors que l'on retrouve le symbiote chez tous les individus de l'espèce hôte. Chez les insectes ces associations sont souvent retrouvées chez des espèces ayant une source alimentaire pauvre en nutriments essentiels tel que la sève des plantes ou le sang (Buchner, 1953a; Duron and Gottlieb, 2020; Manzano-Marín et al., 2015; Moran et al., 2008a; Wernegreen, 2012). Pour ces insectes, les symbiotes permettent une complémentation nutritionnelle en participant à la synthèse d'éléments essentiels manquant à leur alimentation (Baumann, 2005; Moran et al., 2008). Ces symbioses mutualistes impliquent souvent des bactéries que l'on retrouve hébergées à l'intérieur de cellules spécialisées de l'organisme hôte, les bactériocytes (Baumann, 2005b). Ces structures cellulaires permettent le maintien de la symbiose mutualiste (Braendle et al., 2003; Koga et al., 2012; Luan et al., 2016) et sont retrouvées dans de nombreux groupes :

Blattoptères (López-Sánchez et al., 2009), Coléoptères (Login et al., 2011), Diptères (Hosokawa et al., 2012), Hémiptères (Meseguer et al., 2017; Nakabachi et al., 2006), Hyménoptères (Stoll et al., 2010), Tiques (Duron et al., 2018; Perotti et al., 2006). Elles peuvent former un organe spécialisé, le bactériome.

Dans tous les cas, que l'association soit facultative ou obligatoire, son maintien a un coût pour l'hôte, par exemple en terme de durée de vie ou de fécondité (Fleury et al., 2000; Oliver et al., 2006; Polin et al., 2014; Pons et al., 2019; Russell and Moran, 2006). Dans le cas des symbioses facultatives, ce coût (en terme de valeur sélective) doit être inférieur aux bénéfices afin que la symbiose soit qualifiée de mutualiste et que la sélection favorise les individus porteurs des symbiotes. Dans le cas de symbiotes obligatoires, l'association a également un coût mais les individus sans symbiotes ont une valeur sélective nulle, la sélection favorise cependant l'évolution de mécanisme de contrôle de la densité des symbiotes par les hôtes (Parker, 2021; Whittle et al., 2021). Ce contrôle est facilité par la compartimentation des symbiotes dans des structures spécialisées (Chomicki et al., 2020) et est également fortement dépendant du mode de transmission des symbiotes.

En effet, une des caractéristiques les plus importantes pour comprendre l'évolution des associations est **le mode de transmission de ces symbiotes**. De nombreux symbiotes sont transmis de manière verticale (c'est-à-dire des parents aux descendants), parfois par les deux parents, parfois par la mère uniquement. Ce mode de transmission a pour conséquence d'aligner les intérêts reproducteurs des partenaires et joue un rôle primordial dans l'évolution des associations vers le mutualisme (Leigh, 2010; Werren and O'Neill, 1997; Yamamura, 1993). Il est souvent observé dans les symbioses obligatoires mais également dans beaucoup d'exemples de symbioses facultatives. Il est facilité par une localisation intra-hôte voire intracellulaire des symbiotes. La persistance de la transmission verticale sur plusieurs générations permet le maintien des associations sur de larges échelles de temps et peut générer des patrons de cospéciation (ou cocladogénèse) (Moran et al., 2008a).

De façon alternative les symbiotes sont transmis aux insectes de manière horizontale, c'est-à-dire acquis dans l'environnement à chaque génération (Bright and Bulgheresi, 2010). Pour de nombreux symbiotes facultatifs des insectes, les deux modes de transmission coexistent. En effet, l'absence de patrons de cospéciation observée pour de nombreux symbiotes suggère que la transmission verticale n'a pas lieu à chaque génération sur de larges échelles de temps. En ce qui concerne plus particulièrement les symbiotes associés aux insectes

phytophages, on suppose assez fréquemment que la transmission horizontale peut avoir lieu via la plante hôte (Caspi-Fluger et al., 2012). Mais les preuves expérimentales des événements de transmission sont assez rares et les études montrant ces événements sont peu répétées. Ce mode de vie et de transmission des symbiotes est à l'origine de multiples modifications chez ceux-ci (Moran et al., 2008a; Perreau and Moran, 2021; Wernegreen, 2002; Werren and O'Neill, 1997). Ces modifications sont visibles très rapidement dans les génomes des symbiotes.

II. Evolution des génomes endosymbiotiques bactériens

Les pressions de sélection qui s'exercent sur les endosymbiotes sont très particulières. En conséquence, leurs génomes sont fortement modifiés et façonnés d'une manière caractéristique par rapport à leurs proches apparentés. Ainsi, on observe chez ceux-ci une convergence vers des génomes fortement réduits, avec de nombreux gènes perdus par rapport à leurs parents à vie « libre », ils présentent également une composition nucléotidique biaisée avec un faible contenu en GC (c'est-à-dire des fréquences de Guanine et Cytosine) (Degnan et al., 2009; McCutcheon et al., 2019; McCutcheon and Moran, 2012; Moran et al., 2008a; Perreau et al., 2021; Perreau and Moran, 2021; Toft and Andersson, 2010; Wernegreen, 2015, 2002). Dans la diversité existante des systèmes endosymbiotiques décrits à ce jour, le degré de réduction génomique, les taux de GC, ainsi que les catégories fonctionnelles des gènes perdus varient grandement et peuvent dépendre de la fonction apportée par le symbiote (Chong et al., 2019; Moran et al., 2008a; Vasquez and Bennett, 2022; Wernegreen, 2002). Ils dépendent également fortement du mode de transmission de ces symbiotes. Le degré de réduction (voir de dégradation) des génomes peut alors être vu comme reflétant un gradient de l'organisme libre, à l'association facultative jusqu'au symbiote obligatoire à transmission verticale (et même jusqu'à l'organite) et ainsi être un indicateur de la nature de l'association hôte-symbiote (McCutcheon and Moran, 2010, p. 201; Toft and Andersson, 2010) (Figure 1).

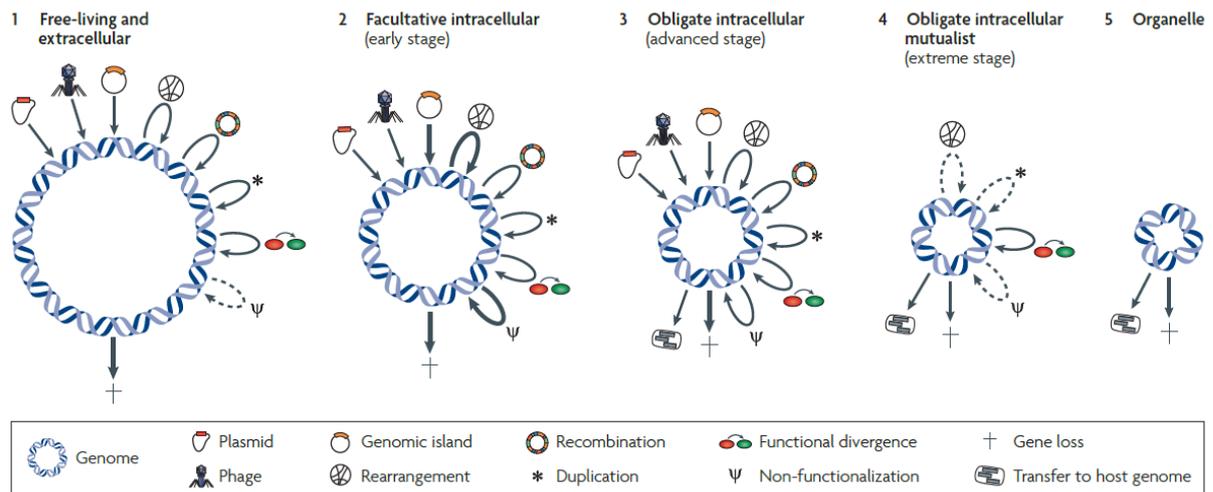


Figure 1 : Figure reprise de la publication de Toft & Andersson (2010). Montrant les différents étapes d'adaptation à la vie intra-hôte d'une bactérie libre à un organelle.

Dans le cadre de ma thèse et compte tenu que la grande majorité d'études a été conduite sur les symbioses nutritionnelles, les exemples utilisés par la suite dans cette introduction concerneront celles-ci. En effet, grâce à l'étude de différentes lignées d'endosymbiotes bactériens particulièrement étudiés dans les symbioses nutritionnelles nous connaissons aujourd'hui les principaux mécanismes de réduction et de spécialisation de leur génome (résumés sur la figure 2).

Les changements génomiques observés ne sont pas toujours aléatoires, certains sont directement liés au mode de vie particulier des bactéries. Les gènes associés à la virulence, la défense, la réponse aux stress environnementaux, à l'enveloppe cellulaire, et autres fonctions qui sont indispensables pour les lignées bactériennes au mode de vie libre, deviennent « inutiles » dans l'environnement confiné et stable des bactériocytes (Moran et al., 2008a; Moya et al., 2008; Wernegreen, 2015). Par exemple, les gènes codant pour l'enveloppe cellulaire sont potentiellement compensés par des gènes du génome nucléaire de l'hôte, (Husnik et al., 2013; Smith et al., 2022). Par conséquent, ces gènes ne sont plus sous sélection purifiante et on observe une pseudogénéisation et une perte des gènes associés à ces fonctions et cela de façon indépendantes chez divers endosymbiotes **sous l'action de la dérive** (Bennett and Moran, 2015; McCutcheon et al., 2019). La dérive agit également sur l'ensemble du génome des endosymbiotes du fait de leur démographie. En effet, la vie intra-hôte induit une réduction de la taille efficace de la population des endosymbiotes bactériens (N_e) par rapport à des bactéries à stades de vie libre. Par ailleurs, lorsque les endosymbiotes sont transmis de façon verticale, seulement une partie des bactéries est transmise à la descendance, cela crée des goulets d'étranglement successifs dans les populations de symbiotes à chaque événement de

reproduction qui augmentent l'effet de la dérive. Probablement sous l'effet de cette dérive, on observe que les gènes responsables de la réparation de l'ADN ou des mécanismes de réplication de l'ADN ont eux aussi été perdus indépendamment dans plusieurs lignées bactériennes endosymbiotique (Moran et al., 2008a; Moran and Wernegreen, 2000). Ces pertes ainsi que l'absence de recombinaison due à la ségrégation des bactéries dans les bactériocytes, renforcent l'accumulation de mutations neutres ou délétères dans des séquences codantes et les rendent non-fonctionnelles, augmentant le nombre de pseudogènes dans les génomes (un phénomène que l'on nomme cliquet de Muller) (Gabriel et al., 1993; McCutcheon and Moran, 2012; Moran, 1996; Moran et al., 2009), mais aussi l'interdépendance entre les bactéries et leur hôte. Suite à la pseudogénisation d'une grande partie du génome sous l'effet de la dérive, les mécanismes conduisant à la réduction de la taille des génomes seraient dus à un biais mutationnel en faveur des délétions chez l'ensemble des Procaryotes (Danneels et al., 2018; Kuo and Ochman, 2009; Mira et al., 2001), bien que ce biais soit difficile à mettre en évidence (Kuo and Ochman, 2009; Lynch, 2006). Compte tenu du relâchement de la sélection sur une grande partie du génome (pseudogénisé par l'accumulation de mutations), ce biais conduirait à éliminer de nombreuses séquences. Les mécanismes à l'origine de cette diminution de la taille des génomes ne seraient donc pas un processus adaptatif des bactéries (Mira et al., 2001). De la même façon, les faibles taux de G-C observés dans les génomes des endosymbiotes semblent avoir pour origine un biais favorisant l'adénine et la thymine (A-T). En effet, le relâchement des pressions de sélection couplé à une détérioration de la machinerie de réparation de l'ADN semble pouvoir causer un appauvrissement des génomes en GC %, par un arrêt de correction de la désamination des Cytosines en Uracile et de l'oxydation des Guanines (Klasson and Andersson, 2006; Lind and Andersson, 2008).

Par opposition, les gènes qui sont impliqués dans la symbiose sont sous forte sélection purifiante. C'est le cas des gènes impliqués dans la synthèse des acides aminés essentiels et des vitamines dans les symbioses nutritionnelles des Hémiptères (Chong et al., 2019; Clayton et al., 2016; Hansen and Moran, 2011; Tamas et al., 2002; Toh et al., 2006). Ceux-ci ont en effet, un rôle majeur de complémentation de l'hôte au sein de la symbiose et sont donc indispensables au bon fonctionnement de celle-ci et donc à la survie de chacun des partenaires de la symbiose (Akman Gündüz and Douglas, 2009; Blow et al., 2020; Douglas, 1998b; Russell et al., 2014). Il existe alors une forte pression de sélection purificatrice qui agit sur ces gènes (Chong et al., 2019; McCutcheon et al., 2009; Tamas et al., 2001; Vasquez and Bennett, 2022). Néanmoins, l'effet de la dérive est tellement fort que certains de ces gènes sont tout de même perdus. En

conséquence de ces pertes de gènes et de leur réduction, les génomes d'endosymbiotes obligatoires sont les plus petits génomes bactériens connus (seulement 132 kb chez *Nasuia* (Bennett and Moran, 2013)), avec une forte proportion de séquences codantes et principalement spécialisées dans les fonctions symbiotiques (Luan et al., 2015; Manzano-Marín et al., 2015; Manzano-Marín et al., 2020a; Mao et al., 2018; McCutcheon and von Dohlen, 2011; Tamas et al., 2001; Weglarz et al., 2018; Zientz et al., 2004) (Figure 2).

Dans certain cas, la dérive peut cependant amener à la perte de certains gènes aux fonctions essentielles à la survie de l'hôte. Dans ce cas particulier le symbiote obligatoire primaire peut être remplacé par un nouveau micro-organisme déjà présent dans l'environnement et capable de remplir les fonctions perdues par le symbiote primaire. Une autre possibilité observée chez plusieurs systèmes symbiotiques est l'acquisition d'un co-symbiote obligatoire qui lui aussi compense les fonctions perdues par le symbiote primaire. Dans de tels systèmes di-symbiotiques, l'hôte dépend de deux partenaires bactériens obligatoires (Bennett and Moran, 2015).

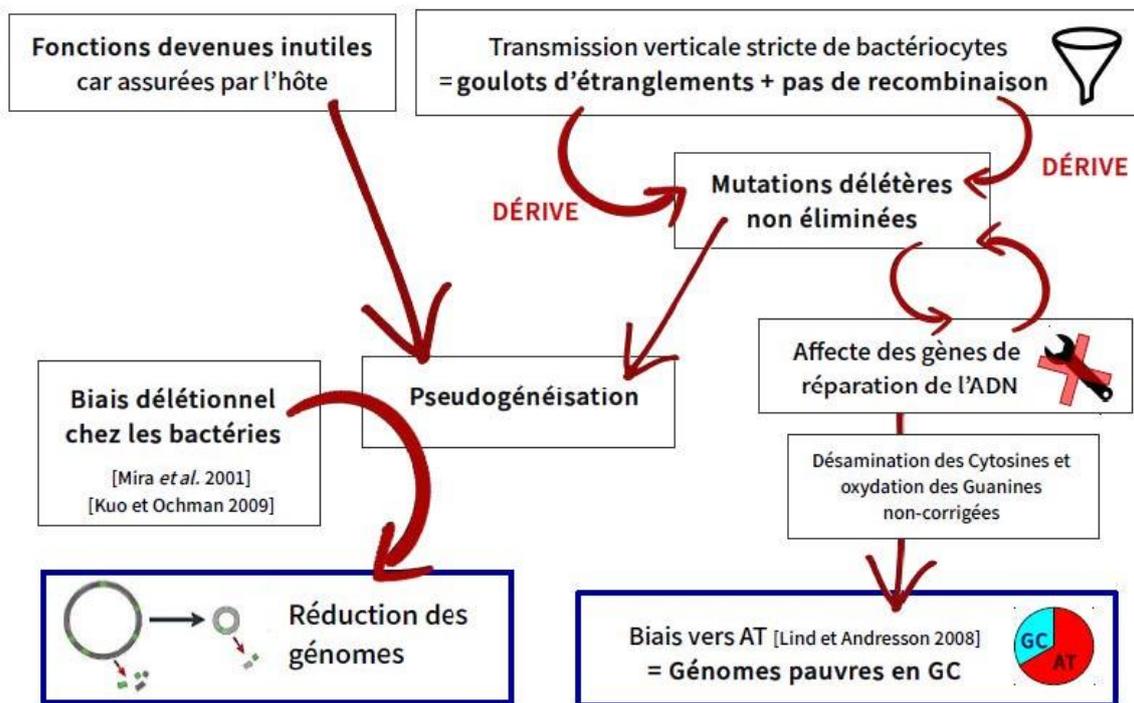


Figure 2 : Résumé schématique des différentes forces sélectives et étapes menant à la réduction des génomes endosymbiotiques (adapté d'après Annonier 2018)

III. Les systèmes di-symbiotiques dans lesquels les symbiotes assurent la nutrition de leurs hôtes

La dépendance à l'égard de multiples symbiotes héréditaires obligatoires se retrouve chez de nombreux insectes, reflétant une acquisition successive au cours de l'évolution de leurs hôtes. Les processus impliqués dans la mise en place de telles symbioses sont encore peu décrits. Cependant, il semble que dans la majorité des exemples décrits à ce jour soit impliquée dans des complémentations nutritionnelles suggérant une acquisition de co-symbiote obligatoire pour compenser des pertes de fonctions du symbiote primaire (Douglas, 2016; Duron and Gottlieb, 2020; Husnik and McCutcheon, 2016). Cette complémentation nutritionnelle suggère que les interactions entre endosymbiotes sont importantes et peuvent jouer un rôle majeur dans le façonnement des traits fonctionnels et de l'histoire évolutive de ces endosymbioses multipartenaires. Les symbiotes co-obligatoires ont évolué à partir de différentes origines telles que des pathogènes, des bactéries environnementales, des symbiotes facultatifs ou même des bactéries phytopathogènes (Douglas, 2016; Husník et al., 2011). De la même manière que les symbiotes primaires, on les retrouve localisés dans des bactériocytes, le plus souvent différents de ceux qui contiennent les symbiotes primaires (Douglas, 2016) et de manière plus inhabituelle, imbriqués dans le symbiote primaire comme chez certaines cochenilles (McCutcheon and von Dohlen, 2011). Leur transmission se fait aussi de façon verticale similairement au symbiote primaire et en conséquence, on observe chez eux des génomes réduits avec de forts taux de GC (voir partie précédente).

Un des exemples les mieux décrits de symbiose double est celui des Cicadelles, chez qui le symbiote primaire *Sulcia* cohabite avec le co-symbiote *Baumannia* et se complémentent mutuellement dans leur capacité à fournir des nutriments essentiels à leurs hôtes (McCutcheon and Moran, 2007; Wu et al., 2006). Ces deux symbiotes se sont co-diversifiés avec leurs hôtes sur une période représentant des millions d'années (Takiya et al., 2006). Plus généralement il semble que les symbioses multipartenaires soient très répandues chez les hémiptères (Bennett and Moran, 2015), ainsi chez certains insectes hématophages tels que les tiques (Buisse et al., 2021) ou encore la mouche tsétsé (Nováková and Hypša, 2007; Rio et al., 2006).

Parmi les divers exemples de symbioses multi-partenaires décrites aujourd'hui, il semble exister une multitude de cas de remplacements ou acquisitions indépendantes des symbiotes secondaires (Bennett and Moran, 2015; Dial et al., 2022; Husnik and McCutcheon, 2016; Koga et al., 2013; Meseguer et al., 2017; Michalik et al., 2021; Rosenblueth et al., 2012; Szabó et al.,

2020, 2017; Elena R. Toenshoff et al., 2012; Urban and Cryan, 2012). En effet, on observe dans certains groupes une forte diversité de souches bactériennes présentes comme co-symbiotes (Bennett and Moran, 2015; Meseguer et al., 2017). Par opposition, quelques exemples montrent l'existence d'une co-spéciation à long terme (Manzano-Marín et al., 2020a; Vasquez and Bennett, 2022), mais aucune étude n'a encore permis de déterminer quels facteurs sont impliqués dans ces différences de dynamique de remplacement symbiotique.

IV. Les pucerons (Hemiptera : Aphididae)

Les Aphides (Aphididae Latreille, 1802) sont un groupe appartenant à l'ordre des hémiptères ; il comprend plus de 5000 espèces réparties en 24 sous-familles (<http://aphid.speciesfile.org>) (Favret et al., 2015; Peccoud et al., 2010). Ils sont caractérisés par un cycle de vie particulier avec une alternance de reproduction sexuées et asexuée selon les conditions environnementales (Figure 3).

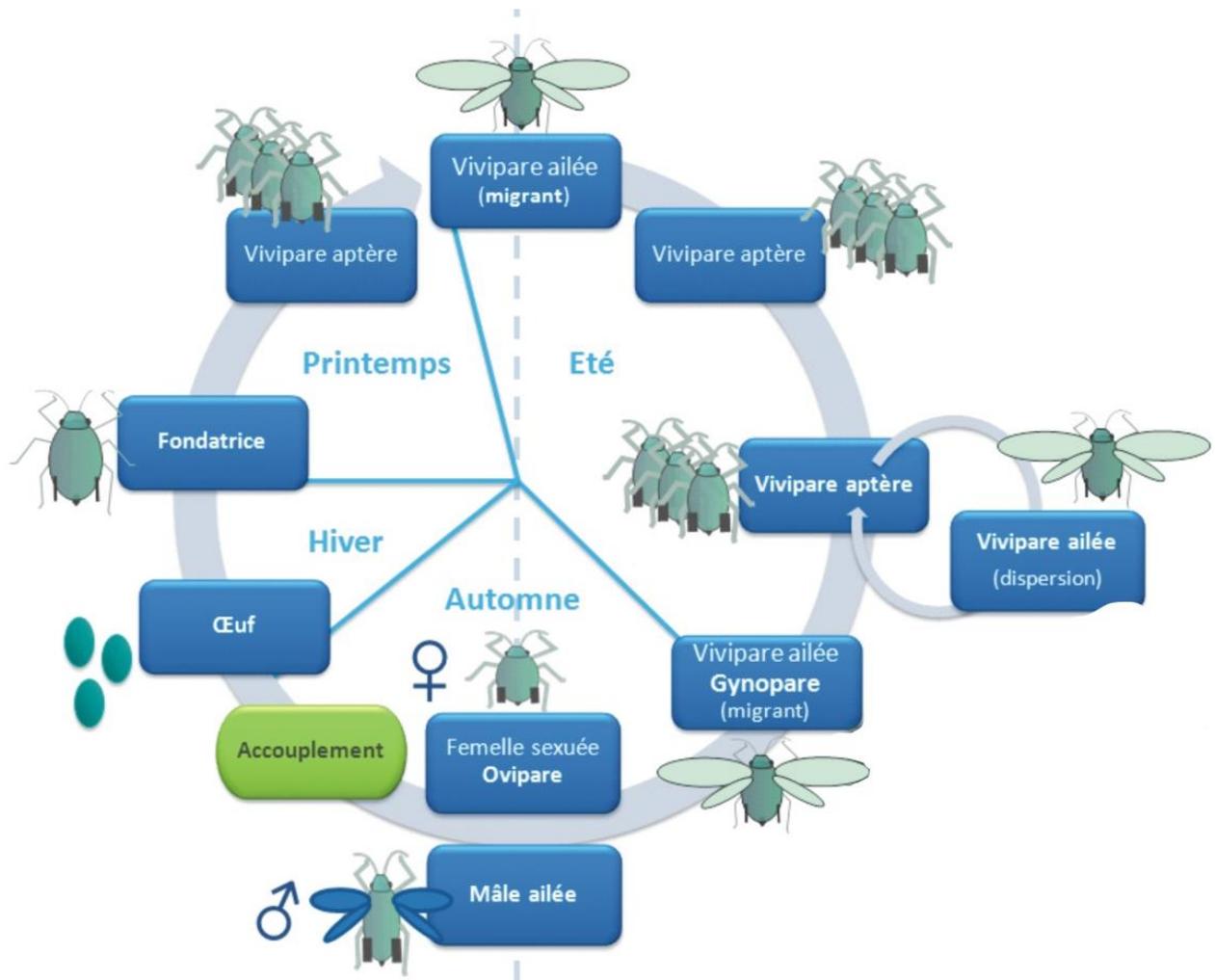


Figure 3: Le cycle de vie classique d'un puceron, la parthénogénèse cyclique, consiste en une succession de phases de reproduction sexuée et asexuée. La phase de reproduction asexuée a lieu durant la belle saison (printemps et été) et consiste en plusieurs générations parthénogénétiques apomictiques (sans recombinaison) : des femelles asexuées vivipares produisent des descendants génétiquement identiques à elles-mêmes.

Ce sont des insectes exclusivement phytophages et la plupart des espèces sont très spécifiques vis-à-vis de leur gamme de plantes hôtes. Il existe cependant des espèces polyphages, on compte parmi elles des ravageurs des cultures notoires tels *Myzus persicae*. Leur alimentation constituée de sève élaborée (sève circulant dans le phloème), contient une faible densité d'acides aminés essentiels et de vitamines, ils sont donc associés à des bactéries intracellulaires obligatoires qui supplémentent leur alimentation (Akman Gündüz and Douglas, 2009; Douglas, 1992; Mittler, 1971; Nakabachi and Ishikawa, 1999; Poliakov et al., 2011; Prosser and Douglas, 1991; Shigenobu and Wilson, 2011; Wilkinson and Ishikawa, 2000). Chez la plupart des espèces de pucerons c'est une gama-protéobactérie *Buchnera aphidicola* qui est retrouvée comme symbiote nutritionnel (Buchner, 1953a). Ces bactéries sont retrouvées dans des cellules spécialisées de l'hôte, les bactériocytes, qui sont eux même regroupés pour former un organe : le bactériome (Figure 4). Leur caractère obligatoire et leur localisation particulière sont associés à une transmission exclusivement verticale. Lors des épisodes de reproduction clonale le transfert des *Buchnera* a lieu d'un bactériocyte de la mère à un embryon en développement (Braendle et al., 2003; Koga et al., 2012; Michalik, 2010; Michalik et al., 2014; Wilkinson et al., 2003).

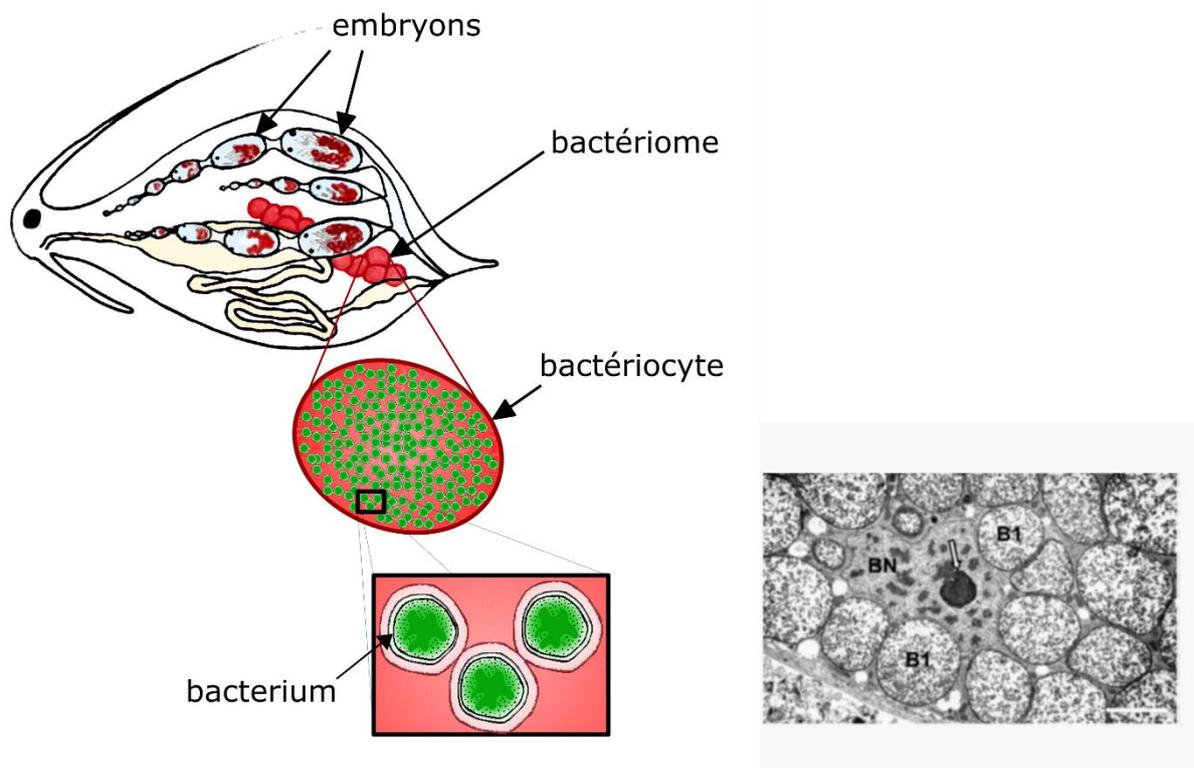


Figure 4 : A gauche, schéma d'une coupe transversale de puceron montrant les bactériocytes contenant *Buchnera aphidicola*, la chaîne embryonnaire dont les embryons contiennent déjà leur propre bactériome. A droite, une photo en TEM reprise de Michalik et al. 2014, montrant chez *Pemphigus spyrothecae* (Eriosomatinae) un fragment d'un bactériocyte B1 – bacterium *Buchnera aphidicola*, BN – nucleus du bactériocyte, (barre en blanc = 4µm)

Cette ségrégation intracellulaire de *Buchnera* implique que celle-ci soit régulée par son hôte afin d'éviter l'invasion ou l'augmentation du coût de sa présence (Whittle et al., 2021). Comme il a été décrit chez *Acyrtosiphon pisum*, la dynamique intra-hôte de la bactérie endosymbiotique *Buchnera* semble corrélée aux besoins nutritionnels du puceron. En effet, il semble y avoir une forte augmentation de la densité du symbiote primaire jusqu'au dernier stade larvaire puis une diminution lente à l'âge adulte qui s'accélère après la dernière reproduction (Lu et al., 2014; Nishikori et al., 2009; Simonet et al., 2016a). De plus, cette dynamique semble être totalement sous le contrôle de l'hôte avec une régulation du nombre de bactériocytes via la dégénérescence cellulaire et la dégradation de *Buchnera* par le système lysosomal de la cellule (Simonet et al., 2018b, 2016b).

Depuis le séquençage du premier génome de *Buchnera* (Shigenobu et al., 2000a), de nombreuses autres lignées ont été séquencées et étudiées, révélant que celles-ci avaient un génome relativement réduit (de 420 Mb à 644 Mb, contenant de 381 à 644 gènes), (Manzano-marín, 2022), des taux de G-C bas (de 22 à 27% de GC) et préservaient dans une large mesure une synténie. De plus, les études phylogénétiques faites sur *Buchnera* montrent qu'il existe une congruence quasi parfaite avec la phylogénie de l'hôte. On estime que celle-ci co-évolue depuis 84-164 millions d'années avec les pucerons, en faisant ainsi l'endosymbiote nutritionnel acquis ancestralement par le groupe (Douglas, 1998a; Jouselin et al., 2009; Moran, 1993; Sandström and Moran, 1999; Von Dohlen and Moran, 2000). Compte tenu de ces études, il semble qu'une réduction rapide du génome et une perte des éléments mobiles ait eu lieu avant la diversification des pucerons (Tamas et al., 2002; van Ham et al., 2003).

De nombreux symbiotes facultatifs ont également été mis en évidence (Shigenobu et al., 2000a) chez les pucerons (Oliver et al., 2010). Ces symbiotes facultatifs ne sont pas nécessaires à la survie de l'hôte mais peuvent tout de même conférer des avantages aux individus qui les hébergent dans certaines conditions environnementales (Burke et al., 2009; Henry et al., 2013; Oliver et al., 2010, 2003; Sakurai et al., 2005). Par exemple, la bactérie endosymbiotique *Hamiltonella defensa*, connue pour être présente chez plusieurs espèces d'hémiptères, confère une protection contre les guêpes parasitoïdes chez les pucerons (Brandt et al., 2017; Oliver et al., 2009). Cette protection est apportée par des gènes de toxines portés par des phages : les APSEs (pour *Acyrtosiphon pisum secondary endosymbiont*) premièrement décrit chez les endosymbiotes du puceron du pois (Asplen et al., 2014; Degnan and Moran, 2008; Hopper et al., 2018; Lenhart and White, 2017; Leybourne et al., 2020; Moran et al., 2005; van der Wilk et al., 1999).

V. Les pucerons du genre *Cinara*

Le genre *Cinara* Curtis, 1835 (Hemiptera : Aphididae : Lachninae) représente plus de la moitié de la diversité de la famille des Lachninae, avec plus de 248 espèces décrites, il est le deuxième genre de pucerons le plus diversifié et l'un des rares à se nourrir sur des conifères (Chen et al., 2016; Favret et al., 2015; Meseguer et al., 2015) (Figure 5). En effet, ce groupe semble s'être diversifié par une radiation rapide à la fin du Cénozoïque (45 Ma), après le passage d'un ancêtre se nourrissant d'angiospermes à un hôte gymnosperme (Chen et al., 2016). Les espèces de *Cinara* décrites jusqu'à présent se rencontrent presque exclusivement sur deux familles de plantes, les Pinaceae et les Cupressaceae ; la plupart des espèces se nourrissent sur *Pinus*, *Abies* et *Picea*, et présentent un haut niveau de spécificité envers une ou quelques espèces hôtes d'un de ces genres (Figure 5). De précédents travaux moléculaires sur les membres du genre *Cinara* ont confirmé que ce dernier était monophylétique, et ont divisé le genre en trois clades phylogénétiques majeurs, appelés simplement A, B et C (Meseguer et al., 2017, 2015).

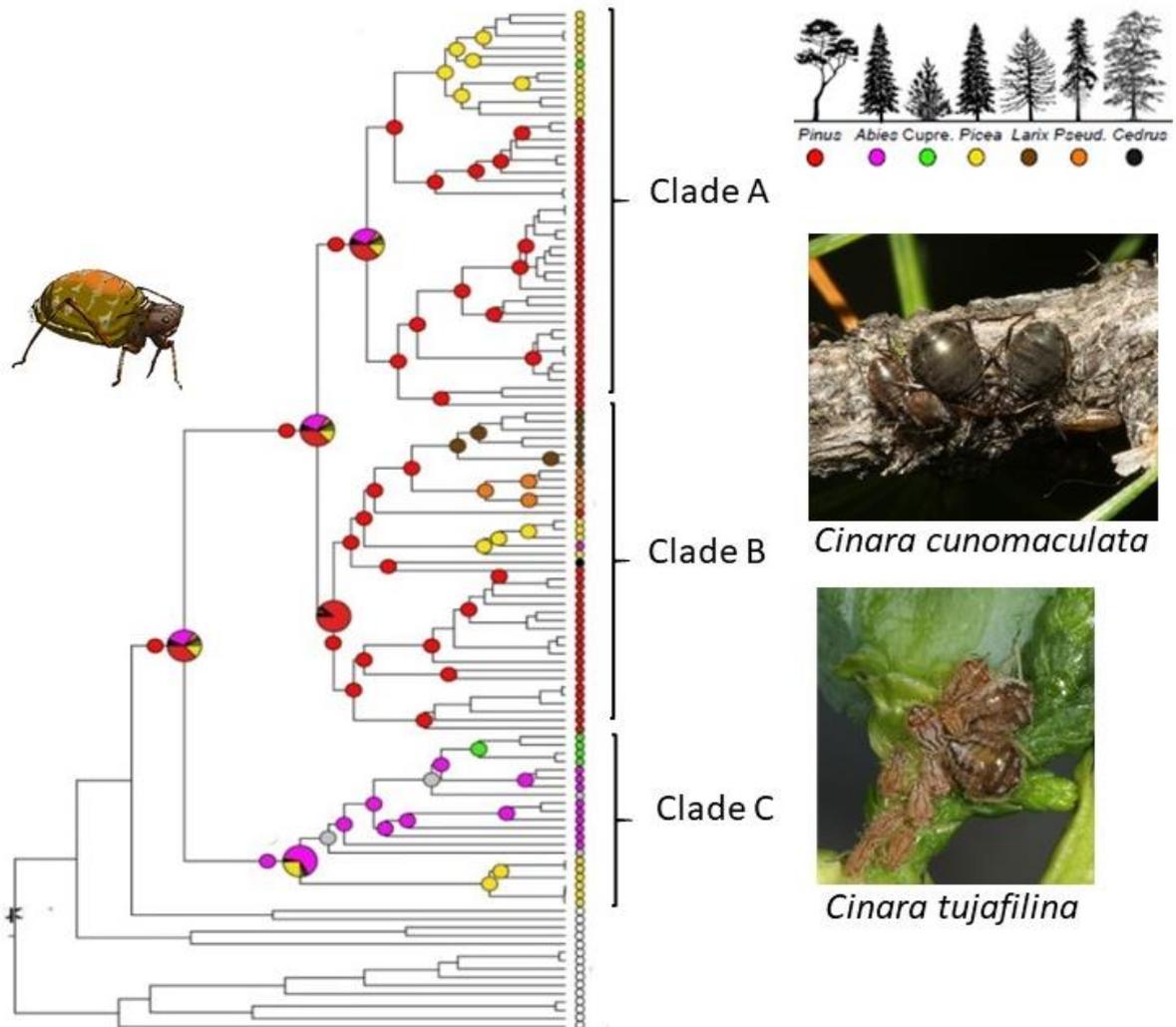


Figure 5 : A gauche représentation de la phylogénie des *Cinara* (à partir de 92 espèces) et de l'histoire des associations avec les genres hôtes (d'après Meseguer et al. 2015). Le dessin à gauche représente *Cinara splendens* (d'après Jorge Mariano Collantes Alegre).

Plusieurs études ont montré que *Buchnera* n'est pas le seul fournisseur d'acides aminés essentiels et de vitamines chez les *Cinara*, mais que dans cette sous famille, le puceron s'appuie sur un système di-symbiotique : un deuxième symbiote coexiste avec *Buchnera* et est essentiel pour la nutrition de l'hôte (Lamelas et al., 2008; Meseguer et al., 2017; Monnin et al., 2020). Grâce au séquençage de plusieurs génomes de *Buchnera* au sein de ce groupe (les génomes de *Buchnera* associées à 14 espèces de *Cinara* ont été publiés avant le début de ma thèse), il a été observé chez chacune des espèces étudiées une perte de la capacité de synthétiser la biotine (B7) et la riboflavine (B2) par *Buchnera* (Manzano-Marín et al., 2016; Manzano-Marín and Latorre, 2016; Meseguer et al., 2017; Pérez-Brocal et al., 2006). Par ailleurs, plusieurs lignées bactériennes coexistant avec *Buchnera* dans leur hôte ont été mises en évidence chez les *Cinara* (Manzano-Marín et al., 2017; Meseguer et al., 2017) grâce au séquençage du gène bactérien 16S rRNA (Figure 6).

Parmi ces partenaires endosymbiotiques de *Cinara* et *Buchnera*, on retrouve diverses gammaprotéobactéries *Fukatsua symbiotica*, et *Serratia symbiotica* (Manzano-Marín et al., 2020a; Meseguer et al., 2017) (Figure 6) ; qui sont habituellement trouvées comme symbiotes facultatifs des pucerons (Zytynska and Weisser, 2016). Ainsi que des espèces moins fréquemment retrouvés tel *Sodalis* ou des apparentées à des bactéries phytopathogènes telles *Erwinia*. Parmi ces symbiotes certains sont « fixés » à l'intérieur des espèces (c'est-à-dire qu'ils sont retrouvés chez tous les individus de l'espèce) (Meseguer et al., 2017). L'étude du génome de plusieurs de ces symbiotes et l'inférence de leur capacité métabolique chez différentes espèces de *Cinara* montrent que : 1) ces symbiotes ont un génome extrêmement réduit symptomatique de symbiose obligatoire 2) qu'ils ont les capacités métaboliques permettant de suppléer aux déficits de *Buchnera* (Figure 7).

Par ailleurs, plusieurs de ces symbiotes diffèrent de ceux retrouvés dans d'autres groupes de pucerons de par leur localisation et leur forme cellulaire (Buchner, 1953a; Fukatsu and Ishikawa, 1998; Manzano-Marín et al., 2017; Michalik et al., 2014; Pyka-Fosciak and Szklarzewicz, 2008). Ils sont en effet trouvés le plus souvent dans des bactériocytes indépendant de ceux de *Buchnera* que l'on retrouve localisés à proximité des bactériocytes de *Buchnera* (Manzano-Marín et al., 2017). La présence de cette diversité de co-symbiotes, leurs caractéristiques génomiques et phénotypiques montrent qu'ils forment des systèmes di-symbiotiques avec *Buchnera* et qu'ils se sont probablement établis à travers plusieurs acquisitions et remplacements indépendants tout au long de la diversification de leurs hôtes (Manzano-Marín et al., 2017; Meseguer et al., 2017).

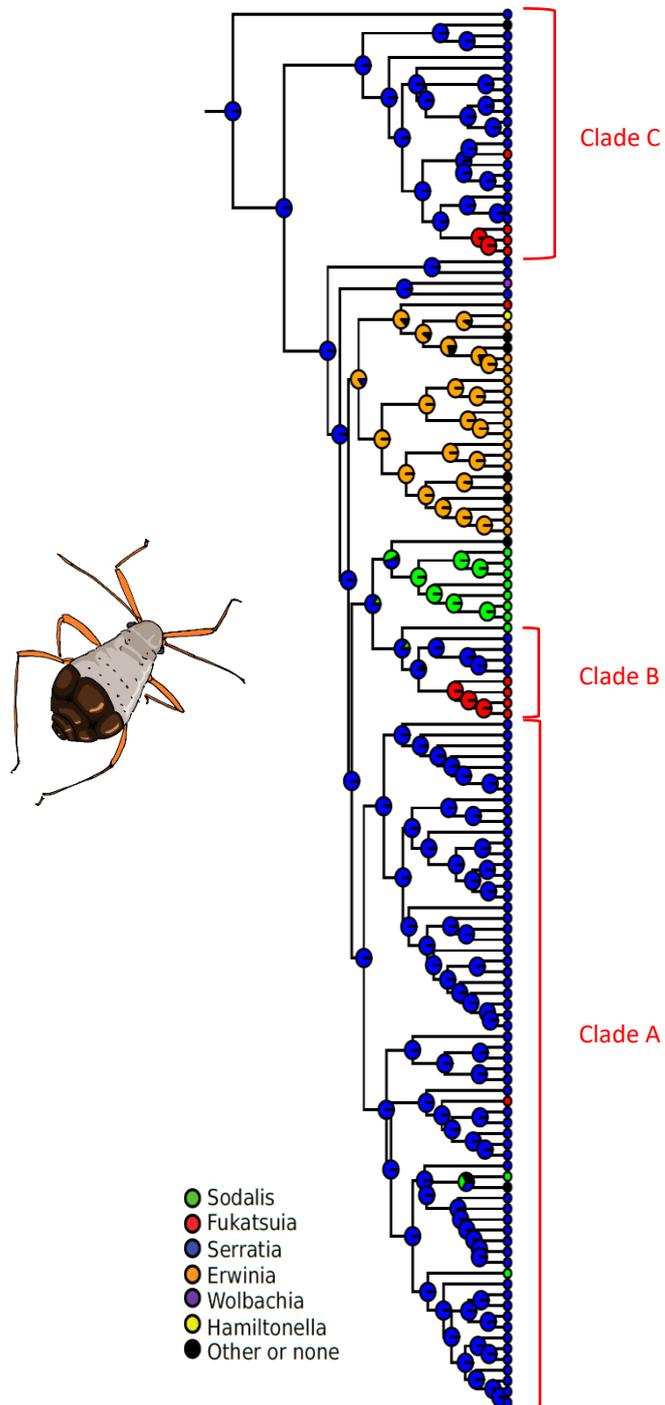


Figure 6 : A gauche représentation de la phylogénie des *Cinara* et de l'histoire de leurs associations avec leur co-symbiotes (d'après Meseguer et al. 2017). Le dessin à gauche représente *Cinara brauni* (d'après Jorge Mariano Collantes Alegre).

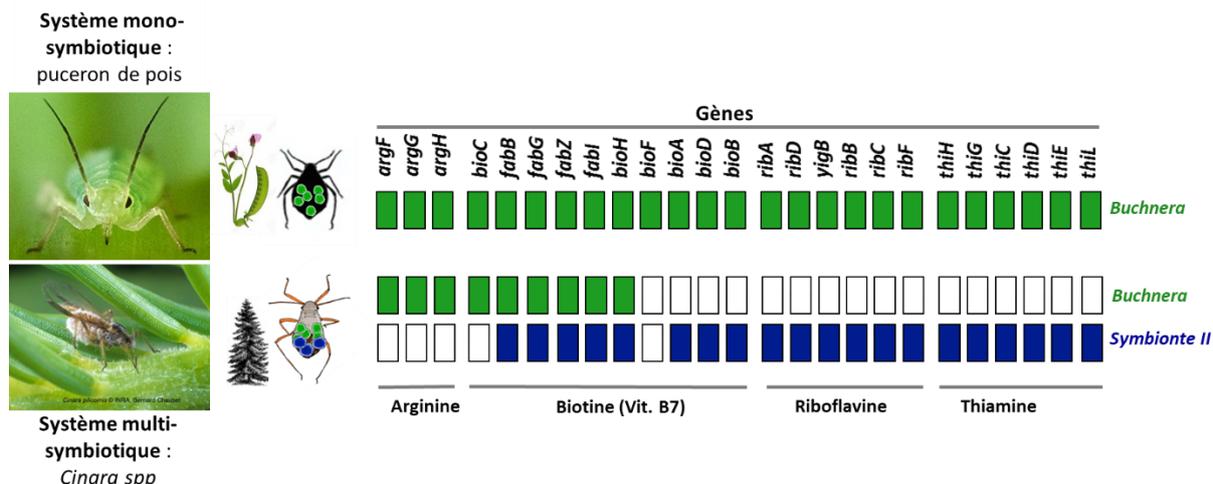


Figure 7 : Représentation schématique de la présence ou absence des différents gènes impliqués dans la biosynthèse des vitamines. À gauche photographie des différentes espèces de puceron, schéma de leur plante hôte et schéma du puceron.

Bien que différentes bactéries aient été trouvées associées aux *Cinara*, de nombreuses espèces de cette sous-famille sont associées à *Serratia symbiotica* (Burke et al., 2009; Lamelas et al., 2008; Manzano-Marín et al., 2017; Meseguer et al., 2017). Un scénario parcimonieux expliquant l'histoire de l'association obligatoire avec cette bactérie serait qu'elle a été acquise avant la diversification des *Cinara*, facilitée par une perte de la voie de biosynthèse de la riboflavine chez *Buchnera*. Elle aurait ensuite été perdue de nombreuses fois et remplacée par d'autres bactéries. Pourtant, l'histoire semble plus compliquée. Ces souches de *Serratia symbiotica*, forment deux groupes phylogénétiquement distincts, l'une phylogénétiquement très proche de *Serratia* que l'on trouve de manière facultative dans l'ensemble des pucerons (c'est le cas de *Serratia symbiotica* retrouvée chez *C. tujaefilina* avec une taille de génome de 2.5 Mb (Manzano-Marín and Latorre, 2014), l'autre qui a des caractéristiques génomiques de type "obligatoire" (taille de génome de 1 à 1.7 Mb) c'est le cas des souches de *Serratia* retrouvées chez *C. cedri* et *C. fornacula* (Lamelas et al., 2011a; Meseguer et al., 2017). Ces différences génomiques sont aussi retrouvées au niveau morphologique, les *Serratia* de *C. cedri* étant de petite taille et rondes alors que celles retrouvées chez *C. tujaefilina* sont de plus grande taille et ont une forme de bacille. De plus, il a été observé chez *C. tujaefilina* que *Serratia* n'était pas localisée uniquement dans des bactériocytes mais en périphérie des bactériocytes de *Buchnera*. L'ensemble de ces éléments semble donc être en faveur d'un scénario évolutif comportant plusieurs acquisitions indépendantes de *Serratia*. Compte tenu de la particularité de ce système *Buchnera-Serratia* chez les *Cinara*, celui-ci constitue un modèle intéressant pour étudier les

symbioses multipartenaires, leur évolution, leur mise en place et l'évolution des génomes bactériens.

VI. Objectifs

Afin d'approfondir l'étude des processus évolutifs associés à l'évolution des symbioses, ma thèse porte sur l'étude des symbioses obligatoires comportant deux bactéries observées chez les pucerons du genre *Cinara*. Mon projet se concentre sur l'origine évolutive de l'acquisition de nouvelles bactéries, en examinant la dynamique de remplacement de symbiote, les processus évolutifs menant à ces remplacements ainsi que la dynamique cellulaire des symbiotes.

Dans un premier chapitre, je présente les analyses menées pour reconstruire l'histoire évolutive de l'acquisition de co-symbiote obligatoire en me focalisant sur le couple de bactéries *Buchnera-Serratia* prédominant chez les *Cinara*. Pour ce faire, les génomes de symbiotes de 45 espèces de *Cinara* ainsi que les génomes mitochondriaux de leur hôte ont été *de novo* séquencés et assemblés (préalablement à ma thèse). A partir de ces génomes, j'ai reconstruit de façon indépendante la phylogénie des *Cinara*, de *Buchnera* et celle de *Serratia*. Nous avons alors retracé les différents scénarii possibles de l'histoire évolutive de *Serratia symbiotica* chez les *Cinara*, en répertoriant les différentes acquisitions et remplacements de symbiote ayant eu lieu. Pour cela des analyses de réconciliation ont été effectuées sur les phylogénies des différentes partenaires de la symbiose.

Dans le deuxième chapitre, je me suis focalisée sur treize espèces de *Cinara* formant un groupe monophylétique, hébergeant *Serratia symbiotica* et *Buchnera aphidicola* comme symbiotes et pour lesquelles il semble y avoir une diversification conjointe avec une parfaite co-spéciation. Pour celles-ci je me suis consacrée à caractériser les taux de substitution des deux bactéries, les changements génomiques, ainsi que les pertes de gènes ou les changements de régimes de sélection qui s'exercent sur les gènes des deux bactéries symbiotiques. En utilisant des datations obtenues dans des travaux précédents ma thèse, j'ai pu mesurer les taux d'évolution des deux symbiotes. L'ensemble des génomes bactériens ont été annotés et caractérisés et l'ensemble des pertes de gènes ayant eu lieu ont été relevées et décrites. Pour caractériser les changements de régime de sélection, plusieurs stratégies ont été utilisées afin de tester la présence de sélection positive, purifiante ou neutre sur chacun des gènes.

Dans le troisième chapitre, dans le but d'étudier si et comment le puceron hôte régule ses symbiotes dans un système di-symbiotique, je présente des analyses de quantification de symbiotes chez deux espèces de *Cinara* qui ont été effectuées à plusieurs stades de développement des pucerons. Pour cela, nous avons dû mettre en place un élevage de pucerons et développer un protocole d'échantillonnage des pucerons à différents stades et de quantification par qPCR de chacun des deux symbiotes. Pour cette partie, nous avons choisi deux espèces de *Cinara* (*C. palaestinensis* et *C. tujaefilina*) étant associées à *Serratia symbiotica* comme co-symbiote mais pour lesquelles les caractéristiques suggèrent des acquisitions indépendantes. En effet, les symbioses avec *Serratia* dans ces deux espèces diffèrent en terme de localisation, caractéristique cellulaire, taille de génome et âge d'acquisition. En parallèle, nous avons mené la même étude de quantification sur une espèce de puceron associée à *Erwinia* pour laquelle il y a eu une acquisition suivie de co-spéciation durant 20 Ma. Dans ce groupe des caractéristiques morphologiques, structurelles et génomiques similaires ont été mises en évidence chez toutes les *Erwinia* observées (Manzano-Marín et al., 2020a). La comparaison de la dynamique bactérienne chez ces modèles nous permet donc d'obtenir des premières informations quant au degré et au type de régulation des symbiotes par leur hôte, selon l'âge d'acquisition de la symbiose, et la souche bactérienne.

Finalement, dans un quatrième chapitre, une analyse de génomique comparative a été effectuée sur des phages APSE présents dans les co-symbiotes *Hamiltonella* de *Cinara* et procurant une défense contre les parasitoïdes. Afin de comprendre la diversité et la trajectoire évolutive de ce phage défensif, nous avons assemblé et annoté les génomes complets de 16 phages APSE et effectué des analyses phylogénétiques sur *Hamiltonella* et ses phages APSE associés. Les différents réarrangements du génome des APSE et l'évolution des toxines présentes ont été décrits.

Chapitre 1
Histoire évolutive de l'acquisition de *Serratia*
***symbiotica* comme co-symbiote obligatoire dans**
le genre *Cinara*.

(projet d'article, rédigé pour l'instant en français)

Contexte de l'étude

Cette étude a émergé de la mise en évidence de plusieurs bactéries co-symbiotique obligatoire chez les pucerons du genre *Cinara*, parmi lesquelles *Serratia symbiotica* est la plus représentée. La description des caractéristiques génomiques et structurales, ainsi que la reconstruction phylogénétique des pucerons hôtes couplée à l'identification par 16S de leurs bactéries symbiotiques, suggèrent plusieurs acquisition et/ou remplacement du co-symbiotes *Serratia* au cours de la diversification des *Cinara*. Cependant, aucune étude à ce jours n'a décrit l'histoire évolutive de l'association des *Cinara* et de leur co-symbiote *Serratia*.

Dans ce cadre-là, un large projet d'échantillonnage de 45 espèces de *Cinara* répartie dans le monde a été effectué préalablement à ma thèse afin de disposer de données représentatives de la diversité de se groupe. Pour chacune de ces espèces, un séquençage des symbiotes et des mitochondries a été effectué et l'ensemble des génomes bactériens et mitochondriaux ont été assemblé eux aussi préalablement à ma thèse.

Ce premier chapitre présente les analyses effectuées à partir de ces génomes symbiotiques et mitochondriaux, ayant pour objectif de décrire l'histoire évolutive de l'acquisition de *Serratia* comme co-symbiote au sein des *Cinara*. Cela permettra d'estimer le nombre d'acquisitions et/ou de pertes indépendantes de *Serratia* en tant que co-symbiotes obligatoires et ainsi inférer le nombre de transitions de bactéries facultatives vers des bactéries obligatoires ayant eu lieu lors de la diversification de l'hôte. Pour cela, les objectifs de cette partie sont donc de : **1)** Reconstruire de façon indépendante la phylogénie des *Cinara*, de *Buchnera* et celle de *Serratia*. **2)** Retracer les différents scenarii possibles d'acquisitions et de remplacements de symbiote par des analyses de réconciliation. **3)** Décrire la localisation des bactéries dans le puceron **4)** Etudier la structure des génome endosymbiotique par une description de ceux-ci en terme de taille de génome, taux de GC et nombre de CDS.

Les hypothèses de cette étude sont que la distribution chez les *Cinara* de *Serratia symbiotica* comme symbiote obligatoire est issu de plusieurs acquisitions indépendantes avec transition d'un mode de vie facultatif vers obligatoire. On suppose que ces acquisitions suivies de transition de mode de vie sont associées à une convergence évolutive de la structure de la symbiose. Mais aussi à une convergence de la structure des génomes endosymbiotiques avec des tailles de génome réduites, des taux de GC bas et un faible nombre de CDS.

I. Introduction

Depuis la découverte des systèmes symbiotiques obligatoires chez les arthropodes, de nombreuses études ont cherché à caractériser ces relations que ce soit leur fonctionnement, leur origine ou leur évolution. La grande majorité des symbioses obligatoires décrites sont des associations avec des bactéries à transmission verticale, généralement intracellulaires parfois localisées dans des cellules spécialisées de l'hôte, les bactériocytes (Braendle et al., 2003; Michalik et al., 2014; Wilkinson et al., 2003). Parmi les endosymbioses obligatoires les associations avec des bactéries qui synthétisent des nutriments essentiels absents de l'alimentation semblent être prédominantes. C'est notamment le cas des arthropodes se nourrissant de sang ou de sève, qui sont des aliments pauvres en vitamines B. Les endosymbioses sont donc intégrées à l'hôte jusqu'au cœur de ses cellules et de sa reproduction, mais ces modes de vie affectent aussi fortement les pressions de sélection sur les symbiotes et façonnent leurs génomes. En effet, des traits tels que la taille des génomes, les remaniements qui s'y opèrent ou le taux en GC sont affectés par le mode de vie particulier des endosymbiotes. On peut alors observer un gradient dans les caractéristiques des génomes bactériens des organismes libres jusqu'aux symbiotes obligatoires avec pour stade intermédiaire les associations facultatives. Ainsi, ces caractéristiques génomiques peuvent être de bons indicateurs de la force de l'association hôte-symbiote (McCutcheon and Moran, 2012; Toft and Andersson, 2010).

Dans certains cas, la dégénérescence des génomes peut amener à la perte de gènes aux fonctions essentielles à la symbiose. En effet, la perte de gènes tels que ceux impliqués dans la biosynthèse des acides aminés essentiels ou des vitamines par les symbiotes obligatoires peut mettre en péril la survie de l'hôte et donc de ses symbiotes obligatoires. Dans ces systèmes particuliers on peut alors observer l'existence de remplacements de symbiotes primaires par une nouvelle bactérie possédant encore toutes les capacités métaboliques nécessaires à compléter l'hôte (Bennett and Moran, 2015). Mais on retrouve également des cas dans lesquels de nouveaux symbiotes sont acquis par l'hôte formant ainsi des systèmes multi-symbiotiques. Dans ces systèmes, l'hôte dépend alors de deux partenaires bactériens obligatoires pour assurer son alimentation, ceux-ci pouvant se compléter pour compléter les voies de biosynthèse des nutriments. Les cas de remplacements de symbiotes et/ou de systèmes multi-symbiotiques ont été observés dans plusieurs groupes d'arthropodes (voir synthèses de (Bennett and Moran, 2015; Moran et al., 2008).

Dans ce contexte, l'une des questions évolutives les plus fondamentales concernant la symbiose insecte-bactérie est l'origine et les relations phylogénétiques des différentes lignées symbiotiques. En effet, cette connaissance est indispensable à la compréhension de la dynamique et des mécanismes d'établissement et de maintien des symbioses. Chez les pucerons (Hemiptera : Aphididae), l'association avec le symbiote primaire *Buchnera aphidicola* (γ -protéobactérie) est très stable, celle-ci est maintenue depuis son acquisition estimée à près de 80 à 150 millions d'années (Dohlen and Moran, 2000). Cette association, est indispensable pour la compensation des carences alimentaires de l'hôte qui se nourrit exclusivement de sève élaborée (Bermingham et al., 2009; Blow et al., 2020; Buchner, 1953; Douglas, 1998)). Comme décrit précédemment, le génome de *Buchnera* est fortement réduit, comptant seulement 0.6 millions de paires de bases chez le puceron du pois, *Acyrtosiphon pisum* (Shigenobu et al., 2000). Cependant, dans la sous-famille des Lachninae l'analyse des génomes de différentes lignées de *Buchnera* (associées à 15 espèces dans le genre *Cinara* et une espèce du genre *Tuberolachnus*, voir Table 1, S1) a permis de mettre en évidence la perte de plusieurs gènes impliqués dans la synthèse de certaines vitamines (la Biotine, la Riboflavine) et d'acides aminés essentiels indispensables à la survie des pucerons hôtes (Lamelas et al., 2011; Manzano-Marín et al., 2017; Manzano-Marín and Latorre, 2014; Manzano-Marín et al., 2020; Meseguer et al., 2017). De plus, dans cette lignée de pucerons, les génomes de *Buchnera* avec 0.44 Mb en moyenne (Gil et al., 2002; Manzano-Marín et al., 2020) sont plus petits que ceux de la souche associée aux pucerons du pois. Ces systèmes comportent des co-symbiotes obligatoires compensant les fonctions perdues par *Buchnera* (Lamelas et al., 2011, 2008; Manzano-Marín et al., 2017; Manzano-Marín et al., 2020; Meseguer et al., 2017).

Les deux tiers de la diversité des Lachninae appartiennent au genre *Cinara* qui compte plus de 200 espèces (Chen et al., 2016; Meseguer et al., 2015). Dans ce groupe, la caractérisation de la diversité des co-symbiotes selon les lignées hôtes a révélé la présence de différentes bactéries suggérant un système dynamique aux multiples remplacements et/ou acquisitions indépendantes (Meseguer et al., 2017). Parmi les co-symbiotes obligatoires décrits chez les *Cinara*, le plus fréquent est la γ -protéobactérie Enterobacteriaceae *Serratia symbiotica*. Cette bactérie connue pour être présente chez de nombreuses espèces de pucerons en tant que symbiote facultatif (Zytynska and Weisser, 2016) semble avoir subi une transition vers une forme obligatoire dans de nombreuses lignées de *Cinara*. L'étude des génomes de quatre lignées de *Serratia* co-obligatoires chez des espèces de *Cinara* montrent que celles-ci ont les capacités métaboliques pour compléter les fonctions perdues par *Buchnera* (Lamelas et al.,

2011; Manzano-Marín et al., 2018, 2016; Manzano-Marín and Latorre, 2016; Meseguer et al., 2017). L'histoire évolutive de l'acquisition de *Serratia* comme co-symbiote obligatoire au cours de la spéciation du genre *Cinara* est cependant encore mal connue. En effet, les différentes études génomiques existantes sur *Serratia* semblent montrer l'existence de plusieurs lignées aux origines et caractéristiques différentes. On retrouve ainsi la présence d'une *Serratia* intracellulaire au génome hautement réduit chez *C. cedri*, *C. fornacula* ainsi que chez *Tuberolachnus salignus* (Lachninae proche des *Cinara*) (Lamelas et al., 2011; Manzano-Marín et al., 2016; Meseguer et al., 2017) (Table 1). Ces données pourraient suggérer une acquisition ancestrale de cette bactérie dans la sous-famille. Cependant, la présence d'une *Serratia* présentant un génome plus gros que celle de *C. cedri* chez *C. tujafilina* ainsi que chez *C. strobi* avec une forme cellulaire filamenteuse plutôt que ronde, et hébergée à la périphérie des bactériocytes de *Buchnera* rappelle les formes facultatives (Manzano-Marín and Latorre, 2016). De plus, l'existence d'autres lignées bactériennes présentes comme co-symbiote obligatoire (par exemple, *Erwinia*, *Sodalis* ou *Fukatsuia*), laissent supposer que *Serratia* a été acquise à de multiples reprises de façons indépendantes par les *Cinara*.

Table 1 : Récapitulatif des caractéristiques génomiques et histologiques des *Serratia symbiotica* « co-obligatoires » associées à la famille des Lachninae. Dans les autres genres de Lachninae décrits : *Stomaphis*, *Lachnus*, *Maculolachnus*, , *Nippolachnus*, , *Essigella*, *Eulachnus* d'autres endosymbiotes obligatoires ont été généralement identifiés par séquençage du gène 16S rDNA chez quelques espèces de chaque genre et parois des études FISH (Manzano-Marín et al. 2017). Dans trois genres, *Longistigma*, *Protrama*, *Pyrolachnus*, aucune donnée n'existe sur l'identité du co-symbiote comblant une perte ancestrale de la capacité de biosynthèse de vitamines B et de Riboflavine chez *Buchnera*. cellules coocoides,  cellules bacilliformes ○

Genre(nb d'espèces décrites)	Espèce	Taille Génome de <i>Serratia</i>	Références	Localisation de <i>Serratia</i> et forme des cellules	Références
<i>Cinara</i> (> 240 sp)	<i>cedri</i>	1.78 Mb	(Lamelas et al. 2011)	bactériocytes, ○	(Gomez-Valero et al. 2004), (Manzano-Marín et al. 2017)
	<i>cupressi</i>	?	(Manzano-Marín et al. 2017)	Périphérie, 	(Manzano-Marín et al. 2017)
	<i>tujafilina</i>	2,5 MB	(Manzano-Marín & Latorre 2014)	Périphérie, 	(Manzano-Marín et al. 2017)
	<i>fornacula</i>	1.17 MB	(Meseguer et al. 2017)	?	
	<i>strobi</i>	2,24 MB	(Manzano-Marín et al. 2018).	?	
<i>Tuberolachnus</i> (2)	<i>salignus</i>	0,65 MB	(Manzano-Marín et al. 2016)	bactériocytes, ○	(Manzano-Marín et al. 2017)
<i>Trama</i> (15)	<i>caudata</i>	?		bactériocytes, ○	(Manzano-Marín et al. 2017)
<i>Pterochloides</i> (1)	<i>persicae</i>	?		bactériocytes, ○	(Manzano-Marín et al. 2017)

Les études phylogénétiques de *Serratia symbiotica* ne concernent que quelques espèces. L'étude de Manzano-Marín et al (2018), sur la souche de *Serratia symbiotica* associée à *Cinara strobi*, montre une phylogénie basée sur des génomes complets de deux *Serratia* facultatives et quatre *Serratia* co-obligatoires. Elle révèle un regroupement des espèces à génomes réduits mais avec beaucoup d'incertitudes et de très longues branches. Les phylogénies basées sur un seul gène (gène 16S rRNA), des *Serratia symbiotica* associées à une plus large diversité de pucerons, sont-elles aussi très incertaines (Figure 1). Les réductions, remaniements, évolution rapide des génomes et taux de GC faibles favorisent les convergences moléculaires et induisent des biais tels que l'artefact d'attraction des longues branches (Husník et al., 2011; Philippe and Roure, 2011), ce qui rend les relations de parenté des bactéries endosymbiotiques difficiles à reconstruire et donc les différents scénarii d'acquisitions difficiles à décrypter.

Ici nous avons cherché à reconstruire l'histoire évolutive de l'acquisition de *Serratia symbiotica* au sein des *Cinara*, afin de déterminer le nombre d'acquisitions et/ou de pertes indépendantes en tant que symbiotes obligatoires et ainsi inférer le nombre de transitions de bactéries secondaires facultatives vers des bactéries obligatoires. Pour cela nous avons reconstruit les phylogénies de *Cinara*, *Buchnera* et *Serratia* à l'aide de méthodes probabilistes et de modèles d'évolution moléculaire. A partir de ces inférences phylogénétiques nous avons effectué des analyses de réconciliation pour analyser la plausibilité de différents scénarii de co-spéciation entre l'hôte et ses symbiotes. Parallèlement à cette analyse, la localisation des bactéries dans le puceron a été décrite chez plusieurs espèces afin de vérifier l'existence d'une corrélation avec l'histoire évolutive de l'association. Cela permettra notamment de vérifier l'existence d'un gradient de « domestication » du symbiote variant avec l'âge d'acquisition, mais aussi une convergence de la structure de la symbiose chez des bactéries acquises indépendamment. On peut en effet supposer que les symbiotes récemment acquis ne sont pas toujours dans des bactériocytes propres comme chez *Cinara tujaefilina* contrairement à ce que l'on observe pour des symbioses plus anciennes. Finalement, nous étudierons la convergence de la structure des génome endosymbiotique par une description de ceux-ci en terme de taille de génome, taux de GC et nombre de CDS. Pour effectuer ces analyses, nous avons utilisé un jeu de données génomiques représentatif de la diversité des *Cinara*. Pour chacune des espèces de pucerons utilisées étaient associés le génome mitochondrial complet de l'hôte, le génome complet de *Buchnera aphidicola* et, lorsqu'elle était présente, le génome de *Serratia symbiotica*.

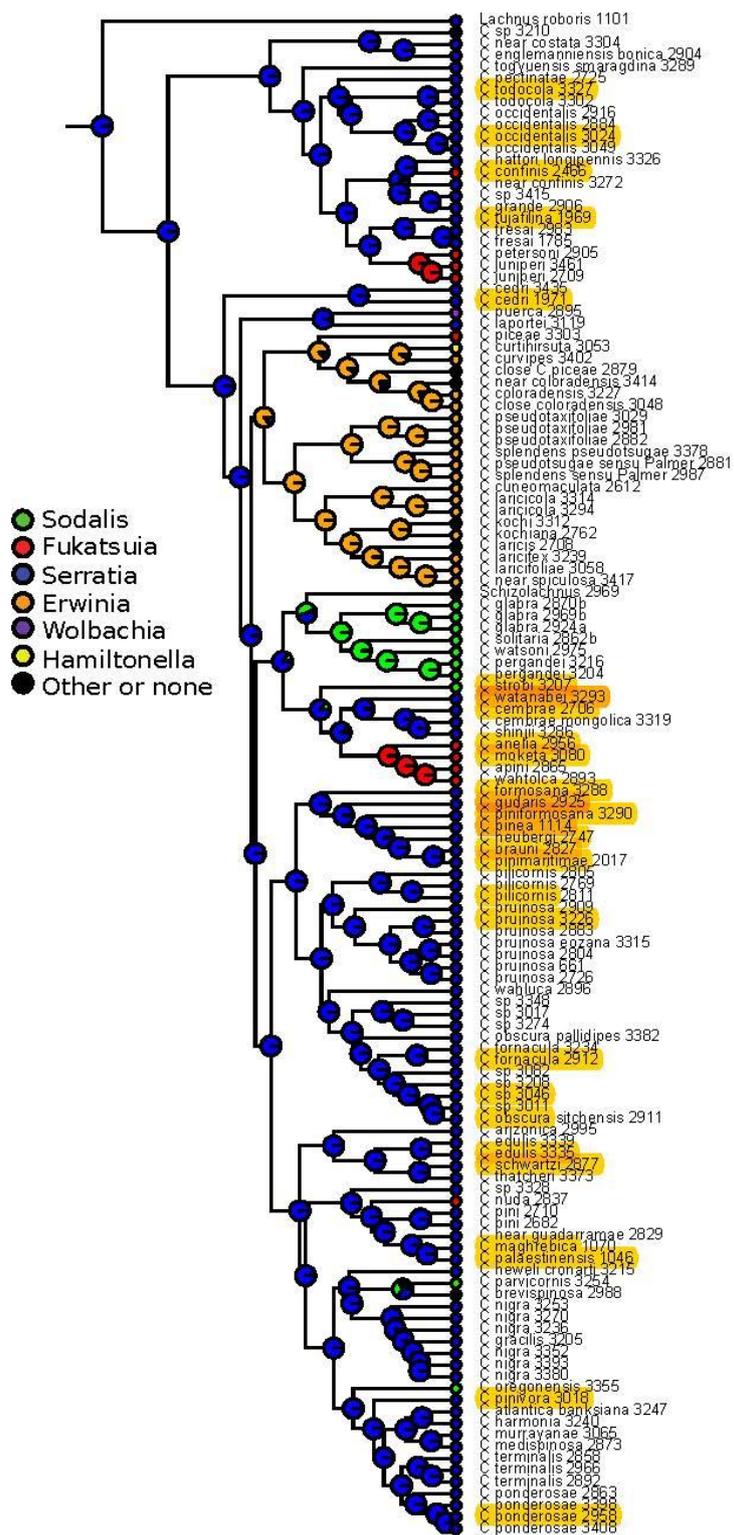


Figure 1 : Figure extraite et modifiée de Meseguer et al. 2017 : les espèces surlignées en orange sont les espèces de *Cinara* pour lesquelles nous avons pu obtenir un génome pour *Buchnera*, *Serratia* et un génome mitochondrial de pucerons. Les espèces ont été choisies afin d'être représentatives de la diversité des *Cinara* ayant *Serratia symbiotica* pour co-symbiote obligatoire et pour lesquelles nous disposons de suffisamment de matériel en alcool pour construire les bibliothèques de séquençage.

II. Matériel et méthode

Collecte des pucerons, extraction de l'ADN, séquençage et assemblage

Préalablement à ma thèse, des échantillonnages de colonies de pucerons du genre *Cinara* (Curtis 1835 ;Lachninae : Eulachnini) ont été effectués comme rapportés dans (Meseguer et al., 2015) et (Meseguer et al., 2017) et stockés dans de l'éthanol à 70% à 6 °C dans la collection CBGP. Ces colonies ont été utilisées pour préparer des échantillons d'ADN enrichis en bactéries comme décrit dans (Jousselin et al., 2016) pour le séquençage du génome entier par le projet France Génomique MicroPhan. Les espèces utilisées dans cette étude sont soulignées dans la figure 1 extraite de l'article de Meseguer et al. 2017. Pour cela, 15 à 20 pucerons d'une même colonie ont été regroupés. L'ADN extrait a été utilisé pour préparer deux librairies. En résumé, 5 ng d'ADN génomique ont été fragmentés, en utilisant l'appareil E220 de Covaris (Covaris, USA). Les fragments ont été réparés en bout de chaîne et des adaptateurs de barcode PCR libres NEXTflex (Bioo Scientific, USA) ont été ajoutés en utilisant le kit de préparation de librairie NEBNext® Ultra II pour Illumina (New England Biolabs, USA). Les produits de ligation ont ensuite été purifiés par Ampure XP (Beckman Coulter, États-Unis) et les fragments d'ADN (>200 pb) ont été amplifiés par PCR (2 réactions PCR, 12 cycles), en utilisant des amorces spécifiques aux adaptateurs d'Illumina et le kit NEBNext ® Ultra II Q5 (NEB). Après l'analyse du profil de la librairie par le bioanalyseur Agilent 2100 (Agilent Technologies, États-Unis) et la quantification par qPCR, à l'aide du kit de quantification de bibliothèque KAPA pour les librairies Illumina (Kapa Biosystems, États-Unis), les librairies ont été séquencées, en read appariée de 251 pb sur un séquenceur Illumina HiSeq2500. L'ensemble de ces librairies et du séquençage a été réalisé par une équipe du génoscope.

Après séquençage, pour chaque librairie, les reads obtenus ont été filtrés et assemblés suivant le protocole de Manzano-Marin et al. 2020. Ces étapes ont été réalisées avant le début de ma thèse. En résumé, seuls les reads Illumina d'une qualité minimum de 20 ont été conservés en utilisant FASTX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/). Les reads de moins de 75 pb ont été éliminés. PRINSEQ v0.20.4 (Schmieder and Edwards, 2011) a été utilisé pour supprimer les reads contenant des nucléotides non définis ainsi que ceux qui n'avaient pas de paires après le processus de filtrage. Les reads résultants ont été assemblés, en utilisant SPAdes v3.10.1 (Bankevich et al., 2012), avec les options --only-assembler et des

tailles de k-mer de 33, 55, 77, 99 et 127. Parmi les contigs résultantes, celles qui étaient plus courtes que 200 pb ont été éliminées. Les contigs restantes ont été classées, en utilisant les résultats d'une recherche BLASTX (Altschul, 1997) (meilleur résultat par contig) par rapport à une base de données constituée du protéome du puceron du pois et d'une sélection de bactéries symbiotiques du puceron trouvées précédemment dans les protéomes des espèces de *Cinara* et des espèces apparentées vivant librement (voir (Manzano-Marín et al., 2020; Meseguer et al., 2017) pour plus de détails sur les protéomes de référence). Les contigs attribuées ont été examinées manuellement à l'aide du serveur web BLASTX (recherche dans la base de données nr) afin d'assurer une attribution correcte. Ce processus d'assignation a confirmé la présence de *Serratia symbiotica*, telle que déduite du séquençage des fragments du gène de l'ARNr 16S de plusieurs spécimens (Jousselin et al., 2016; Meseguer et al., 2017) et du séquençage des endosymbiotes de *C. fornacula* rapporté dans Meseguer et al. (2017). Nous avons également toujours récupéré une contig d'environ 16000 pb représentant les génomes mitochondriaux des hôtes des pucerons. Les contigs obtenues ont ensuite été utilisées comme référence pour mapper les reads et l'assemblage de génomes individuels a été fait par A. Manzano-Marín à l'aide de SPAdes, comme décrit ci-dessus, avec correction des erreurs de lecture.

Phylogénie

Parmi les génomes assemblés, 43 espèces représentatives de la diversité phylogénétique des *Cinara* et de leurs co-symbiotes obligatoires ont été sélectionnées en gardant uniquement celles pour lesquelles nous disposions des génomes du symbiote primaire *Buchnera aphidicola* et du symbiote co-obligatoire associé (Table S1, Figure 1). L'ensemble des paires de génomes du duo *Buchnera*+*Serratia* pour lesquelles l'assemblage était satisfaisant a été utilisé. Les génomes de *Buchnera* des lignées associées aux *Cinara* dépourvues de *Serratia* mais abritant une autre bactérie co-obligatoire ont été inclus dans nos analyses de réconciliation (Table S1) afin d'inférer correctement les événements de pertes de *Serratia* au cours de la spéciation des *Cinara*. Des espèces externes aux *Cinara* ont également été utilisées pour enraciner les phylogénies. Ainsi nous avons utilisé les génomes complets de 43 *Buchnera* associées au *Cinara*, et de quatre *Buchnera* associées à d'autres genres comme groupes externes. Parmi les groupes externes nous avons sélectionné deux Lachninae *Lachnus roboris* et *Eulachnus mediterraneus* et trois espèces de groupes plus éloignés *Acyrtosiphon pisum*, *Aphis fabae* et *Aphis gossypii* totalisant 48 espèces. Les génomes de *Buchnera* hébergées par *A. pisum*, *Aphis fabae* et *Aphis gossypii*, ainsi que ceux de 13 espèces de *Cinara* étaient préalablement publiés,

les 30 autres sont nouvellement séquencés. Pour les *Serratia* nous avons utilisé les génomes complets appartenant à 32 espèces, dont 28 associées aux *Cinara* (24 nouvellement séquencées et 4 génomes préalablement publiés). Toutes les *Serratia* associées à *Cinara* séquencées pour cette étude sont obligatoires, exceptée celles de *Cinara curtihirsuta* qui est facultative : elle n'est pas fixée dans l'espèce, et il a été démontré que c'était la bactérie *Erwinia* qui complémentait *Buchnera* chez ce puceron et non *Serratia* (Meseguer et al., 2017). Trois espèces de pucerons d'un autre genre que *Cinara* ont été utilisées comme groupes externes (*Acyrtosiphon pisum*, *Aphis fabae* et *Aphis gossypii*) et dans lesquelles *Serratia symbiotica* est facultative. Finalement, la bactérie au mode de vie libre *Serratia marcescens* a été choisie comme groupe plus externe.

Phylogénie mitochondriale de *Cinara*

Afin de reconstruire la phylogénie des pucerons hôtes, nous avons utilisé les treize gènes mitochondriaux (atp6, atp8, cob, cox1, cox2, cox3, nad1, nad2, nad3, nad4, nad4L, nad5, nad6) pour chacune des espèces de *Cinara* séquencée ou récupérée sur NCBI. Pour cela, les génomes mitochondriaux ont été annotés à l'aide de MitoS Web (Bernt et al., 2013) puis vérifiés manuellement sur Geneious v11.1.5 (<https://www.geneious.com>). Nous avons sélectionné cinq espèces utilisées comme outgroup (*Adelges tsugae* (*Adelgidae*), *Acyrtosiphon pisum*, *Aphis gossypii*, *Aphis fabae*,) rassemblées à partir du NCBI et *Lachnus roboris* un outgroup de la famille des Lachninae nouvellement séquencé dans cette étude (Table S1). Pour chaque gène, les alignements de séquences ont été réalisés à l'aide de Macse v2.03. (Ranwez et al., 2018) et vérifiés visuellement à l'aide de SeaView v4 (Gouy et al., 2010). Nous avons ensuite supprimé les blocs divergents et aligné de manière ambiguë en utilisant Gblocks v0.91b (Talavera and Castresana, 2007) avec les paramètres par défaut et en veillant à conserver le cadre de lecture. Les longueurs nucléotidiques des alignements par gène ont été relevées afin de créer un fichier de partitions pour l'analyse phylogénétique et ont ensuite été concaténées. Pour tenir compte des variations possibles dans la vitesse d'évolution entre les différents gènes ainsi que les différentes positions de codon, un partitionnement de l'alignement a été effectué à l'aide du logiciel PartitionFinder 2.1.1 (Lanfear et al., 2016) avec des blocs de données définies pour chaque position de codon (première, deuxième et troisième) dans chaque gène. Une première phylogénie a par la suite été reconstruite par maximum de vraisemblance avec IQtree multicore 1.6.2 (Nguyen et al., 2015) avec la meilleure partition trouvée par PartitionFinder et 1000 itérations bootstraps ultra-rapides pour estimer le soutien statistique des clades.

Par la suite, des analyses phylogénomiques par inférences bayésiennes ont été faites avec: MrBayes v3.2.7 (Ronquist et al., 2012) avec la meilleur partition sélectionnée par PartitionFinder disponible dans MrBayes (selon la métrique du critère d'information bayésien (BIC)). Deux analyses indépendantes avec quatre chaînes chacune pour 3 000 000 de générations ont été réalisées et la convergence de celle-ci a été vérifié.

Phylogénies des symbiotes : *Buchnera* et *Serratia*

Les séquences codantes (CDS, CoDing Sequences) ont été recherchées dans les génomes endosymbiotiques à l'aide de Prodigal 2.6.3 (PROkaryotic Dynamic programming Gene-Finding Algorithm, (Hyatt et al., 2010)). Puis une recherche de groupes d'orthologues a été réalisée à partir des résultats en acides aminés (AA) de Prodigal avec OrthoFinder-2.5.4 (Emms and Kelly, 2015). Enfin, seuls les gènes les plus conservés, présents chez toutes les espèces séquencées en un seul exemplaire ont été utilisés pour constituer les jeux de données.

Nous avons ainsi obtenu 480 gènes partagés par toutes les espèces de *Serratia* et 304 gènes partagés par toutes les espèces de *Buchnera*. Pour chaque ensemble de données, les séquences d'acides nucléiques ont été alignées à l'aide de MAFFT v7.450 (avec un maximum de 1 000 itération) (Kato and Standley, 2013) et concaténées. Les blocs divergents et alignés de manière ambiguë ont été supprimés à l'aide de Gblocks (Talavera and Castresana, 2007) (paramètres par défaut et option `-b5 h`, pour éliminer une position que si au moins la moitié des séquences comportent un gap). Compte tenu de la vitesse d'évolution élevée des endosymbiotes, seul les alignements en AA ont été utilisés pour la suite des analyses. En effet, les taux de substitution trop élevés des positions nucléotidiques risquent d'augmenter les homoplasies, et donc de fausser les inférences phylogénétiques en rendant difficile la détection de substitutions multiples. Le choix des modèles de substitution sur les ensembles de données pour chaque symbiote a été conduit à l'aide de ModelFinder (Kalyaanamoorthy et al., 2017) (option `-MFP+MERGE` pour tester les modèles d'évolution protéiques classiques et `-madd` pour l'explorer l'ensemble des modèles additionnels implémentés dans IQ-TREE 1.6.8) en utilisant le critère d'information bayésien (BIC). Les inférences phylogénétiques par ML ont ensuite été réalisées à l'aide de IQ-TREE 1.6.8 (Nguyen et al., 2015) avec 1000 itérations bootstraps ultra-rapides pour estimer le soutien statistique des clades.

Pour chacun des symbiotes, les inférences bayésiennes ont également été faites avec le modèle CAT, implémenté dans Phylobayes 4.1c (Lartillot et al., 2009) avec une distribution des vitesses d'évolution entre sites suivant une loi Γ discrétisée en 4 catégories (Γ_4). Afin de

tenir compte des taux d'évolution rapides des endosymbiotes et pour éviter de masquer des changements passés dans les séquences, nous avons effectué un recodage des acides aminés en 6 catégories selon la matrice de Dayhoff6. Celui-ci a permis de les regrouper selon leurs caractéristiques physico-chimiques. En effet, lorsque les taux de substitution sont élevés, les séquences en acides aminés, peuvent atteindre la saturation, c'est à dire que les substitutions passées risquent de ne pas pouvoir être recouvrées. Quatre chaînes ont été lancées pour 20 000 cycles (3200 au moment du rendu du manuscrit) pour chacun des symbiotes. Finalement, « la phase d'allumage » a été retirée (20% des cycles) et la convergence des runs a été regardée à l'aide de la fonction tracecomp de Phylobayes afin de vérifier que la différence d'estimations des paramètres entre 2 chaînes est bien inférieure à 0.3, et que la taille efficace (ESS) est supérieure à 50.

Réconciliation

Afin de tester les différents scénarii de cospéciation entre les *Cinara*, *Buchnera* et *Serratia*, des analyses de réconciliations d'arbres phylogénétiques par maximum de parcimonie ont été réalisées (Encadré). Pour cela le programme Mowgli (Doyon et al., 2011) a été utilisé. Pour chaque analyse, la phylogénie de *Buchnera* réalisée par inférence bayésienne avec Phylobayes a été utilisée comme référence. Les tests des scénarii de cospéciation ont alors été effectués avec les deux phylogénies de *Serratia* obtenues par l'analyse en Maximum de Vraisemblance (IQtree) et par Inférences Bayésiennes (Phylobayes). Afin d'orienter les réconciliations, l'option -b de Mowgli a été utilisée permettant de contraindre les racines des symbiotes à correspondre entre elles. Le groupe servant à raciner les phylogénies représente un groupe externe hypothétique (*Serratia marcescens* pour les *Serratia* associées de façon artificielle à *Eulachnus mediterraneus* utilisé pour enraciner la phylogénie de *Buchnera*) et permet de proposer des sources de transferts de *Serratia* externes aux pucerons étudiés. En effet, l'origine d'un co-symbiote étant généralement inconnue, une *Serratia* pourrait provenir de l'environnement (autre insecte, sol, sève, etc...) et non d'une co-spéciation avec les *Cinara*. Pour cette raison des groupes externes de *Serratia* qui ne sont pas issus des *Cinara* sont indispensables. Pour permettre aux réconciliations d'associer la racine de l'arbre des *Serratia* à un ancêtre d'une lignée extérieure qui ne provient pas des *Cinara* présents dans notre jeu de données. Les réconciliations ont ensuite été visualisées dans l'interface graphique SylvX (Chevenet et al., 2016).

Encadré 1 : Réconciliations par Maximum de Parcimonie (MPR) :

Les méthodes de co-phylogénie visent à éclairer l'histoire co-évolutive entre un hôte et un associé (parasite ou mutualiste) sur la base de phylogénies. Une des premières méthodes, l'Analyse de Parcimonie de Brooks, (BPA : Brooks, 1981), consistait à considérer les symbiotes comme des états de caractères de l'hôte et à minimiser le nombre de changements d'états (transitions entre présence et absence du symbiote). Page, dans les années 1990, proposa l'analyse de réconciliation, qui considère cette fois les associés comme des lignées évolutives. Cette méthode tente de faire correspondre la phylogénie des associés à celle des hôtes pour reconstruire le scénario le plus parcimonieux, c'est à dire le moins coûteux en événements. Ces événements étant de 4 grands types, chacun étant associé à un coût (voir tableau) :

	Événement	Coût
●	Cospéciation	0
✗	Extinction	1
◇	Spéciation intra hôte	1
←	Changement d'hôte	1

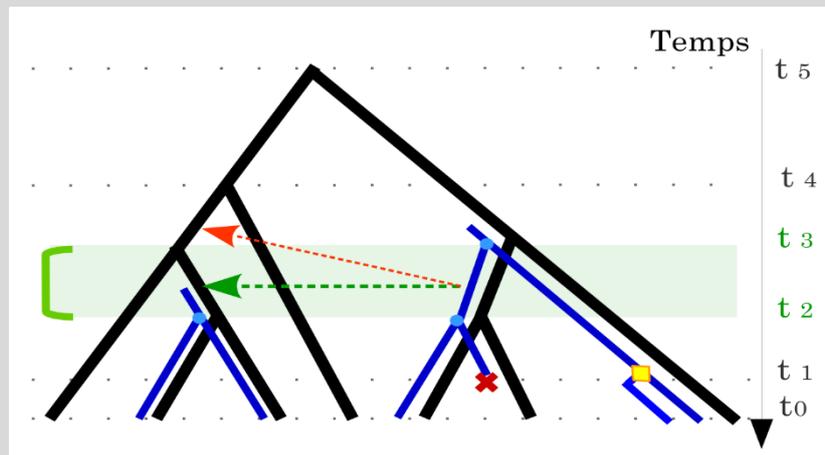
1) cospéciation : spéciation quasi simultanée des hôtes et de leurs associés, 2) extinction : la lignée du symbiote s'éteint, 3) spéciation ou duplication intra-hôte: les associés divergent en deux ou plusieurs lignées dans leurs hôtes qui eux ne subissent pas de spéciation et 4) changement d'hôtes : un symbiote va coloniser un hôte plus ou moins apparenté à son hôte d'origine.

Une des difficultés rencontrées par ces méthodes est la congruence temporelle lors des changements d'hôtes. En effet, un échange de symbiotes entre deux lignées ne peut exister que si celles-ci sont contemporaines : par exemple, un transfert ne peut s'effectuer vers une branche disparue car cela reviendrait à revenir vers le passé (flèche orange ci-dessous).

Pour éviter ce problème, l'utilisation d'algorithmes de programmation dynamique peut être une solution : le temps dans lequel évoluent les partenaires est découpé en courts intervalles, et la réconciliation est réalisée par pas de temps, du présent (t_0) en remontant vers le passé.

De cette manière, des changements d'hôtes

entraînant des incongruences temporelles ne peuvent être proposés (accolade et flèche vertes). De plus, cette décomposition facilite la résolution de ce problème et limite ainsi le temps de calcul : les coûts sont optimisés à chaque pas de temps au lieu de l'être globalement. Cette technique permet normalement de trouver le scénario de réconciliation le plus parcimonieux.



Adapté de Annonier 2018

Localisation et morphologie des co-symbiotes obligatoires

La localisation des bactéries dans le puceron (bactériocytes spécifiques/partagés ou extracellulaires) ainsi que la forme des bactéries ont été observés chez plusieurs espèces à l'aide de FISH (fluorescence in-situ hybridation). Ces observations ont été réalisées, chez *C. cedri* (espèce témoin pour laquelle on connaît déjà les localisations), *C. palaestinensis* et *C. pinea* (deux espèces du clade A), *C. confinis*, *C. pectinatae*, (clade C sur la phylogénie pucerons). En plus de celles-ci, des FISH ont été réalisées dans des études précédentes sur d'autres espèces de *Cinara* telle que *C. tujafilina* (cf chapitre 3 (Manzano-Marín et al., 2017)). Des individus de chacune des espèces nommées ci-dessus ont d'abord été fixés lors de la collecte dans un fixateur de Carnoy modifié (6 chloroforme : 3 éthanol absolu : 1 acide acétique glacial) et laissés pendant la nuit, en suivant le protocole de (Koga et al., 2009). Les individus ont ensuite été transférés dans une solution de 6% de H₂O₂ diluée dans de l'éthanol absolu et ont été laissés dans cette solution pendant 2 semaines (en changeant la solution tous les 3 jours). Les individus ont ensuite été lavés deux fois avec de l'éthanol absolu et disséqués dans de l'éthanol absolu pour extraire les embryons. L'hybridation a été réalisée pendant une nuit à 30 °C dans un tampon d'hybridation standard (20mM Tris-HCl [pH 8,0], 0,9M NaCl, 0,01% SDS et 30% formamide), puis lavée (20mM Tris-HCl [pH 8,0], 5 mM EDTA, 0,1M NaCl et 0,01% SDS) avant la préparation des lames (Manzano-Marín et al., 2020). Les embryons d'un maximum de 10 individus ont été observés sous un microscope confocal. Nous avons utilisé plusieurs sondes spécifiques compétitives pour *Buchnera* et *Serratia* selon les espèces (Table 2). Les images au format TIFF ont été exportées à l'aide de la suite d'applications X de Leica. Les images exportées ont été importées dans Inkscape v0.92.4 pour construire les figures.

Table 2 : Liste des amorces utilisées pour la microscopie à hybridation in situ par fluorescence.

Espèce hôte	Bactérie	Sonde	Séquence	Origine
<i>Cinara tujaefilina</i>	<i>Buchnera</i>	BLach-2x6-FAM	5'-2x6-FAM- CCCGTTYGCCGCTCGCCGTCA-3'	(Manzano-Marín et al., 2017)
<i>Cinara tujaefilina</i>	<i>Serratia</i>	Ssym-Cy3	5'-Cy3-CTCGTCACCCAGAGAGCA-3'	(Manzano-Marín et al., 2017)
<i>Cinara palaestinis</i>	<i>Buchnera</i>	BLach-FITC	5'-FITC- CCCGTTCGCCGCTCGCCGGCA-FITC- 3'	(Manzano-Marín et al., 2017)
<i>Cinara palaestinis</i>	<i>Serratia</i>	SeJR5Cinara-Cy3	5'-Cy3-CCGTYCGCCGCTCGTCACC- Cy3-3'	Cette etude
<i>Cinara pinea</i>	<i>Buchnera</i>	BLach-FITC	5'-FITC- CCCGTTCGCCGCTCGCCGGCA-FITC- 3'	(Manzano-Marín et al., 2017)
<i>Cinara pinea</i>	<i>Serratia</i>	SeJR5Cinara-Cy3	5'-Cy3-CCGTYCGCCGCTCGTCACC- Cy3-3'	Cette etude
<i>Cinara cedri</i>	<i>Buchnera</i>	BccFITC	5'-FITC- CCCGTTCGCCGCTCGCCGGCA-FITC- 3'	(Gosalbes et al., 2008)
<i>Cinara cedri</i>	<i>Serratia</i>	SCcCy3	5'-Cy3-CCGCCGCTCGTCACCCAAA-3'	(Manzano-Marín et al., 2017)
<i>Cinara confini</i>	<i>Buchnera</i>	BuconfITC	5'-FITC- CCGTTTGCCGCTCGCCGTCAGAA- FITC-3'	Cette etude
<i>Cinara confini</i>	<i>Serratia</i>	SeJR5Cinara-Cy3	5'-Cy3-CCGTYCGCCGCTCGTCACC- Cy3-3'	Cette etude
<i>Cinara pectinatae</i>	<i>Buchnera</i>	BLach-FITC	5'-FITC- CCCGTTCGCCGCTCGCCGGCA-FITC- 3'	(Manzano-Marín et al., 2017)
<i>Cinara pectinatae</i>	<i>Serratia</i>	SeJR5Cinara-Cy3	5'-Cy3-CCGTYCGCCGCTCGTCACC- Cy3-3'	Cette etude

III. Résultats

Caractéristiques génomiques

Nous avons généralement obtenu des génomes complets pour *Buchnera* avec une taille moyenne de 0.45 Mb contre 1.68 Mb pour *Serratia* (Table S1). Les résultats montrent que les génomes de *Buchnera* sont très stables, en terme de nombre de taille et de pourcentage en GC, alors que ceux de *Serratia* montrent des caractéristiques génomiques variables (Figures 2 et 3, Table S1).

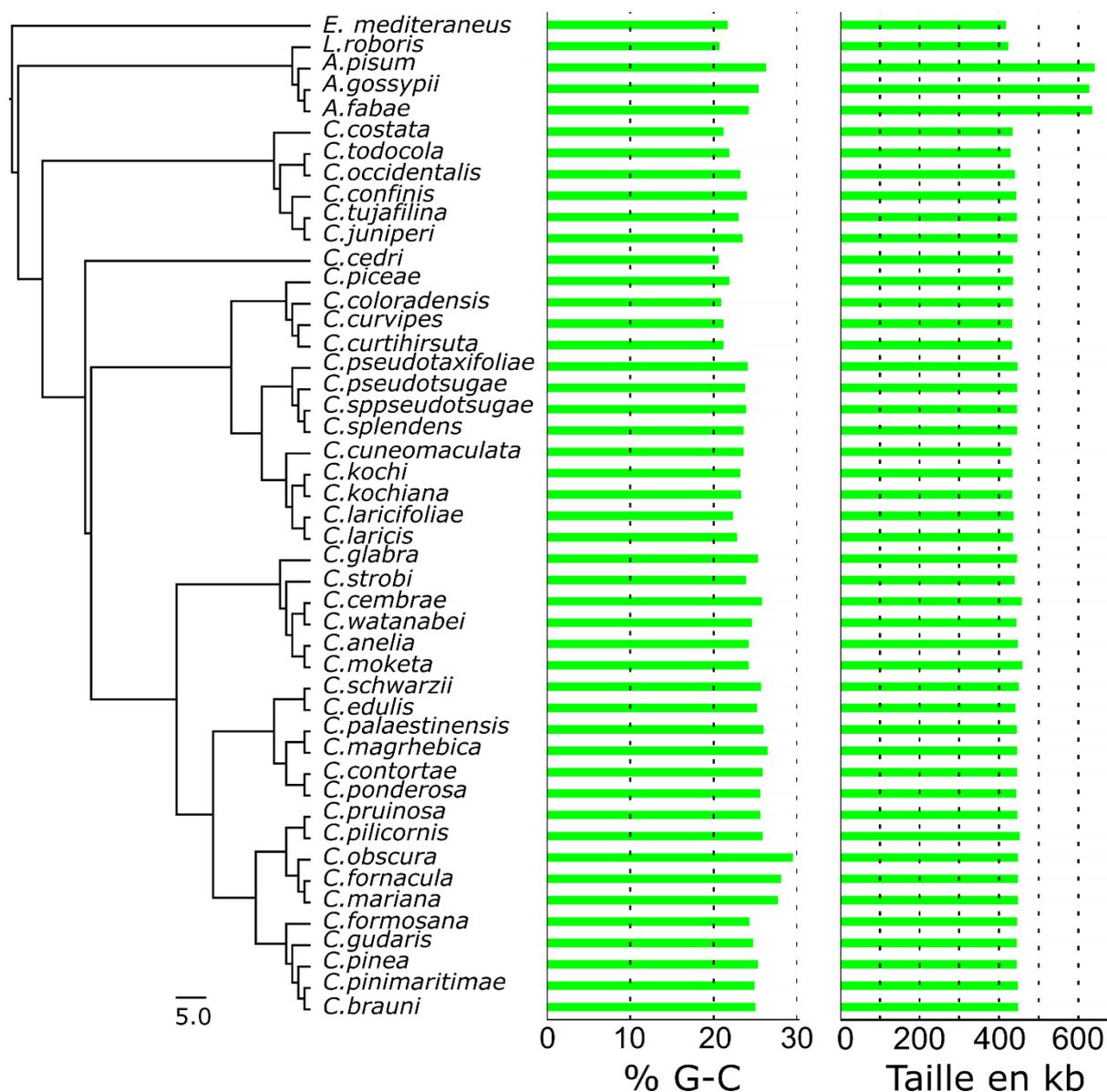


Figure 2 : Caractéristiques génomiques de *Buchnera aphidicola* le long de leur phylogénie avec à gauche la phylogénie reconstruite en Maximum de Vraisemblance (IQtree1.6.2) et représentée en cladogramme (ultramétrique). Les noms des espèces-hôtes (pucerons) sont indiqués aux feuilles. Les diagrammes en barres représentent les taux de Guanine+Cytosine génomiques et la taille des génomes de *Buchnera* en kilobases.

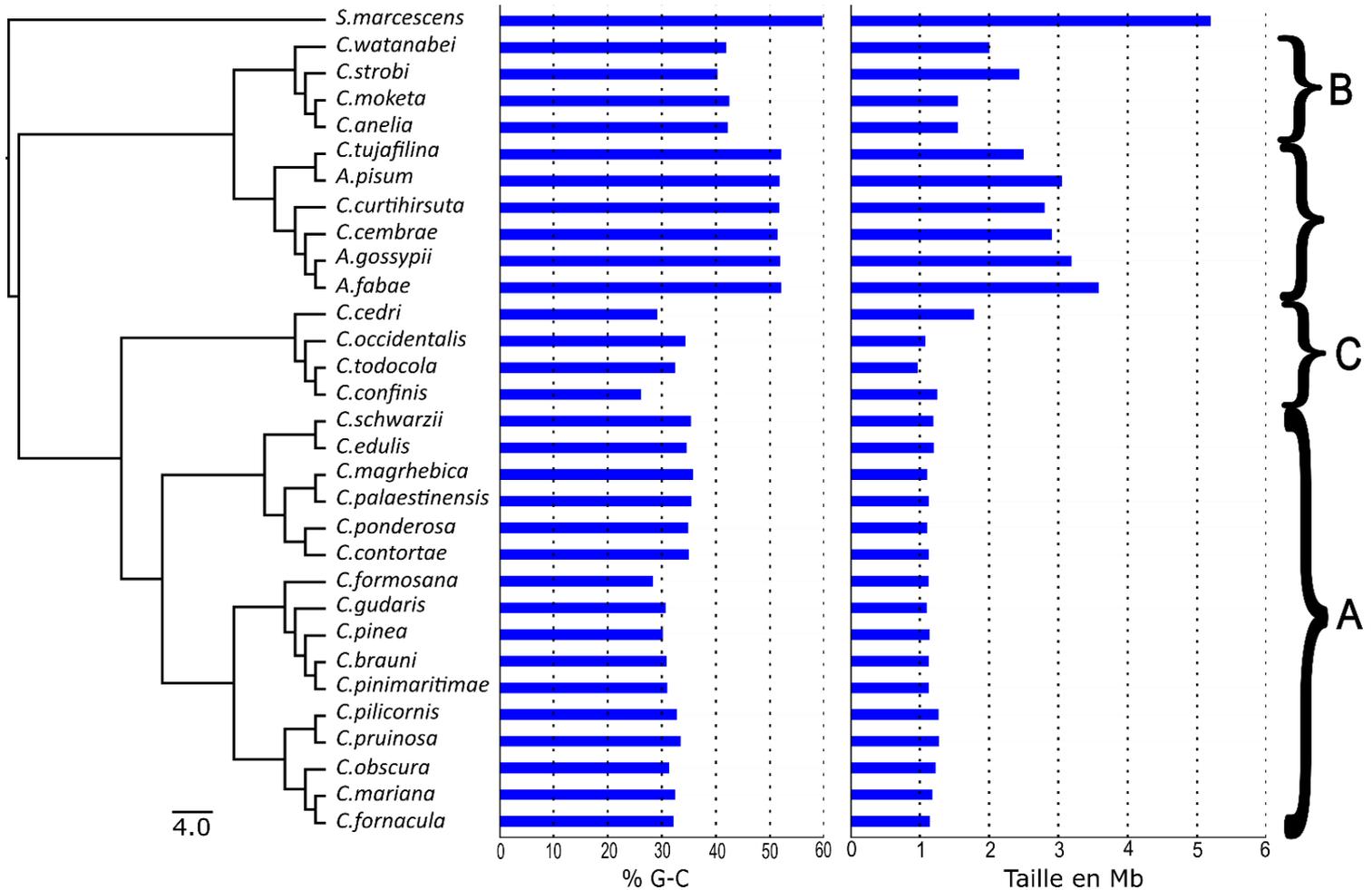


Figure 3 : Caractéristiques génomiques des *Serratia symbiotica* le long de leur phylogénie avec à gauche la phylogénie reconstruite en Maximum de Vraisemblance (IQtree1.6.2) et représentée en cladogramme (ultramétrique). Les noms des espèces-hôtes (pucerons) sont indiqués aux feuilles). Les diagrammes en barres représentent les taux de Guanine+Cytosine génomiques et la taille des génomes de *Serratia* en Megabases.

Phylogénies

Les analyses phylogéniques de *Buchnera* obtenues sont congruentes entre les différentes méthodes avec celle obtenues pour les *Cinara*. Ces résultats sont conformes avec l'hypothèse déjà bien étayée par d'autres études d'une cospéciation entre *Buchnera* et les pucerons. On retrouve ainsi les différents clades décrit précédemment chez les *Cinara* (Burke et al., 2009; Lamelas et al., 2008; Meseguer et al., 2017) (Figure 2, S1, S3).

En ce qui concerne *Serratia*, nous obtenons quatre groupes bien distincts (Figure 3, S2, S4), le clade A dont les relations phylogénétiques sont parfaitement congruentes avec celles de leur puceron hôte (voir chapitre 2). Un groupe composé de *Serratia* obligatoires anciennement acquises (clade C) pour lesquelles la présence de longues branches pourrait indiquer un artefact. Enfin les lignées restantes forment le clade B dont les positions des nœuds profonds changent selon la méthode utilisée (Figure.S2, S4). Un groupe frère du clade B, regroupe les souches de

Serratia symbiotica connues comme bactéries facultatives chez les pucerons en dehors des *Cinara* ainsi que les souches associées à *Cinara tujafilina* et *Cinara cembrae*.

Réconciliation

La visualisation des réconciliations dans l'interface graphique SylvX (Chevenet et al., 2016) nous a permis de décrire différents scénarii possibles de la cospéciation entre les *Cinara* et leurs endosymbiotes obligatoires. Etant donnée la phylogénie peu stable de *Serratia* dans les nœuds profonds, il existe plusieurs scénarii. Dans tous les scénarii il y a cospéciation parfaite entre les *Cinara* du clade A et leurs symbiotes, certaines souches de *Serratia* ont été acquises très récemment (celle de *C. tujafilina* et de *C. cembrae*) et finalement le clade B semble être issu d'une acquisition récente avec de la cospéciation (Figure 4, 5). Pour les autres associations (avec *C. cedri* ainsi que les espèces du clade C : *C. occidentalis*, *C. confinis* et *C. todocola*) étant donnée les incertitudes sur leurs positions phylogénétiques, les variations dans leurs caractéristiques génomiques il est difficile de conclure sur le nombre d'acquisitions indépendantes.

Les scénarii les plus parcimonieux ; inférés par les réconciliations sont dans tous les cas à modérer, certains transferts sont en effet biologiquement impossibles. Par exemple, pour ce qui est de *C. occidentalis*, il semble impossible qu'un transfert ait eu lieu depuis *Cinara cedri* (Figure 4), de même pour un transfert entre *C. strobi* et *C. anelia* et *C. moketa* (Figure 4, 5) puisque la transmission d'un symbiote obligatoire d'une espèce à l'autre parait très improbable. Le scénario figure 5 (inféré à partir de la phylogénie de *Serratia* obtenue avec PhyloBayes), semble biologiquement plus probable (car il y a peu de transferts horizontaux de symbiotes obligatoires) et il suggère ainsi que les associations avec *C. occidentalis*, *C. confinis* et *C. todocola*) résultent d'une ancienne acquisition ancestrale à la diversification du clade C (Figure 4).

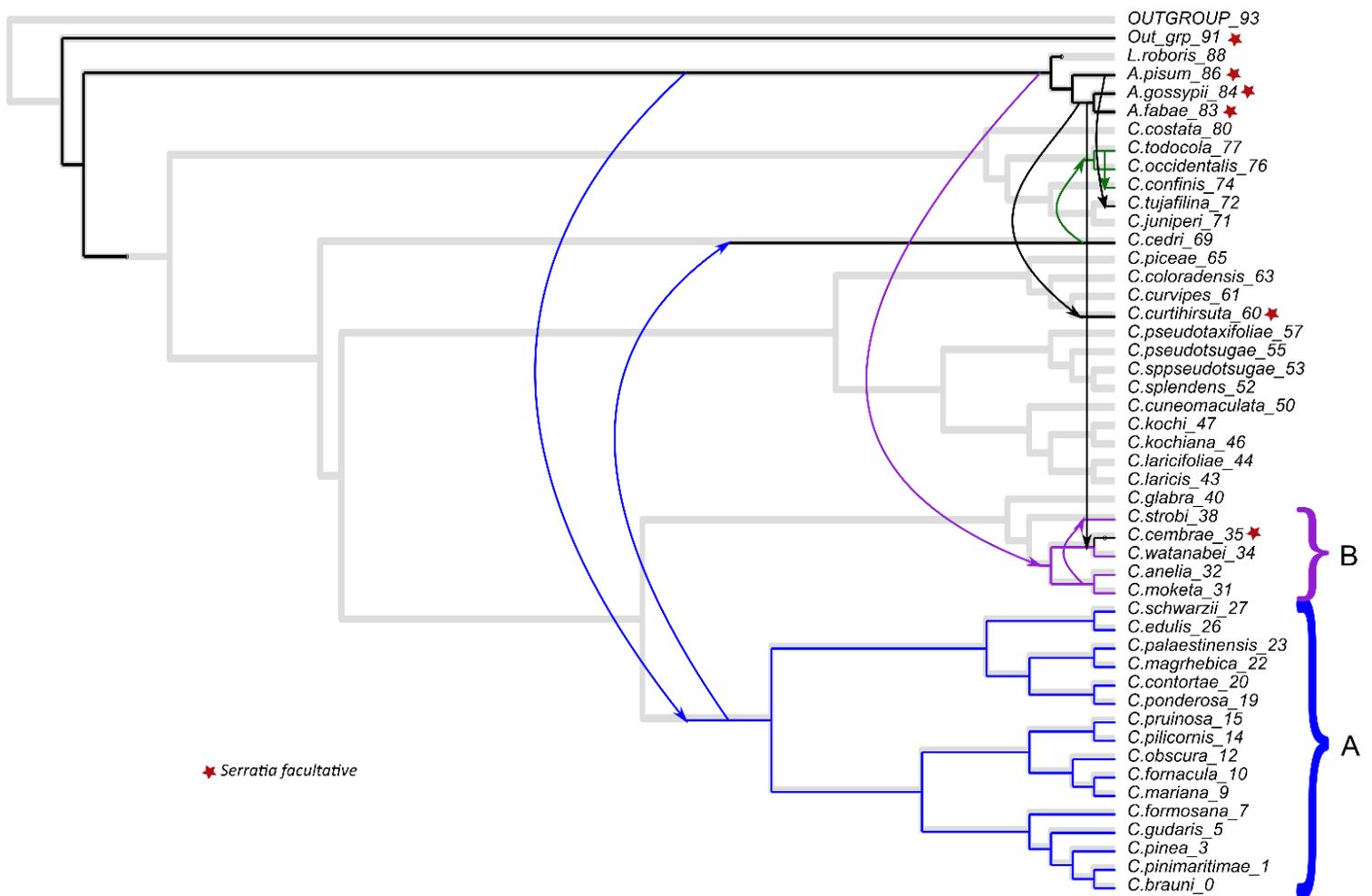


Figure 4 : Réconciliations de la phylogénie des *Serratia* sur la phylogénie des *Buchnera* de pucerons (arbre gris en fond) réalisées avec Mowgli, et visualisées avec SylvX. La topologie de *Serratia* est issue de l'inférence en Maximum de Vraisemblance (IQtree) et celle de *Buchnera* de l'inférence Bayésienne (Phylobayes) Les *Serratia* facultatives sont représentées avec une étoile rouge. Les transferts de symbiote sont représentés par des flèches. Les clades A et B sont représentés par des accolades.

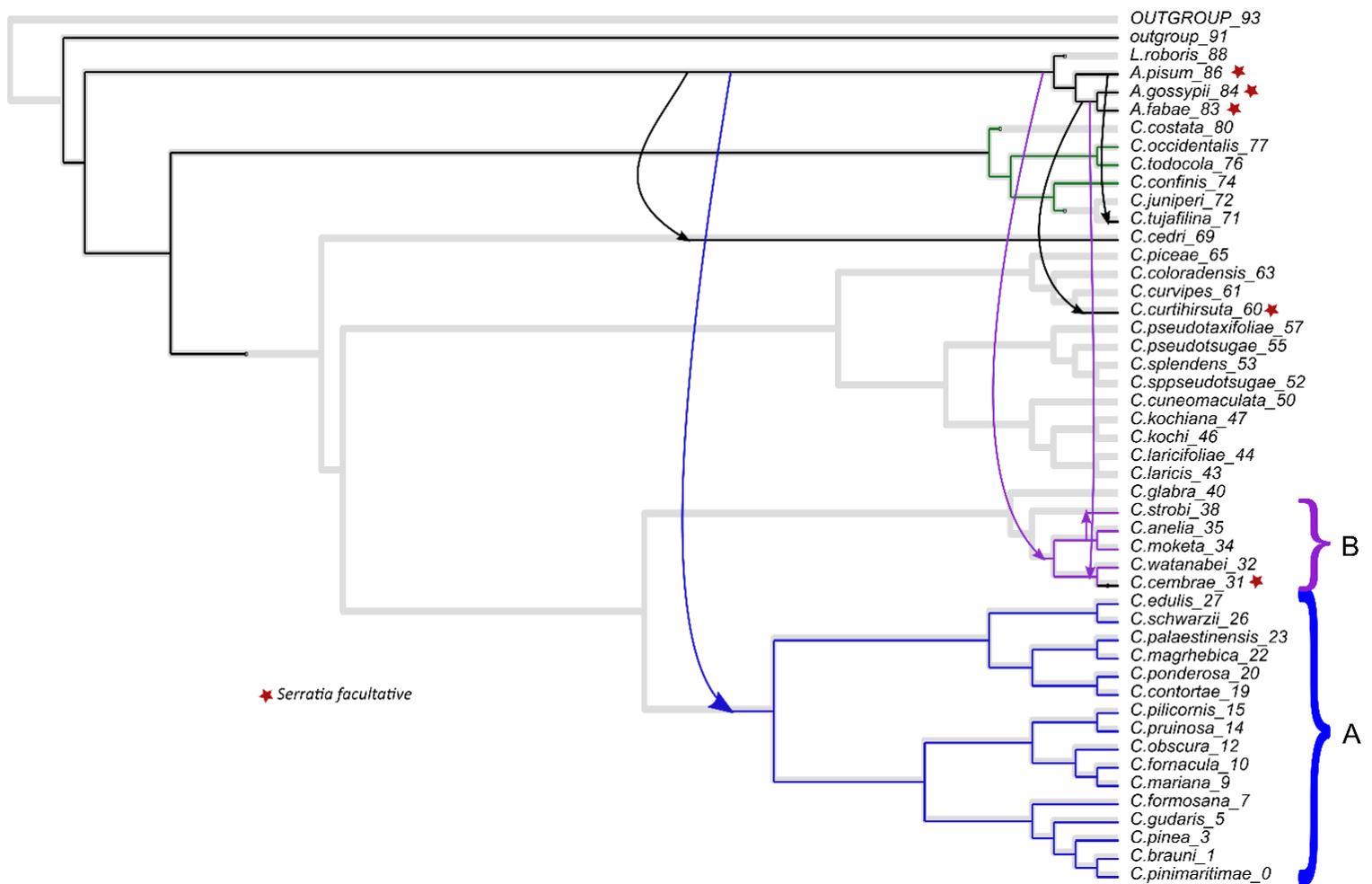


Figure 5 : Réconciliations de la phylogénie des *Serratia* sur la phylogénie des *Buchnera* de pucerons (arbre gris en fond) réalisées avec Mowgli, et visualisées avec SylvX. La topologie de *Serratia* est issue de l'inférence Bayésienne (Phylobayes) et celle de *Buchnera* de l'inférence Bayésienne (Phylobayes). Les *Serratia* facultatives sont représentées avec une étoile rouge. Les transferts de symbiote sont représentés par des flèches. Les clades A et B sont représentés par des accolades.

FISH

Les microscopies par FISH ont révélé que les *Serratia symbiotica* dans le clade A sont dans des bactériocytes indépendants des bactériocytes de *Buchnera* et présentent une forme de cellule coccoïde plus petite que les cellules de *Buchnera* (Figure 6). Ces résultats sont en accord avec nos attentes, puisque les *Serratia* ont des génomes très réduits (~1 Mbp) dans ce groupe et sont similaires à ce qui a pu être décrit dans les endosymbiotes d'autres espèces de pucerons avec un génome de symbiote fortement réduit (Manzano-marín, 2022; Manzano-Marín et al., 2017).

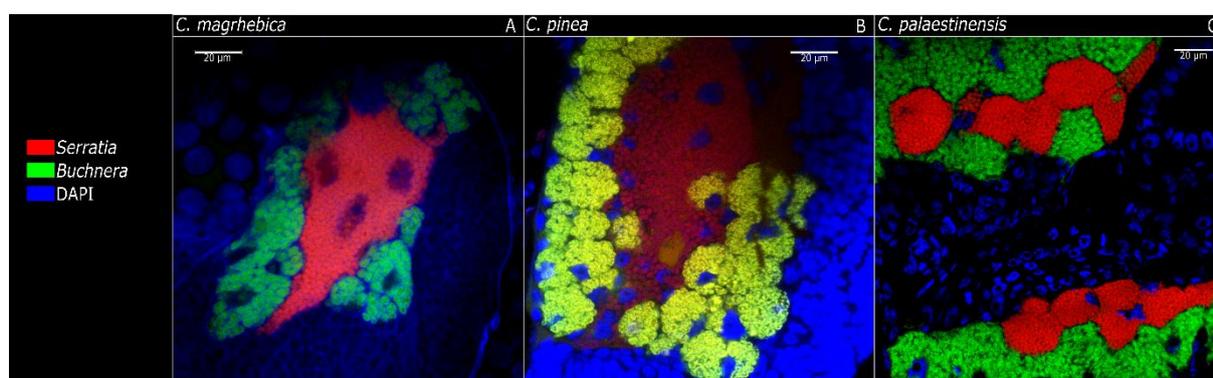


Figure 6 : Images microscopiques FISH fusionnées d'embryons de pucerons provenant de pucerons *Cinara* sélectionnés. Le signal de *Serratia* est indiqué en rouge, celui de *Buchnera* en vert. Le DAPI (qui colore l'ADN et met en évidence les noyaux de l'hôte) est visible en bleu. Le nom scientifique de chaque espèce est indiqué en haut de chaque panneau et le code couleur de chaque sonde fluorescente et de son groupe cible est indiqué dans chaque cadre.

Pour ce qui est du clade C, les *Serratia* observées dans des bactériocytes sont coccoïdes, mais leur taille semble différente selon l'espèce. Nous avons mis en évidence la présence d'une *Serratia* de très grande taille dans des bactériocytes spécifiques chez *C. confinis* (environ 4 µm ; Figure 7). Pour *C. pectinatae* (proche de *C. occidentalis*, *C. costata* et *C. todocola*), nous avons mis en évidence des *Serratia* de petite taille dans des bactériocytes spécifiques (environ 2 µm, comme chez *C. cedri* ; Figure 7), cela pourrait suggérer que celle-ci est plus anciennement acquise. Ces analyses semblent en accord avec nos analyses phylogénomiques montrant plusieurs acquisitions indépendantes de *Serratia symbiotica*. De plus, de précédentes observations sur *C. tujafilina* ont révélé la présence de *Serratia* localisées hors des bactériocytes à la périphérie des bactériocytes de *Buchnera*, et celles-ci ont la particularité d'être bacilliformes (Manzano-Marín et al., 2017).

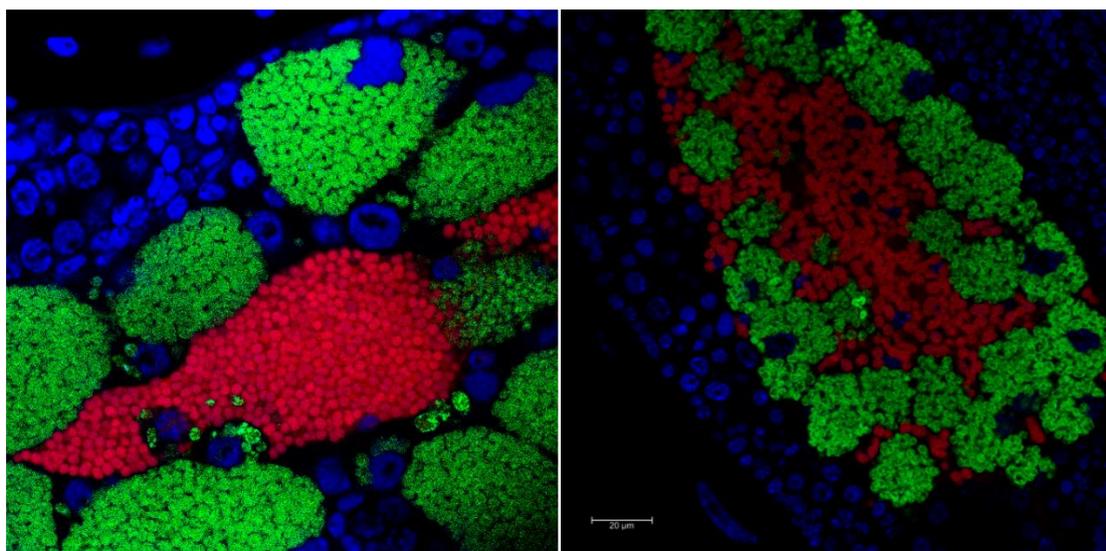


Figure 7 : Images microscopiques FISH fusionnées d'embryons de pucerons provenant de pucerons *Cinara pectinatae* à droite et *Cinara confinis* à gauche sélectionnés. Le signal de *Serratia* est indiqué en rouge, celui de *Buchnera* en vert. Le DAPI (qui colore l'ADN et met en évidence les noyaux de l'hôte) est visible en bleu.

IV. Discussion

Y-a-t-il eu cospéciation entre les trois partenaires de cette symbiose tri-partite ?

Pour les 49 *Buchnera aphidicola* étudiées ici, nous reportons des tailles de génome et des teneurs en G+C faibles et relativement similaires entre elles (entre 0.42 et 0.44 Mb, et entre 20 et 26 % de G+C). Ces résultats confirment qu'il y a eu une réduction des génomes ancestrale dans le genre *Cinara* et probablement avant la diversification des Lachninae. En effet, une étude sur *Tuberochnaus salignus*, une autre espèce de Lachninae montre que la *Buchnera* associée est de taille réduite (0.44 Mb) (Manzano-Marín et al., 2016). Dans de précédentes études, dans lesquelles les génomes de *Buchnera* de différentes espèces de *Cinara* ont été partiellement annotés et comparés à la *Buchnera* associée au puceron du pois, il a été montré que cette réduction était concomitante à la perte de fonctions essentielles chez *Buchnera* (synthèse de Biotine et Riboflavine) qui a accompagné l'acquisition de co-symbiote obligatoire dans cette sous-famille (Manzano-Marín et al., 2016; Meseguer et al., 2017).

Les analyses phylogénomiques utilisant les génomes mitochondriaux des pucerons et les génomes entiers de *Buchnera* ont permis de confirmer que les deux partenaires ont des histoires phylogénétiques parallèles avec une congruence parfaite de leur phylogénie respective. Ces résultats étaient attendus, puisque l'acquisition de *Buchnera* est considérée comme ancestrale aux Aphididae (Munson et al., 1991) et toutes les études phylogénétiques ayant testé

l'hypothèse de cospéciation entre les deux partenaires obligatoires l'ont confirmée dans différentes lignées de pucerons (Jousselin et al., 2009; Liu et al., 2013; Moran, 1993).

Similairement, nous montrons que la plupart des *Serratia* présentes comme co-symbiote des *Cinara* ont des génomes réduits avec un faible taux de G-C symptomatique d'une association obligatoire à long terme avec leurs hôtes (Moran et al., 2008). Cependant, leurs niveaux de réduction et leur taux de G-C observés sont différents selon l'espèce hôte révélant alors la présence de plusieurs souches assez divergentes de *Serratia*. Les reconstructions phylogénomiques de la bactérie *Serratia symbiotica* ainsi que les analyses de réconciliations entre *Buchnera* et *Serratia* ont quant à elles révélé une histoire évolutive complexe. Les lignées de *Serratia* sont ainsi probablement issues de plusieurs acquisitions indépendantes plus ou moins récentes au sein des *Cinara*. Une de ces acquisitions indépendantes de *Serratia* est à l'origine du clade A pour lequel il semble y avoir diversification parallèle de l'hôte et de ses deux symbiotes. Cette histoire et ses conséquences sur l'évolution du génome des deux symbiotes sera l'objet du chapitre 2.

D'après nos analyses de réconciliation, une acquisition de *Serratia* indépendante suivie d'une cospéciation semble avoir eu lieu dans un autre clade de *Cinara* (le clade B : Figure 4, 5). Au sein de ce clade B, les bactéries ont aussi des génomes réduits et des taux de G-C faibles ce qui suggère comme dans le clade A une longue histoire co-évolutive avec leurs hôtes.

Par contre, les *Serratia* obligatoires associées à deux espèces de *Cinara* sont très similaires aux facultatives et apparaissent phylogénétiquement proches dans nos reconstructions phylogénétiques. Elles forment ainsi un clade (clade C) qui est le plus proche de la bactérie libre *Serratia marcescens* dans nos phylogénies avec des branches extrêmement courtes comparées aux autres bactéries. C'est le cas des *Serratia* associées à *C. tujafilina* et à *C. cembrae* chez lesquelles la taille des génomes (environ 2.5Mb) est proche des facultatives décrites chez *A. pisum*, *C. curtihirsuta*, et *A. fabae* (respectivement 2.8 Mb, 2.8 Mb et 3.6 Mb) et dont les taux de G-C sont identiques (environ 50%). Cela suggère que la transition d'un mode de vie facultative à obligatoire de ces *Serratia* chez ces hôtes est probablement récente. Alternativement, la présence de ces courtes branches dans la phylogénie et le regroupement de ces *Serratia* avec les symbiotes facultatifs pourraient être un artéfact issu d'échanges plus importants entre ces bactéries qui homogénéiseraient davantage leurs génomes (cela pourrait être favorisé par une localisation extracellulaire) ou encore un taux de mutation plus faible chez ces bactéries comparées aux taux élevés caractéristiques des anciens symbiotes. Il est cependant

plus probable, étant donnée la dynamique évolutive de ce système et les multiples remplacement de *Serratia symbiotica* par d'autres symbiotes, que ces *Serratia* représentent une acquisition indépendante de celles du clade A ou du clade B.

Finalement, pour ce qui est des souches de *Serratia* associées à *C. confinis*, *C. occidentalis*, *C. cedri*, *C. todocola*) leurs positions dans les phylogénies sont variables selon les méthodes de reconstructions utilisées. Ces reconstructions génèrent ainsi des scénarii de réconciliations alternatifs complexes (et parfois peu probables biologiquement car supposant des transferts de symbiotes obligatoires entre espèces) avec des origines potentiellement différentes de *Serratia* chez ces espèces de pucerons et de potentiels évènements de co-spéciation avec les pucerons hôtes. Les génomes des endosymbiotes obligatoires étant en quelque sorte aberrants (très biaisés, très petits), ils peuvent se regrouper artificiellement dans des phylogénies par convergence dans leurs caractéristiques (Husník et al., 2011; Murray et al., 2016). Malgré l'utilisation de modèles d'évolution adaptés à pallier au phénomène d'attraction des longues branches, l'histoire de leur association avec leurs hôtes est donc toujours complexe à inférer. Une façon de consolider nos résultats serait de pouvoir mieux explorer l'architecture des génomes de *Serratia* : la synténie des génomes indique en effet que les souches sont très proches. Malheureusement, certains génomes étaient trop fragmentés pour conduire cette analyse à l'échelle du genre *Cinara*.

Ces inférences issues de nos réconciliations sont en partie corroborées par les images FISH obtenues sur sept espèces (*C. pinea*, *C. magrhebica*, *C. palaestinensis*, *C. cedri*, *C. confini*, *C. pectinatae* et *C. tujafilina*) qui montrent l'existence de caractéristiques cellulaires différentes pour *Serratia* selon l'espèce hôte. A l'exception de *C. tujafilina*, ces images montrent la présence de deux types de bactériocytes proches l'un de l'autre, l'un contenant des cellules de *Buchnera* et l'autre des cellules de *Serratia*. Les *Serratia* observées dans des bactériocytes sont coccoïdes, mais leur taille semble différente selon l'espèce, avec dans le clade A et dans *C. cedri* et *C. pectinatae* des bactéries de petite taille (environ 2 µm) comparée aux grandes bactéries de *C. confinis* (clade C) (environ 4 µm). Ainsi ces différences de taille et de forme observées peuvent être également le reflet d'acquisitions indépendantes de *Serratia*.

Une histoire de co-diversification complexe : quels processus expliquent cette dynamique ?

Ainsi, nous suggérons dans cette étude que plusieurs acquisitions indépendantes de *Serratia* ont eu lieu au sein des *Cinara*, mais aussi probablement des remplacements de *Serratia*. En effet, plusieurs autres bactéries, *Erwinia*, *Fukatsuia*, *Sodalis* sont associées à diverses lignées de *Cinara* (Meseguer et al., 2017) et nos reconstructions suggèrent qu'une *Serratia* « ancienne » est parfois remplacée par une plus jeune. Afin de comprendre l'histoire évolutive de ces associations, et en particulier l'origine des transferts il serait intéressant de mieux comprendre les sources de transfert potentielles en échantillonnant à fois plus d'espèces de *Cinara* mais aussi en appréhendant la diversité intra-espèces des symbiotes chez les *Cinara* ainsi que la diversité des bactéries qu'ils peuvent rencontrer dans leur environnement. En effet, Meseguer *et al.* (2017), ont montré que certaines espèces de *Cinara* pouvaient héberger deux souches de *Serratia* (deux haplotypes 16S bien distincts). La cooccurrence de deux souches (une obligatoire, une facultative) peut expliquer que l'une remplace l'autre assez facilement. Cela nécessite que la nouvelle *Serratia* puisse supplanter la première dans le bactériome du puceron.

Dans les systèmes di-symbiotiques des pucerons étudiés jusqu'à présent, la similarité des structures et localisations des « nouveaux » symbiotes avec les bactériocytes de *Buchnera*, suggère que les co-symbiotes suivent tous les mêmes voies de transmission verticale que celle de *Buchnera*. Cela interroge alors sur le mode d'acquisition et de mise en place de ces co-symbiotes dans les différentes espèces, celui-ci pouvant être par colonisation de bactériocytes de *Buchnera*, ou encore par remplacement d'un ancien co-symbiote dans ses bactériocytes.

L'histoire évolutive complexe des *Cinara* n'est pas unique puisque des résultats similaires démontrant l'instabilité des systèmes di-symbiotiques ont été observés dans plusieurs clades d'hémiptères. En effet, dans plusieurs systèmes d'études des remplacements de co-symbiotes ont été observés à de grandes échelles de temps évolutif (Dial et al., 2022; Husnik and McCutcheon, 2016; Koga et al., 2013; Rosenblueth et al., 2012; Szabó et al., 2017; Toenshoff et al., 2012; Urban and Cryan, 2012).

Les scénarii évolutifs expliquant cette dynamique de multiples remplacement de co-symbiotes sont encore incertains. Meseguer et al (2017) a exploré si différents facteurs écologiques (i.e. le genre de plante hôte, le régime alimentaire, la niche climatique) étaient corrélés à l'identité du co-symbiote présent mais aucun facteur significatif n'a été identifié par

ces analyses comparatives. Afin d'approfondir cette hypothèse, une analyse des capacités de chaque co-symbiote permettrait de vérifier s'il existe des différences dans les capacités de complémentation de la symbiose en nutriments essentiels par les différentes lignées de *Serratia* ou encore si celles-ci peuvent amener d'autres compétences (e.g. une résistance à des ennemis naturels). Ainsi, il est possible qu'une souche apporte une nouvelle fonction à l'hôte, qui favorise son installation. Une hypothèse alternative à celle-ci serait que les remplacements puissent se faire avec toute bactérie présente dans le microbiote et capable de remplir les fonctions de complémentation des pertes de gènes de *Buchnera*. Ainsi, à partir du moment où *Buchnera* aurait perdu certaines fonctions essentielles, seule la « disponibilité » relative des différentes bactéries au sein des pucerons engendrerait des remplacements successifs de souches en compétition pour la même niche (Meseguer et al., 2017).

Conclusion

Dans l'ensemble, nos résultats démontrent que la symbiose entre *Buchnera* et la bactérie endosymbiotique *Serratia symbiotica* est apparue à de multiples reprises de façon indépendante au sein du genre *Cinara*. Plus généralement, nous pouvons affirmer que l'acquisition de nouveau co-symbiote a évolué à de multiples reprises dans ce genre de puceron. Cela montre que l'association des pucerons *Cinara* avec leur co-symbiote est encore plus dynamique que ne le laissaient penser les travaux de Meseguer et al. (2017) et Manzano-Marín et al. (2017). Ces résultats sont cohérents avec ce qui est observé pour d'autres hémiptères (Cigales : Husnik and McCutcheon, 2016 ; Adelgidae : von Dohlen et al. 2017) montrant que même dans des endosymbioses très anciennes, de nouveaux co-symbiotes peuvent être acquis et que ces systèmes sont particulièrement dynamiques. Cependant, l'origine et les mécanismes de ces remplacements successifs de symbiotes restent à explorer que ce soit le résultat de compétition entre symbiotes et/ou d'une sélection par l'hôte.

Chapitre 2
**Dual nutritional symbionts in conifer feeding
aphids: cospeciation and parallel evolution in
symbiont genomes**

Contexte de l'étude

Lors du premier chapitre, nous avons décrit l'histoire évolutive de l'acquisition du co-symbiote *Serratia* chez les 45 espèces de *Cinara* étudiées. Dans les résultats de celui-ci, nous avons observé la présence d'un clade de treize espèces de *Cinara* pour lequel il semble y avoir eu une co-spéciation des trois partenaires de la symbiose (puceron, *Buchnera* et *Serratia*). Nous avons donc utilisé pour ce chapitre 2, ce clade de treize espèces comme sous jeu de données, pour lequel j'ai annoté l'ensemble des génomes bactérien et mitochondriaux.

Ce deuxième chapitre présente donc les analyses effectuées à partir de ces treize génomes symbiotiques et mitochondriaux, ayant pour but de décrire l'évolution des génomes de *Serratia* et de *Buchnera* dans le contexte d'une co-spéciation au sein d'un clade de *Cinara*. Cela permettra de déterminer si les deux bactéries symbiotiques ont évolué de la même manière au cours du temps, mais aussi d'observer les conséquences sur les génomes et leurs compositions en gènes d'une telle symbiose.

Pour cela, les objectifs de cette partie sont donc de : **1)** Vérifier l'existence d'une co-spéciation à l'échelle d'un clade de 13 espèces de *Cinara*. **2)** Décrire les changements génomiques au sein des génomes bactériens (taille, taux de GC, inversion). **3)** Annoter et décrire l'ensemble des pertes de gènes ayant eu lieu. **4)** Mesurer et comparer les taux de substitution des deux bactéries. **5)** Caractériser les régimes de sélection qui s'exercent sur les gènes des deux bactéries symbiotiques.

Les hypothèses de cette étude sont que les génomes de *Serratia symbiotica* comme ceux de *Buchnera* sont relativement stables dans le temps avec des tailles, taux de GC similaires et pas d'inversions. On suppose que la présence de deux bactéries symbiotiques est associée à la perte de nombreux gènes et notamment les gènes redondants du métabolisme. Compte tenu de la bibliographie, on s'attend à observer des taux d'évolutions différents entre les deux bactéries symbiotiques au cours de leur co-spéciation. Finalement, nous espérons observer un relâchement des pressions de sélection purifiante sur certains gènes bactériens du fait de leur redondance fonctionnelle ou de leur inutilité dans un environnement symbiotique.

Abstract :

Nutritional obligate symbionts have evolved many times with insects that live on nutrient poor diets. For instance, most aphids are engaged into an obligate symbiosis with a gammaproteobacterium, *Buchnera aphidicola*, on which they rely for essential amino acids and B vitamins provisioning. During the long-term co-evolution between *Buchnera* and aphids, *Buchnera* has undergone genome erosion and specialization towards metabolic functions that benefit their hosts. In several aphid clades, *Buchnera* has lost some of these functions and dual symbiosis systems have evolved between aphids, *Buchnera* and a younger bacterial co-symbiont, often *Serratia symbiotica*, which complements *Buchnera*. Though *Serratia* is found as a co-obligate partner in many aphid species, it is unknown if and how long it co-speciates with its hosts. It is also unknown whether this co-symbiont evolves at the same pace as *Buchnera* and whether its presence, that creates gene redundancy with the primary symbiont, modifies the selective regime of the primary symbiont and promotes further genome erosion. Here we used - endosymbiont whole genome sequences of thirteen species in a clade of *Cinara* aphids that harbour *Serratia symbiotica* as a co-obligate partner of *Buchnera*. Through phylogenomics and datation analyses, we reveal a perfect cospeciation pattern and inferred identical substitution rates in both symbionts. The newly sequenced *Serratia* strains are characterized by small (~1Mbp), A+T rich and syntenic genomes. Fluorescence in-situ hybridisation reveals that *Serratia* and *Buchnera* inhabit independent bacteriocytes. Investigating both interspecific and intraspecific variations in both symbionts in two *Cinara* species, we observe low rate of indels and nonsynonymous substitution for almost all coding genes in the two symbionts. We found no evidence for relaxed selection on protein-coding genes that are shared between symbionts but accumulation of indels in non-coding sequence were found in *Serratia* genome, suggesting an ongoing genomic reduction in this new nutritional symbionts. Our results show that this di-symbiotic system has undergone long-term co-evolution and that two symbionts with different ages of acquisitions and stages of degeneration can evolve at the same pace. We can therefore hypothesize that both symbionts are subject to the same demographics events and the same selection regime.

I. Introduction

Microbial symbioses have independently appeared countless times in distant insect clades bringing new physiological processes; favouring the colonization of ecological niches by their hosts (McFall-Ngai et al., 2013; Moran, 2007; Sudakaran et al., 2017, 2015; Wernegreen, 2012). Colonisation of nutrient-poor substrate, like blood or phloem-sap, is a well-known case where the metabolic capacity brought by endosymbionts have enabled insect ecological diversification. Indeed, blood and phloem- sap being poor in essential amino acid and/or vitamins; arthropods that are obligate blood or sap feeders depend on nutritional obligate endosymbiont association for nutrient supplementation (Baumann, 2005b; Douglas, 2009; Duron and Gottlieb, 2020; Meseguer et al., 2017; Moran et al., 2008b; Wernegreen, 2012).

Over evolutionary time, host and obligate endosymbionts have coevolved and become inextricably linked leading to extreme modifications of symbiotic microbes and their inaptitude to live outside of their host (Moran et al., 2008b). Bacterial symbionts of arthropods are generally localised inside specific host cells (bacteriocytes) that can be grouped as an organ (bacteriome) and this compartmentalization is coupled with vertical transmission (Aksoy, 1995; Braendle et al., 2003; Koga et al., 2012; Toenshoff et al., 2012; Buchner, 1953). As a result of this particular lifestyle, endosymbiont genomes have converged towards very specific characteristics (Moran et al., 2009; Wernegreen, 2002). They harbour greatly reduced genomes having lost many genes in comparison to their free-living relatives and display biased nucleotide composition with low GC content (McCutcheon et al., 2019; McCutcheon and Moran, 2012; Toft and Andersson, 2010).

Gene losses in endosymbionts are not random, genes associated with free-living habit such as genes involved in the production of cell envelope components stress response synthesis (Moran, 2002; Moran et al., 2008b; Moran and Wernegreen, 2000), are often lost which then reinforces the interdependency between the bacteria and their host. In opposition genes that serve symbiotic functions such as genes involved in essentials amino acids and vitamins synthesis are conserved (Chong et al., 2019; Clayton et al., 2016; Hansen and Moran, 2011; Tamas et al., 2002; Toh et al., 2006). Consequently, obligate symbiont genomes are among the smallest known bacterial genomes, with a high proportion of coding sequences needed to complement the need of the host (Luan et al., 2015; Manzano-Marín et al., 2015; Manzano-Marín et al., 2020a; Mao et al., 2018; McCutcheon and von Dohlen, 2011; Tamas et al., 2001; Weglarz et al., 2018; Zientz et al., 2004). Within the existing diversity of endosymbiotic

systems, the degree of genome reduction and the functional categories of the genes that have been lost vary greatly and may depend on the role carried by the symbiont (Chong et al., 2019; Moran et al., 2008b; Plague et al., 2008; Vasquez and Bennett, 2022; Wernegreen, 2002). The study of different endosymbiont lineages has allowed a description of the processes leading to their genome reduction and specialisation. These mechanisms include purifying selection that acts on genes involved in essential functions for the bacterium and its host, and positive selection on genes involved in the adaptation to the host (e.g. genes involved in host-symbiont exchange functions) (Chong et al., 2019; McCutcheon et al., 2009; Tamas et al., 2001; Vasquez and Bennett, 2022). In contrast, genes unnecessary for intracellular life-style and not needed in the symbioses are lost through strong genetic drift induced by lack of selection. Vertical transmission that creates successive bottlenecks in endosymbiont populations also results in radical changes in population structures that result in lowered efficacy of selection even on required genes. The consequence is a greater rate of fixation of mildly deleterious mutations. The lack of recombination due to segregation in bacteriocytes, and loss of genes (like DNA repair) accelerates the accumulation of irreversible mutations (Muller ratchet) (Gabriel et al., 1993; Moran, 1996; Moran et al., 2009; Woolfit, 2003)

In some cases, this process of genomic erosion can lead to the loss of essential metabolic functions with negative consequences on host fitness. One way to compensate for this loss is to acquire a secondary obligate symbiont that either replaces or complements the primary symbiont metabolic losses (Koga and Moran, 2014; Lamelas et al., 2011b, 2008; Manzano-Marín et al., 2020a), these new symbionts are often called co-obligate symbionts. Nutritional symbioses involving two partners have been found in blood feeders (ticks (Buysse et al., 2021; Liu et al., 2016), and tsetse flies (Dalet and Maudlin, 1999; Nováková and Hypša, 2007)) and various lineages of sap-feeders within Hemiptera such as Psyllidae, whiteflies, Cicadas, coccids (Dial et al., 2022; Douglas, 2016; Meseguer et al., 2017), but also Diptera (Dalet and Maudlin, 1999; Nováková and Hypša, 2007). Despite being common, the origin and evolutionary stability of multi-endosymbiotic systems are not well described. One feature commonly found in these biological systems is a high endosymbiont turnover and multiple independent acquisitions during host diversification (Bennett and Moran, 2015; Koga et al., 2013; Meseguer et al., 2017; Michalik et al., 2021; Szabó et al., 2020, 2017). Nevertheless, the evolutionary span of multi-partner endosymbioses are not often documented: it is not known how long co-obligate symbionts stay on board throughout the evolutionary history of their host and how or why they get replaced or in some cases manage to evict the primary companion. Furthermore, the

consequences of their presence on primary symbiont genome evolution is seldom investigated (but see Vasquez and Bennet 2022).

Aphids (Hemiptera: Aphididae) represent a well-known insect group that exhibits long-term co-speciation with a primary symbiont. Indeed, they have been associated with *Buchnera aphidicola* (Gammaproteobacteria: Enterobacteriaceae) that provides the missing nutrient from their phloem sap diet for more than 150 MY and co-speciation between aphids and this symbiotic partner has been demonstrated at different evolutionary scales (Baumann et al., 1995; Dohlen and Moran, 2000; Funk et al., 2000; Jousselin et al., 2009). These mutualistic partners are localized in bacteriocytes (Buchner, 1953b). The evolutionary dynamics of *Buchnera* genome is well studied ; it exhibits a reduced genome, a low GC content and a fair part of its genome is specialised in the host metabolic complementation (Chong et al., 2019; Dohlen and Moran, 2000; McCutcheon et al., 2009; Tamas et al., 2002, Manzano-Marín et al. 2022). Recently, several studies have shown that *Buchnera* is not always the sole provider of essential amino acids and vitamins in aphids. In several aphid sub-families, a co-obligate symbiont co-exists with *Buchnera* and both of them are essential for their host nutrition (Lamelas et al., 2008; Manzano-marín, 2022; Manzano-Marín et al., 2017; Meseguer et al., 2017; Monnin et al., 2020; Yorimoto et al., 2022). This has first been reported in species of the *Cinara* genus (Aphididae: Lachninae) (Lamelas et al., 2008; Pérez-Brocal et al., 2006). In all *Cinara* species studied so far, and actually all aphid species from the same sub-family, *Buchnera* has lost the ability to synthesise two essential B vitamins: biotin (B7) and riboflavin (B2) (Manzano-Marín et al., 2016; Manzano-Marín and Latorre, 2016; Meseguer et al., 2017). Consequently, aphids and their associated *Buchnera* rely on an additional endosymbiont for B7 and B2 synthesis. These co-symbiotic partners belong to diverse gammaproteobacteria *Fukatsuia symbiotica*, *Erwinia*, *Sodalis* and in most species *Serratia symbiotica* (Manzano-Marín et al. 2017, Manzano-Marín et al., 2020; Meseguer et al., 2017;).

A recent study has shown that a strain of *Erwinia*, named *Erwinia haradae* has co-speciated with a specific clade of *Buchnera* and their host without being replaced for at least 20 millions years, and preserving genes that complements *Buchnera* (Manzano-Marín et al., 2020a). However the historical association with *Serratia symbiotica* remains to be investigated, though ubiquitous as a co-obligate-symbiont of *Cinara* species (Meseguer et al., 2017), we do not know how long it is maintained over evolutionary times. Genome studies of two *Serratia* associated with *Cinara*, (*C. cedri* and *C. tujafilina* (Lamelas et al., 2011b, 2008; Manzano-Marín and Latorre, 2016, 2014) suggest that they can be at different stages of genomic erosion

(they exhibit differences in term of genome size, GC content and gene content) (Gosalbes et al., 2008a; Lamelas et al., 2011a; Manzano-Marín and Latorre, 2014) which might suggest independent acquisitions. Furthermore, little is known about the pattern of genome evolution of these newly acquired symbionts. We do not know if the evolutionary rate of co-obligate symbiont is similar to the observed rate of *Buchnera*. It is also possible that the co-symbiont presence influences *Buchnera* genome, relaxing selective pressures on symbiotic genes (i.e. genes involved in the host metabolism) that are redundant between the two symbionts

In the present study, we focused on a clade of *Cinara* described in (Meseguer et al., 2017, 2015) as Clade A known to be associated uniquely with *Serratia symbiotica* as a cosymbiotic partner of *Buchnera*. We used phylogenomic analyses to investigate whether *Serratia symbiotica* derived from a single or multiple acquisitions of facultative *Serratia*. We further assessed substitution rates in both *Buchnera* and *Serratia* genomes based on interspecific and intraspecific genetic variation in order to determine whether *Buchnera* and *Serratia* evolve at the same pace. We finally looked at genome wide signature of selection in both symbionts in order to investigate which evolutionary forces shape symbiont genome evolution in this tripartite system. Altogether, these analyses aim at providing an overview of the evolutionary processes that govern the evolution of obligate partners in multi-partner endosymbioses.

II. Materials and methods

Aphid collection, DNA extraction and sequencing

From aphid colonies samples reported in (Meseguer et al., 2015) and (Meseguer et al., 2017), stored in 70% ethanol at 6 °C in the CBGP collection, we used samples from twelve species from a clade of *Serratia*-associated *Cinara* aphid (supplementary Table S1) and for two species in this clade, *Cinara mariana*, *Cinara fornacula* we used respectively 7 and 6 independent samples. We prepared DNA samples enriched with bacteria as described in (Jousselin et al., 2016) for whole genome sequencing. For this, 15 to 20 aphids from a single colony were pooled together. Extracted DNA was used to prepare two custom paired-end libraries. Briefly, 5 ng of genomic DNA were sonicated, using the E220 Covaris instrument (Covaris, USA). Fragments were end-repaired, 3'-adenylated, and NEXTflex PCR free barcodes adapters (Bioo Scientific, USA) were added by using NEBNext® Ultra II DNA library prep kit for Illumina (New England Biolabs, USA). Ligation products were then purified by

Ampure XP (Beckman Coulter, USA) and DNA fragments (>200 bp) were PCR-amplified (2 PCR reactions, 12 cycles), using Illumina adapter-specific primers and NEBNext ® Ultra II Q5 Master Mix (NEB). After library profile analysis by Agilent 2100 Bioanalyser (Agilent Technologies, USA) and qPCR quantification, using the KAPA Library Quantification Kit for Illumina Libraries (Kapa Biosystems, USA), the libraries were sequenced, using 251 bp paired-end reads chemistry on a HiSeq2500 Illumina sequencer for the 12 species, and 100 bp paired-end reads chemistry for intraspecific data .

Fluorescence in situ hybridisation microscopy

From the afore mentioned twelve strains of *Serratia*-, we investigated symbiont localisation patterns in individuals from three species: *Cinara palaestinensis*, *Cinara magrhebica*, and *Cinara pinea*. Those were first fixed upon collection in modified Carnoy's fixative (6 chloroform: 3 absolute ethanol: 1 glacial acetic acid) and left overnight, following the protocol of (Koga et al., 2009). Individuals were then dissected in absolute ethanol to extract embryos and transferred into a 6% solution of H₂O₂ diluted in absolute ethanol and were then left in this solution for 2 weeks (changing the solution every 3 days). Embryos were then washed twice with absolute ethanol. Hybridization was performed overnight at 30 °C in standard hybridization buffer (20mM Tris-HCl [pH 8.0], 0.9M NaCl, 0.01% SDS, and 30% formamide) and then washed (20mM Tris-HCl [pH 8.0], 5 mM EDTA, 0.1M NaCl, and 0.01% SDS) before slide preparation (Manzano-Marín et al., 2020a). The embryos of up to 10 individuals were viewed under a confocal microscope. We used two competitive specific probes for Buchnera (BLach-FITC [5'-FITC-CCCGTTCGCCGCTCGCCGGCA-FITC-3'], from (Manzano-Marín et al., 2017)) and *Serratia* symbionts in *Cinara* (SeJR5Cinara-Cy3 [5'-Cy3-CCGTYCGCCGCTCGTCACC-Cy3-3'], this study). TIFF-formatted images were exported using the Leica application suite X. Exported images were imported into Inkscape v0.92.4 for building the published figures.

Genome assembly and annotation

For each library, Illumina reads were right-tail clipped (minimum quality threshold of 20), using FASTX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/). Reads shorter than 75 base pairs (bp) were dropped. Additionally, PRINSEQ v0.20.4 (Schmieder and Edwards, 2011) was used to remove reads containing undefined nucleotides as well as those

left without a pair after the filtering and clipping process. The resulting reads were assembled, using SPAdes v3.10.1 (Bankevich et al., 2012), with the option --only-assembler option and k-mer sizes of 33, 55, 77, 99 and 127. From the resulting contigs, those that were shorter than 200 bp were dropped. The remaining contigs were binned, using results from a BLASTX (Altschul, 1997) search (best hit per contig) against a database consisting of the pea aphid's proteome and a selection of aphid's symbiotic bacteria previously found in *Cinara* species and free-living relatives' proteomes (see Meseguer et al. 2017, Manzano-Marín et al 2020 for details on reference proteomes). The assigned contigs were manually screened using the BLASTX web server (searching against the nr database) to insure correct assignment.

This binning process confirmed the presence of *Serratia symbiotica*, as inferred from 16S rRNA gene fragment sequencing from several specimens (Jousselin et al., 2016; Meseguer et al., 2017) and the sequencing of *C. fornacula* endosymbionts reported in Meseguer et al. (2017). We also always retrieved a contig of about 16000 bp representing the aphid host mitochondrial genomes. The resulting contigs were then used as reference for read mapping and individual genome assembly using SPAdes, as described above, with read error correction.

We generally obtained a complete *Buchnera* in one to 40 contigs (amounting to a scaffold of about 460 Mb as other previously sequenced *Buchnera* strains associated with *Cinara*) plus the leucine and tryptophan plasmid (Table S2). *Serratia* genomes were scattered into 6 to 13 contigs (Table S3) that were scaffolded using as a reference *Serratia* associated with *C. fornacula* a species from clade A sequenced in Meseguer et al. (2017). For the 12 newly sequenced aphid species, and the previously sequenced endosymbionts of *Cinara fornacula*, the resulting *Buchnera aphidicola* and *Serratia symbiotica* genomes were annotated by submitting the data to the MicroScope platform pipeline (Vallenet et al., 2019). This annotation was followed by manual curation of the genes on Geneious v11.1.5 (<https://www.geneious.com>) and using the MicroScope platform (Vallenet et al., 2019) using BlastX against available *Buchnera* and *Serratia* genomes of *Cinara* (Manzano-Marín et al., 2020a). The resulting coding sequences (CDSs) were considered to be putatively functional if all essential domains for the function were found, if a literature search supported the truncated version of the protein as functional in a related organism, or if the CDS displayed truncations but retained identifiable domains (as shown by Tamas et al., 2008). Mitochondrial genome were annotated using MitoS WebServer (Bernt et al., 2013) followed by manual curation of the genes in Geneious v11.1.5.

Intraspecific variation in symbiont genomes in two *Cinara* species

Illumina reads were processed as previously described to obtain *Serratia* and *Buchnera* assemblies for each sample. When several contigs were obtained for a symbiont, genome scaffolding was conducted by whole genome alignment between resulting contigs and reference endosymbiont genomes of *C. mariana* and *C. fornacula* obtained above using Mauve Contig Mover algorithm (implemented in Geneious v11.1.5 (Darling et al., 2004). Finally, genome pair-wise alignments were visually verified and annotations were transferred from reference genome using Geneious tools (“transfer annotation”). Manual curation of start and stop codons of all CDS was conducted in Geneious v11.1.5.

Phylogenetic analyses

Cinara mitochondrial phylogeny

In order to reconstruct the phylogeny of the aphid hosts (clade A of *Cinara* species) we used all thirteen mitochondrial genes (atp6, atp8, cob, cox1, cox2, cox3, nad1, nad2, nad3, nad4, nad4L, nad5, nad6) from the 13 newly sequenced and annotated *Cinara* species. We selected three outgroup species (*A. pisum*, *C. cedri* and *C. tujafilina*) gathered from NCBI (Table S4).

For each gene, sequence alignments were conducted using Macse v2.03 (Ranwez et al., 2018) and visually checked using SeaView v4 (Gouy et al., 2010). We then removed divergent and ambiguously aligned blocks using Gblocks v0.91b (Talavera and Castresana, 2007) and concatenated the resulting alignments. A partitioned scheme was selected using PartitionFinder v2.1.1 (Lanfear et al., 2016) with one data block defined for each codon position (first, second, and third) in each gene. Phylogenetic inferences were performed with Maximum likelihood (ML) as implemented in IQ-TREE 1.6.8 (Nguyen et al., 2015) with the best PartitionFinder scheme and ultrafast bootstrap approximation with 1000 replicates. We also ran a Bayesian phylogenetic inference using MrBayes v3.2.7 (Ronquist et al., 2012) with the best PartitionFinder scheme available in MrBayes (according to Bayesian Information Criterion metric (BIC) under a greedy algorithm) and running two independent analyses with four chains each for 3,000,000 generations and checked for convergence.

Buchnera and *Serratia* Symbiont phylogenies

For performing symbiont phylogenetic inferences, we extracted orthologous protein-coding gene from previously annotated *Buchnera* and *Serratia* genome dataset on Geneious.

We then obtained 691 genes with 545 shared by all *Serratia* species and 372 genes with 357 shared by all *Buchnera* species. For each dataset, nucleic acid sequences were aligned using MAFFT v7.450 (maxiterate 1,000 localpair) (Kato and Standley, 2013) and concatenated. Divergent and ambiguously aligned blocks were removed using Gblocks. Substitution models on the concatenated datasets for each symbiont were selected using ModelFinder (Kalyaanamoorthy et al., 2017) (MF option implemented in IQ-TREE 1.6.8) using the Bayesian Information Criterion. ML phylogenetic inferences were conducted using IQ-TREE 1.6.8 (Nguyen et al., 2015) with GTR+F+R4 model and 1000 ultrafast bootstrap. Bayesian phylogenetic reconstructions were conducted with MrBayes v3.2.7 MPI (Ronquist et al., 2012) under the GTR+I+G model (Ronquist et al., 2012), running two independent analyses with four chains each for 300,000 generations (sampling every 1000 generations and 20% discarded as burn-in) and checked for convergence.

Examining patterns of genome evolution in co-evolving symbionts

Estimating evolutionary rates

Calibration of *Cinara*'s phylogeny was inferred in Beast v.1.8 (Drummond et al., 2012) over the previously described mitochondrial DNA matrix using the best PartitionFinder scheme. The selected priors were a Birth-Death model with an uncorrelated lognormal molecular clock and fixed tree topology using ML tree (obtained as previously described removing outgroup). We used previous studies to define three calibration points for the estimation of absolute ages of divergence and substitution rates. The crown age of *Cinara* clades A was dated as 25 Ma by (Meseguer et al., 2015) and used as root calibration age. Tree root was calibrated using a normal prior distribution: mean of 25, stdev of 3. We used the two oldest nodes of the tree as others calibration points (mean of 21, stdev of 2 and mean of 21.7, stdev of 2). We ran one MCMC chain of 70 million generations and assessed Convergence using Tracer 1.7 (Rambaut et al., 2018). After removing 20% of the samples as burn-in, we used TreeAnnotator 1.8 to generate a maximum clade credibility (MCC) tree with the median age estimate of nodes.

As there was a perfect congruence between the phylogenies of *Serratia*, *Buchnera* and their host (see results); we used host calibration point in order to calibrate symbiont phylogenies. Absolute ages of divergence events in symbiont were calculated as previously

described for the mitochondrial data, on both symbiont datasets, using GTR+I+G4 substitution model, and two MCMC chain of 300 million generations.

Evolutionary rate correlation

Dating analyses resulted in very similar estimate of the substitution rates by sites by millions of years for each symbiont (*Buchnera*, *Serratia*) (see results). In order to compare pace of genome evolution between *Serratia* and *Buchnera* we also investigated correlation of branch lengths along co-paths (Arab et al., 2020; Arab and Lo, 2020) from the ML trees for both symbionts and hosts. Using R (R Core Team, 2022), we plotted branch length (measured as substitution by site) along parallel branches of the phylogenies. We then tested the correlation between parallel branches of *Serratia* and *Buchnera* trees, but also between symbionts and host trees.

We then restricted the analyses to shared genes between *Buchnera* and *Serratia*, to investigate whether genes with the same functions evolved at the same pace. We first inferred genes sharing the same functional annotation between *Serratia* and *Buchnera*, ran trees as previously described, built phylogenies using these set of genes and we then plotted branch lengths of co-path for this data sub-set.

Test of positive and relaxed selection

Using the 13 *Buchnera* genomes and the 13 *Serratia* genomes within clade A, we performed molecular evolution analyses to identify genes in which positive selection has driven ongoing evolutionary change in encoded proteins. For each dataset (*Buchnera* and *Serratia*), we identified protein-coding genes shared by these 13 genomes (357 for *Buchnera* genomes and 545 for *Serratia* genomes). Genes shorter than 200 nucleotides were eliminated (seven genes for *Buchnera* and *Serratia*) and alignments were conducted using Macse v2.05 (Ranwez et al., 2018). Using the ML tree obtained above and gene alignments of each symbiont, we performed tests of diversifying selection on *Serratia* and *Buchnera* genes. BUSTED (Murrell et al., 2015), implemented in the HyPhy (Pond et al., 2005) and codeml program of the PAML v4.9 package (Yang, 2007) were used to compare neutral to positive selection models (Yang, 2007) for each gene (as in (Chong et al., 2019; Vasquez and Bennett, 2022)). BUSTED analysis were conducted using all branches as the foreground branch set to test for gene-wide positive selection across the tree. The M0 model was used to test for selection across all codon sites, ratio of nonsynonymous to synonymous substitutions ($\omega = dN/dS$) were generated for each gene. Then we used the ratio of nonsynonymous to synonymous substitutions to assess the selective

regime, where $\omega > 1$ indicate positive selection, $\omega = 0$ neutral selection and $\omega < 1$ purifying selection (Yang and Bielawski, 2000).

M0 model being not a sufficiently realistic model to detect signatures of positive selection that operate at finer scales (Anisimova et al., 2001). We tested specifically for positive selection. We used two nested models implemented in codeml program, M1a-M2a and M7-M8 (Anisimova et al., 2001). Both test are composed by a constrained model disallowing positive selection (M1a and M7) and an unconstrained model, allowing for positive selection (M2a and M8). We used chi-squared test; (p-value = 0.05) to compare likelihood scores within paired models to assess significant positive selection for genes. We subsequently considered uniquely genes identified as being under significant positive selection by both tests for downstream analysis.

We additionally test if selection is relaxed on genes shared by *Buchnera* and *Serratia* (redundant genes), we performed a test of selection relaxation on the 254 protein-coding genes shared by all symbionts genome (the 13 *Buchnera* and the 13 *Serratia*). As above, genes shorter than 200 nucleotides were eliminated and alignments were conducted using Macse v2.06. An ML tree was inferred by IQ-TREE v1.6.8 (Nguyen et al., 2015) with each concatenated gene alignment including both symbiont and using ModelFinder (Kalyaanamoorthy et al., 2017) to choose the best substitution model. Finally using the ML tree and gene alignment of each shared genes, we performed tests of relaxed selection with RELAX (Wertheim et al., 2015), implemented in HyPhy (Pond et al., 2005). RELAX analysis were conducted using *Buchnera* branches as the background branch and *Serratia* branches as the foreground branch set to test for gene-wide relaxed selection across the tree. In a second step, we set *Serratia* branches as the background branch and *Buchnera* branches as the foreground.

Intraspecific variations in symbiont genomes

For both species for which we had gathered intraspecific data (*Cinara mariana* and *Cinara fornacula*), we reconstructed both symbiont phylogenies to examine patterns of phylogenetic congruence at the intraspecific level. Maximum likelihood phylogenetic analyses were conducted as described for intraspecific data. For each aphid species and for each symbiont, we extracted all ortholog genes from annotated *Serratia* and *Buchnera* genomes using geneious v11.1.5 (<https://www.geneious.com>). For each symbiont orthologous genes were aligned using MAFFT v7.450 (maxiterate 1,000 localpair) (Katoh and Standley, 2013). Divergent and ambiguously aligned blocks were removed using Gblocks and genes alignment

were concatenated. Phylogenetic reconstruction was conducted with IQ-TREE v1.6.8 (Nguyen et al., 2015) (1000 ultrafast bootstraps) using -m MFP+MERGE option for substitution model selection (GTR+F+G4 were used for both symbionts for the species studied here).

We then plotted of these trees, loses of genes if any and inferred deletion/insertion and single nucleotide polymorphism (SNP) patterns as done previously by (Moran et al., 2009) on 9 genomes of *Buchnera* associated with the pead aphid. We then compared patterns of genomes evolution in *Serratia* and *Buchnera* for *C. mariana* and *C. fornacula*.

III. Results

Endosymbiont Gene content

The genomes of thirteen *Buchnera-Serratia* pairs from *Cinara* species were successfully sequenced, assembled and annotated. *Buchnera* genomes are all composed of a single circular chromosome and two plasmids that code for tryptophan (pTrp) and leucine (pLeu) synthesis. *Buchnera* genomes have a size between 461.94 and 465.15 kb with a G+C content between 23.3 and 28.1 % (Supplementary Table S2). Their gene contents are highly conserved with 366 to 372 proteins coding genes, which are almost identical to those encoded by *Buchnera* strains harboured by *Cinara* species associated with *Erwinia* as a co-symbiont (Manzano-Marín et al., 2020a). As for *Serratia* symbionts, once scaffolded we retrieved a chromosome and sometimes a plasmid for all strains; genomes have a size between 1.10 and 1.28 Mb with an average G+C content between 28.3 and 35.8 % (Supplementary Table S3). Their gene content are highly conserved with 631 to 681 protein coding genes. Finally, a single conserved copy of 16S/23S/5S rRNA operon were found in all 13 *Buchnera* genomes and 31 and 33 to 35 tRNAs in *Buchnera* and *Serratia* respectively.

Buchnera genomes are very stable with a variation of up to six genes between species and exhibit a perfect synteny between the thirteen lineages (Figure 1). Those variations were characterized by 12 genes (trmD, murE, hinT, ffs, argB, fliO, flgI, fis, yciC, FlgH, yoaE, orn; Figure 1) whose loss occurred in at least one lineage, representing 3.23% of all genes. Among these genes, three are involved in flagellum biosynthesis (fliO, flgH-I) and one in arginine AAE biosynthesis (argB). Some of these losses appear to be single, independent events, while others have been lost several times in a convergent manner (e.g. murE lost three times is a component of cell walls, Figure 1).

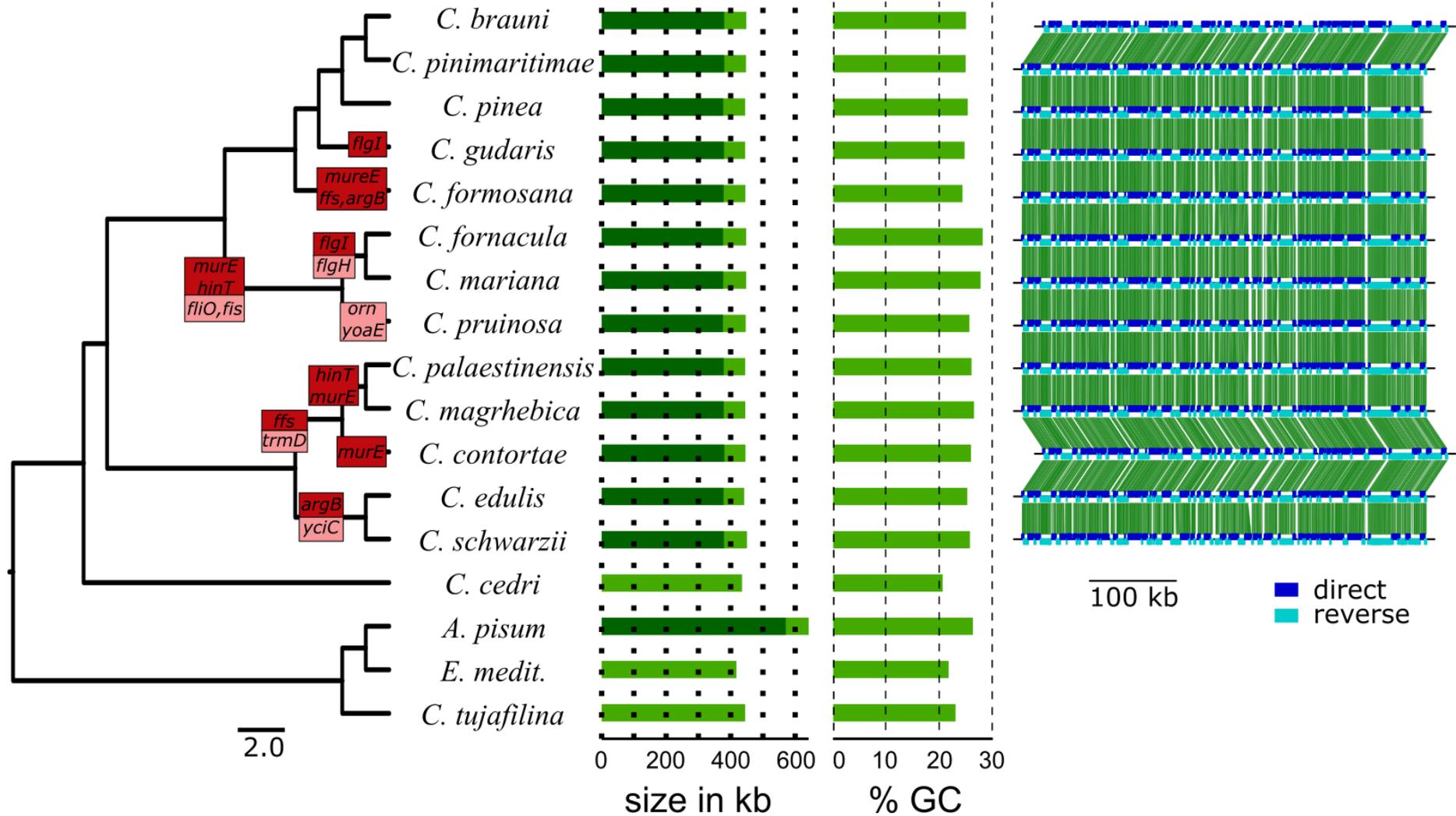


Figure 1: phylogenetic reconstruction for *Buchnera* strains, using the *Buchnera* associated with *C. tujafilina*, *A. pisum* and *E. mediterraneus* as outgroups. The genes lost in the different strains are represented in red. From left to right, the size of each *Buchnera* genomes in kilobase are represented in green. Dark green representing the coding proportion and light green the non coding proportion, the percentage of GC content of each available strain are represented in light green. Genome synteny is represented on the far right.

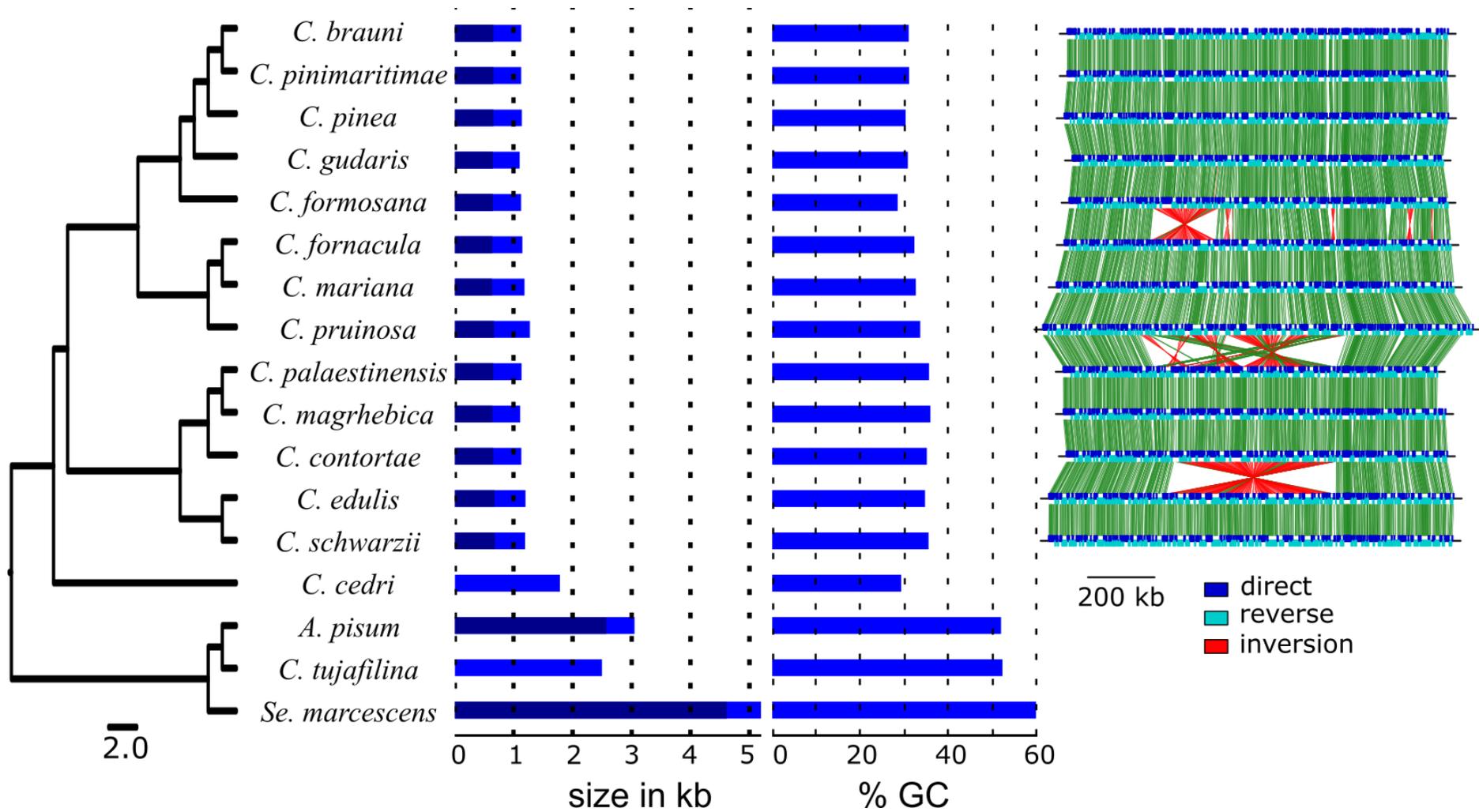


Figure 2: Phylogenetic reconstruction for *Serratia*, using facultative *Serratia* (from *cinara tujafilina* and *Acyrtosyphon pisum*) and *Serratia marcescens* as outgroup. The percentage of GC content of each available strain are represented in light blue. The size of each *Serratia* genomes in megabase are represented in blue. Dark blue representing the coding proportion and light blue the non coding proportion.

For *Serratia* genomes, we found a variation up to 39 genes between species and the presence of inversions shared by several lineages (Figure 2). Those variations were characterized by 145 genes (Figure S1) of *Serratia* whose loss occurred in at least one lineage, representing 21.29% of all genes. These losses are found in all Clusters of Orthologous Genes (COG; Figure 3) with a particularly high proportion of genes implicated in Cog R, S, E, T, C, K P (see Table S5) function lost at least one time. Those results suggest ongoing gene losses in *Serratia*.

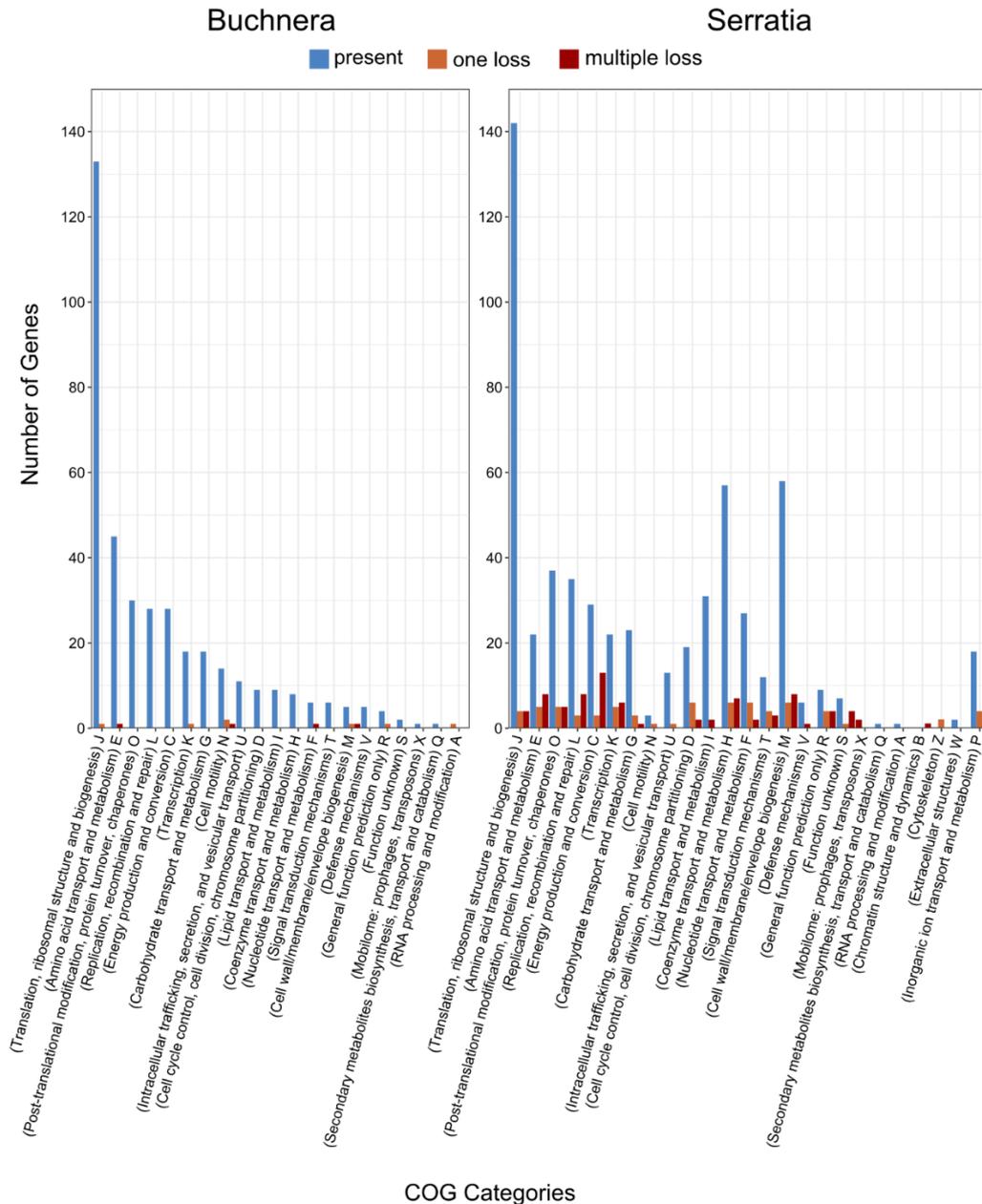


Figure 3: Bar chart showing the amount of genes lost and present across *Buchnera* (right) and *Serratia* (left) genomes from *Cinara* species. Genes are binned into their Clusters of Orthologous Groups (COGs) functional categories (Tatusov et al., 2000). Bars are color coded according to the number of independent losses (i.e., genes lost in no species in blue, genes lost once in orange and genes with multiple independent losses in red).

Endosymbionts nutritional complementation

To infer the dependence of both *Buchnera* and *Serratia* for nutritional complementation of the aphid, we compared the gene repertoire relating to the biosynthesis of EAAs, vitamins, and co-factors in the 13 pairs of genomes from endosymbionts and looked at overlap and complementarity in these nutritional functions (Figure 4) *Serratia* have lost most EAAs' biosynthetic genes, and therefore depend on *Buchnera* for the provision of these nutrients. Similarly to *Serratia symbiotica* co-obligate symbionts of other Lachninae studied so far, *Serratia* strains encode genes that are missing from *Buchnera* to accomplish the biosynthesis of both riboflavin (full pathway) and biotin (bioA, bioD, and bioB genes). In addition, *Serratia* encode for proteins involved in the biosynthesis of thiamine (thiF, thiI, thiH, thiG, thiC, thiD, thiE, and thiL), as *Erwinia* in its associated clade of *Cinara* (Manzano-Marín et al., 2020a).

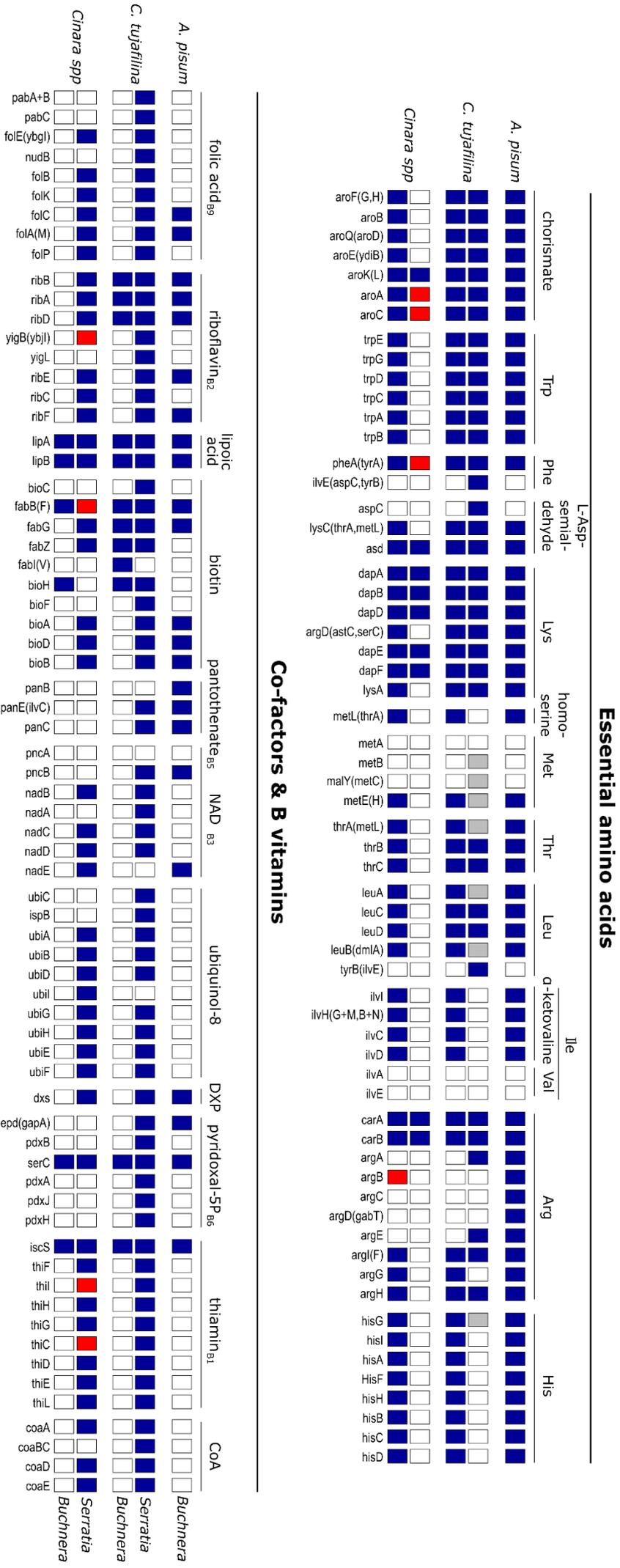


Figure 4. Detailed metabolic complementarity of co-obligate symbiotic systems of aphids. Matrix showing metabolic capacities of obligate symbionts of different aphid species. Blue boxes represent present genes, white boxes represent absent genes, red boxes represent genes lost in at least ones *Cinara* species, grey represent pseudogenes. At the top of each box column, the name of the enzymes or pathway catalysing a reaction. At the bottom, the name of the compound synthesised by the enzymatic steps. At the right of each row of boxes, symbiont taxon name. At the left of each row of boxes, host taxon name.

Phylogenetic history of *Cinara di-symbiotic system*

After filtration by Gblock, we obtain an alignment of 10 583 sites for mitochondrial genes. For *Buchnera* and *Serratia* genomes we obtained alignments of 359451 and 544886 sites respectively. All phylogenetic analyses indicate that *Serratia symbiotica* of the thirteen *Cinara* species used in this study form a well-supported monophyletic group (Figure 2). Moreover, comparison between *Cinara*, *Buchnera* and *Serratia* phylogenies in this monophyletic group showed a perfect congruence of their evolutionary history (Figure 5).

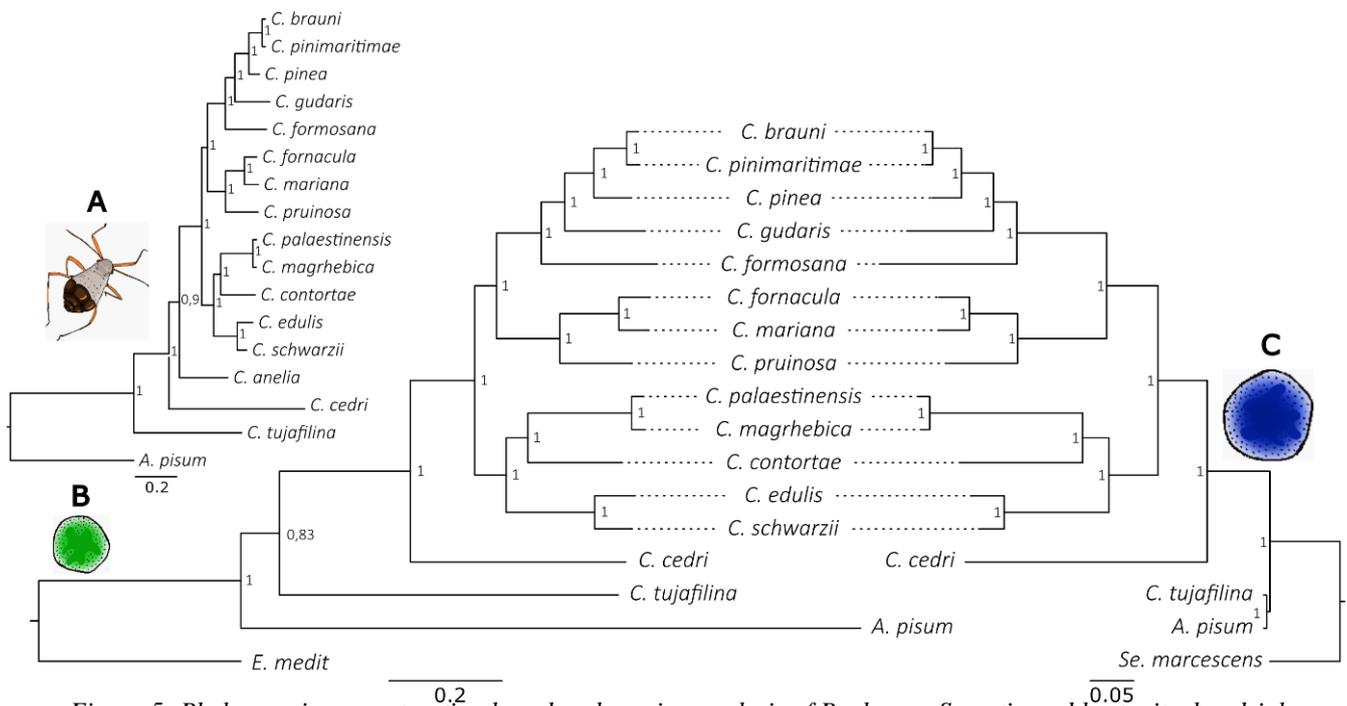


Figure 5: Phylogenetic reconstruction based on Bayesian analysis of *Buchnera*, *Serratia* and host mitochondrial genomes. **A** phylogenetic reconstruction for clade A *Cinara* species, using mitogenomes from *Cinara tujafilina* and *Acyrtosyphon pisum* as an outgroup. **B** phylogeny of the corresponding *Buchnera* strains, using the *Buchnera* associated with *C. tujafilina*, *A. pisum* and *E. mediterraneus* as an outgroup. **C** phylogenetic reconstruction for *Serratia*, using facultative *Serratia* (from *Cinara tujafilina* and *Acyrtosyphon pisum*) and *Serratia marcescens* as an outgroup. Posterior probabilities are written for each node.

Endosymbionts genome evolution

Beast analyses carried out made it possible to measure evolutionary rates for the whole clade for each partner. Thus, respective substitution rates of $10.5 \cdot 10^{-3}$ substitution per site per million years (sbst/ste/My) and $9.25 \cdot 10^{-3}$ sbst/ste/My for *Buchnera* and *Serratia*. For the mitochondrial genomes, a rate of $4.66 \cdot 10^{-9}$ sbst/ste/My (95% HPD interval [$3.95 \cdot 10^{-9}$, $5.44 \cdot 10^{-9}$]) was obtained.

Rate correlation

Comparison of parallel branches in host and symbiont phylogenies (schema Figure S6) revealed a significant correlation between their rates of substitutions by site (host-*Buchnera* : $R = 0.83$, $p = 7.7e-10$; host-*Serratia* : $R = 0.63$, $p = 3.3e-6$; Figure 6). Correlation were also found between branch lengths inferred by ML (in substitution by site) analysis using *Serratia* and *Buchnera* phylogenies ($R = 0.87$, $p = 1.5e-11$; Figure 6). We also found a correlation between lengths of parallel branches in *Serratia* and *Buchnera* phylogenies (Figure S3), when using Bayesian analysis.

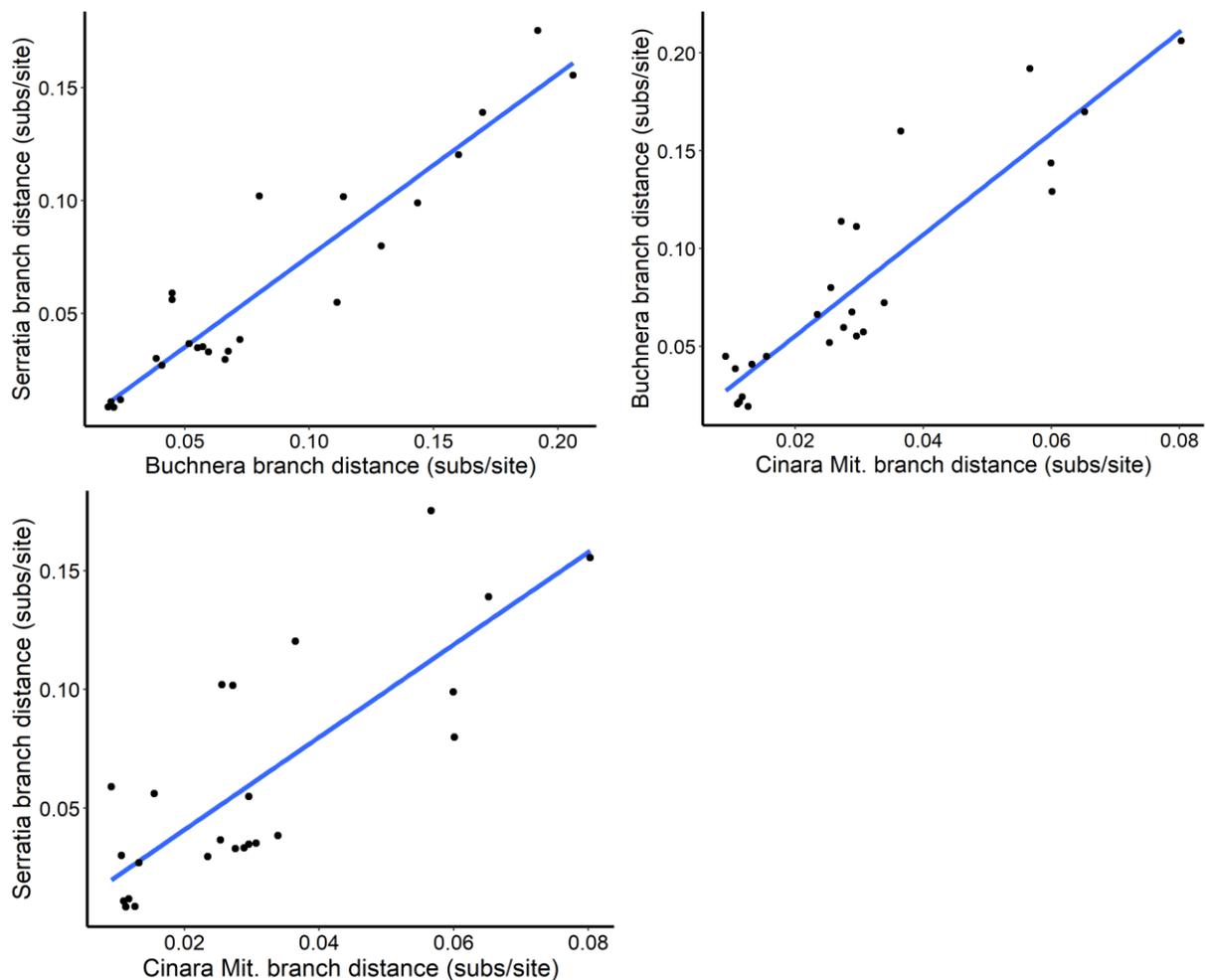


Figure 6: Comparison of evolutionary rates of *Serratia* and *Buchnera* symbionts and their host *Cinara*. Top left correlation of branch distances in phylogenies of *Serratia* and *Buchnera*, inferred using maximum-likelihood analysis of CDS from each data set. Top right correlation of branch distance in phylogenies of *Cinara* mitochondrial genome and *Buchnera*, inferred using maximum-likelihood analysis of *Buchnera* CDS and mitochondrial genes data set. Bottom correlation of branch distance in phylogenies of *Cinara* mitochondrial genome and *Serratia* inferred using maximum-likelihood analysis of *Serratia* CDS and mitochondrial genes data set.

Signature of selection across endosymbiont genomes

Using M0 model in Codeml we found mostly genes undergoing strong purifying selection in *Buchnera* (avg. $\omega = 0.0625$ [range = 0.0053-0.9999, n = 289]), few genes undergoing relaxed purifying selection (avg. $\omega = 0.1328$ [range = 0.1003-0.2556, n = 86]) and no genes were found under positive selection. In *Serratia*, most genes are undergoing strong purifying selection (avg. $\omega = 0.0682$ [range = 0.0001-0.0999, n = 142]), while comparatively few genes are experiencing relaxed purifying selection (avg. $\omega = 0.1590$ [range = 0.1000-0.4025, n = 432]). Additionally, no genes under positive selection were found.

Concerning positive selection tests, all the results were subjected to a correction for multiple testing. For *Buchnera* genes, we found 1 to 51 genes under positive selection ($p < 0.01$) via respectively the use of the M1a-M2a and M7-M8 codeml Nested model respectively. As for the results obtained using Busted, 10 genes under positive selection ($p < 0.01$) were obtained. Among these genes, no genes have strong evidence of positive selection supported by all three tests. Concerning *Serratia* genes, we found 3 to 48 genes under positive selection ($p < 0.01$) via the use of the M1a-M2a and M7-M8 codeml Nested model respectively. As for the results obtained using Busted, 9 genes under positive selection ($p < 0.01$) were obtained. Among these genes, no genes have strong evidence of positive selection supported by all three tests.

Finally, the search for relaxed selection using relax (HYPHY) on genes shared by both symbionts resulted in 92 genes under relaxed selection found in *Buchnera* using *Serratia* genes as reference. In addition, 95 genes under relaxed selection were found in *Serratia* using *Buchnera* genes as reference.

Intraspecific variations in symbiont genomes

Using intraspecific genome dataset we reconstructed phylogenetic tree of both symbionts and found a perfect congruence between *Buchnera* and *Serratia* phylogeny (Figure S7). We then analysed polymorphisms and indels at the intraspecific level (Table S6: *C. mariana* and S7: *C. fornacula*). In the two species, we found strikingly similar patterns of mutations between co-existing symbionts: in the two species, co-existing *Buchnera* and *Serratia* exhibit similar number of SNP/kb in coding and non-coding sequences and in both cases non coding regions exhibiting higher (almost twice as high) number of substitutions. Number of indels are 50 to hundred times higher in non-coding regions than coding regions. In *Serratia* in one of the species higher number of indels/ kb than *Buchnera* were found and may be a sign of ongoing process of reduction in *Serratia*. In both symbiont genomes indels are localized mostly in

intergenic region (more than 90%) which constitute only 17.6% of *Buchnera* genomes and 47% of *Serratia* genomes.

Localisation of Serratia symbiotica in Cinara species

FISH investigations revealed that *Serratia* of the three species examined were exclusively present inside bacteriocytes independent from *Buchnera* bacteriocytes, but distributed in the same body area (Figure 7) *Serratia* also present coccoid cell shape smaller than *Buchnera* cells as previously described in endosymbionts from other *Cinara* species with highly reduced symbiont genome (Manzano-Marín et al., 2017; Manzano-Marín et al., 2020a).

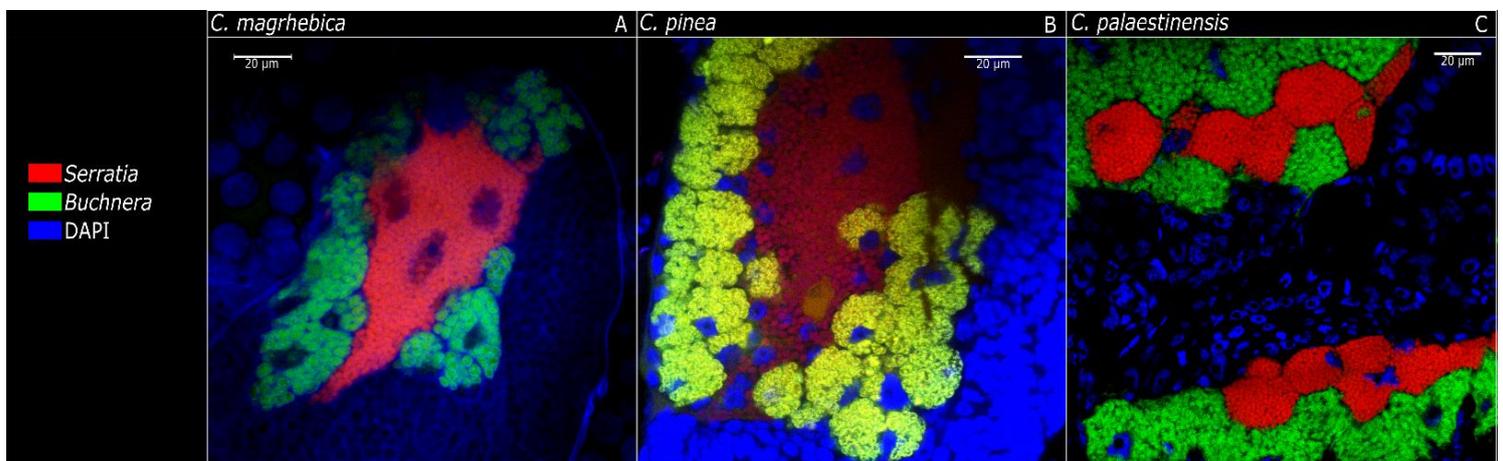


Figure 7: Merged FISH microscopic images of aphid embryos from selected *Cinara* aphids. *Serratia* signal is shown in red, *Buchnera*'s in green, and DAPI's (staining DNA, highlighting host nuclei) in blue. The scientific name for each species are shown at the top of each panel and colour code for each fluorescent probe and its target group are shown on the left part.

IV. Discussion

Previous studies have shown that *Serratia symbiotica* is found as a co-obligate symbiont of *Buchnera* in several subfamilies of Aphididae (Manzano-marín, 2022; Monnin et al., 2020). It is more specifically associated with many species of *Cinara* (Lachninae) (Meseguer et al., 2017). However little is known about its history in this clade of aphids, and though comparisons of different strains of *Serratia* have shown that it undergoes fast genome reduction once it becomes obligate (Manzano-Marín and Latorre, 2016), its pace of genome evolution once associated with aphids over long evolutionary time scales has not been estimated. In addition, the reciprocal effect of the co-symbiont presence on *Buchnera* genome evolution has seldom been investigated. Addressing these issues necessitates a comparative analysis of symbionts genomes in a clade where *Buchnera* and *Serratia* share a long history of coevolution.

Serratia symbiotica has co-speciated with a clade of *Cinara*

Here, we first confirmed the presence of a *Serratia symbiotica* in all *Cinara* species of clade A (Meseguer et al., 2017), showing that each of them have reduced genome and low G-C levels reflecting a long-term obligate association with their aphid hosts (Moran et al., 2008b). Phylogenomic analyses using aphid mitochondrial genomes, *Buchnera* and *Serratia* whole genomes reveal that all three partners have parallel phylogenetic histories. At the intraspecific level, in two species, phylogenetic analyses corroborates the pattern of co-divergence of *Buchnera* and *Serratia* observed at the clade level. Together, this supports the hypothesis of an ancestral acquisition of *Serratia* followed by co-speciation with *Buchnera* and associated aphids for about 25 My according to *Cinara*'s datation analyses (Meseguer et al., 2015). It suggests vertical transmission of *Serratia* at broad and fine evolutionary scale in this clade of *Cinara*. These inferences are corroborated by the FISH images obtained from three species from this clade (*C. pinea*, *C. magrhebica* and *C. palaestinensis*) that show the existence of identical cell structures for *Serratia*. These images show the presence of two types of bacteriocytes close to each other, one containing *Buchnera* cells and another containing *Serratia* cells. This long co-evolutionary history contrasts with the instability of the di-symbiotic systems observed in several clades of Hemipterans where co-symbiont replacements have been observed at large evolutionary time-scales (Dial et al., 2022; Husnik and McCutcheon, 2016; Koga et al., 2013; Rosenblueth et al., 2012; Szabó et al., 2017; Elena R. Toenshoff et al., 2012; Urban and Cryan, 2012).

Few cases of long-term co-speciation in multi-symbiotic system have been uncovered. There is nonetheless examples of co-speciation patterns between planthoppers and their two ancestral symbionts *Sulcia muelleri* and *Nasuia deltocephalinicola* since their acquisition 300 My ago (Michalik et al., 2021; Takiya et al., 2006; Vasquez and Bennett, 2022). Our results mirrors the results of Manzano-Marin et al. (2020) that showed that long-term co-speciation of *Erwinia* and *Buchnera* occurred in another clade of *Cinara*. Altogether, this indicates that durable symbiosis between *Buchnera* and a new symbiont evolved multiple times in *Cinara* species. In both these study systems, the co-symbiont is found in specialised bacteriocytes, very similar in structure and localisation to *Buchnera*'s bacteriocytes, suggesting that the co-symbiont probably follows the same pathways of vertical transmission than the primary symbiont. In addition, to the ability to hitch-hike on the transmission route of the primary symbiont, the relative stability of the association in comparison to other study systems might also stem from the functions carried by the symbiont. Similarly to other cases of di-symbiotic

associations observed now in aphids, we show that the *Serratia* strains sequenced here are able to complement *Buchnera* for the losses of essential genes, here genes involved in the biosynthesis of B vitamin (Manzano-marín, 2022; Monnin et al., 2020) that have been lost in the common ancestor of all Lachninae (Manzano-Marín et al., 2017). It has been hypothesized that any symbiont that can fulfil these functions could be “adopted” as a co-obligate partner of *Buchnera*. The ubiquity of these functions in aphid facultative symbiont genomes might explain instability and successive symbiont replacement (Manzano-Marín et al., 2020a; Meseguer et al., 2017). However, in both *Erwinia* associated *Cinara* and the *Cinara* clade studied here, the co-symbiont brings a metabolic function that is seldom observed in *Buchnera*: the thiamine synthesis capability. In Manzano-Marín *et al.* (2020), it was suggested that the acquisition and maintenance of *Erwinia* in their hosts might be associated with an advantage brought by this new function, it might favour the development of aphids on diet with low levels of thiamine, i.e. here conifer sap, mainly *Abies*, *Larix* and *Picea*. The only *Buchnera* that have retained this function are indeed the ones associated with two aphid genera (*Neophyllaphis* and *Mindarus*) that also feed on these conifer genera (Manzano-marín, 2022). The aphid species studied here share the same range of host plants and also include *Pinus* species. There is no data comparative data on the thiamine content of conifer sap vs angiosperm sap. Testing the role of the thiamine biosynthetic capability in the stability of this di-symbiotic system will necessitate a thorough comparison of the composition of ingested sap of aphids hosting symbionts with and without this capacity. Alternatively the relative stability of the system might also be due to relative rarity of “alternative” symbionts in the clade studied. If symbiont replacements in multi-partner symbioses throughout evolution occurs through successive replacements of strains competing for the same niche (as suggested by Mesguier et al, 2017 in the *Cinara* genus, but also by *Amblyomma* ticks (Binetruy et al., 2020)), a lack of “competitors” might explain why some associations are more stable

Serratia and *Buchnera* evolve at the same pace

Using previous datation analyses on *Cinara*, an average substitution rate for *Buchnera* aphidicola of 10.5×10^{-9} substitution/sites/years (s/s/y) (95% HPD interval [9.19×10^{-9} , 11.9×10^{-9}]) was inferred. The rate of evolution of *Buchnera* genomes exceed that of the mitochondria (average = 4.66×10^{-9} s/s/y; 95% HPD interval [3.95×10^{-9} , 5.44×10^{-9}]). This rate is very similar to *Buchnera* evolutionary rates inferred from, previous studies (inference of substitution rates vary between 4.3×10^{-9} to 8×10^{-9} s/s/y (Clark et al., 1999; Gómez-Valero et al., 2007; Manzano-Marín et al., 2016; Ochman et al., 1999; Pérez-Brocal et al., 2011; Tamas

et al., 2002; van Ham et al., 2003) (table S8). Faster substitution rate for *Buchnera* has been estimated by Moran et al., (2009), but this has been conducted at a different evolutionary scale (within one species). Finally, our estimation shows that *Buchnera* evolved twice as fast as free-living bacteria as *E. coli* which have an evolutionary rate of approximately 0.45×10^{-9} s/s/y (Ochman et al., 1999).

For the first time, due to the co-speciation pattern revealed here, we are able to estimate the evolutionary rates observed in *Serratia symbiotica*. Our analyses also show that *Serratia* evolves at a very similar rates (range = 9.25×10^{-9} s/s/y; 95% HPD interval [7.98×10^{-9} , 10.5×10^{-9}]) and we observe a strong correlation of branch lengths between *Serratia* and *Buchnera* trees. The patterns of intraspecific variation in two species confirms the two symbionts evolve at a very similar pace. Significant correlation with those of the mitochondria and *Buchnera* were found, making this the first di-symbiotic system where correlated evolution of the evolutionary rates of three partners is observed despite symbionts being sequestered in separate bacteriocytes. These results contrast with those observed for the di-symbiotic system of Nesophrosyne leafhoppers that host *Nasuia* and *Sulcia* as obligate symbionts. Indeed, in this group, the evolutionary rates of the two bacteria are relatively different from each other, with a rate of 1.13×10^{-7} s/s/y for *Nasuia*, corresponding to 10.7 and 12.2 times those of *Buchnera* and *Serratia* respectively (Vasquez and Bennett, 2022). As for *Sulcia*, its evolutionary rate is the lowest of all bacterial symbiotic systems observed with a rate of 3.24×10^{-9} s/s/y, which is 0.3 and 0.35 times those of *Buchnera* and *Serratia* respectively (Vasquez and Bennett, 2022).

In addition to the correlation observed between symbiont evolutionary rates, we also reveal a significant correlation between mitochondrial and symbionts evolutionary rates. Our results corroborate those found previously in the *Buchnera*-aphids mono-symbiotic system (Arab and Lo, 2020) but also with those found in other mono-symbiotic systems (*Blochmannia*-Carpenter Ants in Degnan et al., (2004) ; Blattabacterium-Cockroaches in Arab et al., (2020)). Nevertheless, these results are in opposition to those previously found in other di-symbiotic systems of hemipterans, where decoupled substitution rates was found between the three partners (*Sulcia*, *Nasuia*, mitochondria) (Vasquez and Bennett, 2022).

Evolutionary rates of symbiotic bacteria are known to be influenced by generation times (Degnan et al., 2005; Silva and Santos-Garcia, 2015), so that bacteria with shorter and more frequent generations may undergo more mutations and substitutions per unit time. Given our results, we can hypothesise that *Buchnera* and *Serratia* have similar generation times in this

system. In addition to generation times very, these results also suggest that similar selective force are acting on the genomes. This might stem from shared dependency on host genes (e.g. host DNA repair genes taking over from those lost in the symbiont genes; (Mao et al., 2018; McCutcheon et al., 2019; McCutcheon and Moran, 2012)) or shared metabolic pathways (e.g. (Lamelas et al., 2011a; Manzano-marín, 2022; Manzano-Marín et al., 2018, 2016, 2015; Manzano-Marín and Latorre, 2014). Finally, it also strongly suggest that the two symbionts undergo very similar demographic events, such as bottlenecks of similar sizes during vertical transmission. Quantitative measure of symbiont density during aphid development and symbiont transmission events will help testing these assertions

Patterns of gene losses, signature of selection across genomes in aphid di-symbiotic systems

The genomes of vertically transmitted endosymbionts are characterised by numerous gene losses (McCutcheon and Moran, 2012; Moran et al., 2008b; Toft and Andersson, 2010). This process appears to occur rapidly after the acquisition of an obligate life-style (Boscaro et al., 2017; McCutcheon et al., 2019). In line with previous studies (Manzano-Marín et al., 2016; Manzano-Marín et al., 2020a; Pérez-Brocal et al., 2006), we have shown here a high degree of gene conservation and relatively stable genomes between different *Buchnera* lineages. In the *Serratia* sequenced here despite some ancient inversion, we notice a strong conservation of genes (545 genes shared by all lineages) and highly reduced genomes (1.154 Mpb on average) in comparison with its free-living relatives (5.1-5.4 Mpb) or even facultative *Serratia symbiotica* (2.4 Mb). Hence, our analyses support an evolutionary history with rapid genome reduction after the transition from a free-living to obligate symbiont of *Serratia* in this clade. This process mirrors what is assumed for *Buchnera* and other obligate endosymbionts (Chong et al., 2019; Moran et al., 2009, 2008b, p. 200; Patiño-Navarrete et al., 2013; Vasquez and Bennett, 2022; Wernegreen, 2015) and suggests that the evolutionary factors leading to genome shrinkage of vertically transmitted obligate symbionts are common to all bacterial lineages.

The description of a large number of *Buchnera* genomes has highlighted their small genome sizes and the low number of coding genes (354-587 protein-coding genes) (Chong et al., 2019). This is particularly true of the *Buchnera* associated with a co-obligate symbiont (Manzano-Marín et al., 2016; Manzano-Marín et al., 2020a; Pérez-Brocal et al., 2006). It has sometimes been hypothesized that the presence of the co-symbiont might accelerate the process of genome erosion in *Buchnera*, as it makes it redundant for several functions (Lamelas et al., 2011a). On the other hand, the newly acquired symbiont also loses genes that are no longer

necessary for its life inside the host but also genes that encode metabolic functions that are fulfilled by the primary symbionts. This complementarity suggests an ongoing co-evolutionary path that may be similar to the black-queen hypothesis (Morris et al., 2012). We confirm here the complementarity in metabolic functions of the two symbionts. We show that *Serratia* has lost most amino acid biosynthetic genes that are brought by *Buchnera*. Mapping ongoing gene losses for both symbionts on their phylogenies reveal only few losses in metabolic functions (Figures 1 and 2). Therefore, it seems that the process of genome losses is very rapid upon acquisition as a co-obligate symbiont.

The loss of genes involved in cell wall, DNA repair or DNA recombination are functions that are often lost by obligate symbionts in hemipterans but also in other species. These losses are thought to be due to the particular and stable lifestyle of the symbionts which would make the presence of these genes unnecessary and high genetic drift (McCutcheon and Moran, 2007; Moran et al., 2009, 2008b; Tamas et al., 2002; Wernegreen, 2015). We further investigated whether the process of genome erosion could be detected through signatures of selections across symbiont genomes. Our positive selection tests on the genes of each symbiont did not find any genes for which there is strong evidence of positive selection. This result is congruent with the dN/dS measurements, which show that all genes of both symbionts are subject to strong purifying selection. In contrast to our results, previous studies have shown the presence of genes under positive selection in hemipteran symbiotic genomes. For example, in the aphid those *Buchnera* mono-symbiotic system, 29 genes were found to be under positive selection (Chong et al 2019) and 97 and 4 genes in *Sulcia* and *Nasuia* respectively in the leafhoppers di-symbiotic system (Vasquez and Bennett, 2022). However, these studies were conducted at different time scales, among hosts that had diversified in different ecological niches. Among the genes found under positive selection in previous study (Chong et al., 2019; Vasquez and Bennett, 2022), all were under purifying selection or lost in our symbionts genomes. Despite many lost genes, it seems that selection on remaining genes takes place to maintain their function. In this context, gene losses may be more a random process and occurs via rapid accumulation of mutations with high rates of molecular evolution and fixation via high genetic drift (Wernegreen, 2015, 2002). Indeed, genome reduction correlated to increase in mutation rate has been documented for several obligate symbionts (Bourguignon et al., 2020; Kinjo et al., 2021; Vasquez and Bennett, 2022). In this context of overall “drifting system”, it might not be easy to detect signature of selection and we observe that genes are either there or not.

To complement our analyses, we tested whether some genes were under relaxed selection for one of the two symbionts. This analysis recovered very few genes, they did not specifically belong to categories involved in host metabolism (as you would expect under a scenario where *Buchnera* loses functions under relaxed selection caused by redundancy with *Serratia*), some of these genes were sometimes identified as undergoing positive selection in other tests. In such systems, it might be difficult to differentiate relaxed from positive selection. To find the genes under relaxed selection, future analyses might also need free-living bacterial genomes as background for the analysis.

Our intraspecific analyses of two *Cinara* species describe low snps/kb and indels/kilobase in the coding regions of the *Buchnera* and *Serratia* genomes suggesting that purifying selection is acting also at this scale. On the other hand, we observed the presence of a higher rate of indels in the non-coding sequences, especially in *Serratia* genomes. Observation of high levels of indels may indicate that genome reduction is undergoing. Indeed, the process of endosymbiotic genome reduction previously described involve pseudogenisation (by fixation of deleterious mutation) followed deletion in non-coding region (due to deletional bias) (Andersson and Andersson, 1999; Mira et al., 2001; Moran and Mira, 2001; van Ham et al., 2003). Overall, the patterns observed for both obligate symbionts are very similar to the only intraspecific investigation conducted on *Buchnera* genomes on the pea aphid (Moran et al., 2009). The presence of a greatly reduced *Buchnera* has been demonstrated in other sub-families such as the Phyllaphidinae and Calaphidinae, the latter having the smallest *Buchnera* genomes sequenced to date without evidence of association with a co-symbiont (Manzano-marín, 2022). Thus, the data so far suggests that this extreme reduction observed in *Buchnera* may not be due to the presence of co-symbiont but to higher rates of evolution that would increase the likelihood that a random mutation would turn off a gene and then be fixed by the high genetic drift that occurs (Wernegreen, 2015).

Conclusions

Our study here shows that a multi-partner symbiosis can also be stable over long periods with parallel co-speciation of partners, as previously described in other systems (Manzano-Marín et al., 2020a; Vasquez and Bennett, 2022). We also describe the first di-symbiotic model for which both symbionts evolve at the same pace, suggesting relatively similar symbiotic processes. We also support the hypothesis of a convergence of symbiotic genomes towards a stable reduced form where only essential genes remain. Process of genome erosion seems to

happen rapidly in the early stage after co-symbiont acquisition, therefore it is difficult to characterise even at the gene scale. Finally, this work supports the important role of these new symbionts in complementing the primary symbiont *Buchnera* comfort the idea that mechanisms by which free-living bacteria establish themselves as obligate partners in multi-partner endosymbioses the same in *Cinara* species are very similar across lineages.

Chapitre 3

Dynamique des endosymbiotes nutritionnel et régulation de l'hôte.

(projet d'article)

Contexte de l'étude

Nous avons pu observer dans les chapitres précédents l'existence chez les *Cinara* d'une symbiose à deux partenaires obligatoires relativement dynamique, avec de multiples remplacements de co-symbiotes. Il découle de ces remplacements successifs, la présence de co-symbiote dont l'âge d'acquisition est ancien et par opposition d'autres pour lesquelles la transition vers un mode de vie obligatoire est récente. Nous avons également mis en évidence l'existence de localisations différentes des bactéries co-symbiotiques ainsi que des génomes de tailles et taux de GC variables selon l'espèce étudiée. Malgré cette grande diversité au sein des co-symbiotes, les capacités ainsi que les modes de régulation des symbiotes dans ces systèmes multi-symbiotiques restent totalement inconnus.

Nous avons donc voulu dans cette étude effectuer des premières analyses afin de déterminer s'il existe une régulation des symbiotes dans les systèmes di-symbiotiques des *Cinara*, et si celle-ci varie entre espèces. Pour ce faire, deux espèces ont été choisies pour les caractéristiques différentes de leur co-symbiote (en termes de taille de génome, taux de GC, âge d'acquisition et localisation). Celles-ci ont été échantillonnées et élevées en laboratoire pour nous permettre de mesurer la quantité de symbiote au cours de leur vie.

Pour cela, les objectifs de cette partie sont donc de : **1)** Quantifier à différents stades de vie les symbiotes (*Buchnera* et *Serratia*) de deux espèces de *Cinara*. **2)** Comparer la dynamique de population bactérienne de *Buchnera* et de *Serratia* chez chacune des deux espèces. **3)** Comparer les dynamiques bactériennes des deux espèces de *Cinara*.

Les hypothèses de cette étude sont que les co-symbiotes *Serratia* de nos deux espèces (*C. palaestinensis* et *C. tujafilina*) ont des dynamiques différentes du fait de leur acquisition indépendante et de leurs caractéristiques différentes. On suppose que chez *C. tujafilina* pour qui *Serratia* est issu d'une acquisition récente et dont la localisation est extracellulaire, la régulation du symbiote est moins efficace. Par opposition, on s'attend à une dynamique similaire à celle de *Buchnera* dans l'espèce *C. palaestinensis* pour laquelle *Serratia* est intracellulaire et a été acquise il y a plus de 15 Ma.

Abstract

Many insects carry intracellular bacterial symbionts located in bacteriocytes, on which they depend for their nutrition in essential amino acids and vitamins. An increasing number of systems in which multiple obligate symbionts cohabit are described, yet little attention has been paid to the coordination of these systems. However, it is to be expected that they may compete or cooperate for space, nutrients and transmission. In order to determine the nature of these interactions, it is important to know more about how they interact, their location, and the abundance of each symbiont species. Aphids of the genus *Cinara* (Lachninae) typically harbour di-symbiotic systems in which metabolic capabilities of the former obligate symbiont *Buchnera aphidicola* are complemented by those of a more recently acquired nutritional symbiont, often belonging to the species *Serratia symbiotica*. Here, we used real-time quantitative PCR approaches to characterise the abundance and dynamics of symbionts at different developmental stages of the di-symbiotic system formed by *Buchnera aphidicola* and *Serratia symbiotica*. For this purpose, we used two *Cinara* species (*C. palaestinensis* and *C. tujaefilina*) for which *Serratia* was derived from two independent acquisitions and for which the localisation of the bacteria was verified by fluorescence in situ hybridisation. Our results showed that the density of *Buchnera* was lower than that of *Serratia* at all host developmental stages analysed. Indeed, the endosymbiont population increases exponentially throughout nymphal development and then gradually decreases as the insect ages. Finally, our observations show that, the localisation of *Serratia* differs between the two species with an intracellular localisation in *C. palaestinensis* and an extracellular localisation in *C. tujaefilina*. In summary, these results show a coordination of cell dynamics between the two obligate symbionts which might reveal a regulation of these according to the nutritional needs of the host during development. They further suggest that symbionts that are extracellular cannot be as finely regulated

Résumé

De nombreux insectes sont porteurs de symbiotes bactériens intracellulaires localisés dans des bactériocytes et dont ils dépendent pour leur nutrition en acides aminés essentiels et vitamines. De plus en plus de systèmes pour lesquels il y a cohabitation de plusieurs symbiotes obligatoires sont décrits et pourtant, peu d'attention a été portée sur la coordination de ces systèmes. Cependant, on peut s'attendre à ce que ceux-ci puissent entrer en compétition ou coopérer pour l'espace, les nutriments et la transmission. Afin de déterminer la nature de ces interactions, il est important d'en savoir plus sur la façon dont ils interagissent, leur localisation, et l'abondance de chaque espèce de symbiote. Les pucerons du genre *Cinara* (Lachninae) abritent typiquement des systèmes di-symbiotiques dans lesquels les capacités métaboliques de l'ancien symbiote obligatoire *Buchnera aphidicola* sont complétées par celles d'un symbiote nutritionnel plus récemment acquis, appartenant souvent à l'espèce *Serratia symbiotica*. Ici, nous avons utilisé des approches de PCR quantitative en temps réel pour caractériser l'abondance et la dynamique des symbiotes à différents stades de développement du système di-symbiotique formé par *B. aphidicola* et *S. symbiotica*. Pour cela nous avons utilisé deux espèces de *Cinara* (*C. palaestinensis* et *C. tujaefilina*) pour lesquelles *Serratia* est issue de deux acquisitions indépendantes d'âges différents et pour lesquelles la localisation des bactéries a été vérifiée par hybridation in situ en fluorescence. Nos résultats ont montré que la densité de *Buchnera* était plus faible que celle de *Serratia* à tous les stades de développement de l'hôte analysés, même si elles affichaient toutes deux une tendance similaire. En effet, la population d'endosymbiotes augmente de façon exponentielle tout au long du développement nymphal, puis diminue progressivement au cours du vieillissement de l'insecte pour les deux co-symbiotes. Finalement, nos observations montrent que la localisation de *Serratia* diffère entre les deux espèces avec une localisation intracellulaire chez *C. palaestinensis* et extracellulaire chez *C. tujaefilina*. En résumé, ces résultats montrent une coordination de la dynamique cellulaire entre les deux symbiotes obligatoires et révèlent une régulation de ceux-ci selon les besoins nutritionnels de l'hôte au cours du développement.

I. Introduction

Les interactions mutuellement bénéfiques entre les bactéries et les insectes sont omniprésentes dans la nature. Parmi celles-ci on retrouve des symbioses spécifiques et durables dans lesquelles l'association est nécessaire à la survie et à la reproduction de l'hôte et où la bactérie ne peut généralement pas vivre en dehors de l'hôte. Ces symbiotes décrits comme "obligatoires" sont pour la plupart localisés à l'intérieur de cellules hôtes spécialisées appelées bactériocytes (Braendle et al., 2003; Michalik, 2010) et leur acquisition se fait par une transmission verticale du parent aux descendants, le plus souvent par transmission maternelle (Bright and Bulgheresi, 2010) assurant ainsi le maintien de l'association. Ces symbiotes apportent des bénéfices écologiques à leur hôte, tels que l'apport de nutriments essentiels trouvés en quantité insuffisante dans le régime alimentaire des insectes. L'établissement de ces associations a permis à de nombreux insectes de s'adapter à des régimes nutritionnellement limités, comme la sève des plantes dont s'alimentent les pucerons (Douglas, 1998a) ou le sang dont se nourrissent les tiques et mouches Tsétsé (Wang et al., 2013). Mais ces associations ne sont pas sans coûts pour les hôtes, il existe des coûts notamment associés à la fourniture d'énergie et de nutriments pour maintenir la population de symbiotes hébergés par les insectes (par exemple (Engl et al., 2020; Jamin and Vorburger, 2019)). Ainsi, le bénéfice net d'une association symbiotique pour un hôte peut être défini comme la différence entre les bénéfices et les coûts de la présence du symbiote.

La densité de la population de symbiote a un rôle important dans cette balance des bénéfices et des coûts pour l'hôte. En effet plus la population de symbiotes sera importante plus son « hébergement » sera consommatrice d'énergie pour l'hôte. La population doit être cependant suffisamment abondante pour assurer la production des services mutualistes (production de nutriments ou de système de défense par exemple). Il a été notamment montré chez une drosophile que la protection antivirale apportée par la présence de *Wolbachia* dépend de sa densité dans l'hôte (Lu et al., 2012; Osborne et al., 2012). Afin de maintenir cette balance coût-bénéfice à leur avantage, les hôtes doivent donc réguler la densité de leurs symbiotes (López-Madrugal and Duarte, 2019; Ratzka et al., 2012; Whittle et al., 2021; Wilkinson and Douglas, 1998). Il a été montré que la densité des symbiotes est ainsi influencée par plusieurs facteurs tels que la température (Chen et al., 2009), l'âge de l'hôte (Baumann and Baumann, 1994), le sexe de l'hôte (Rio et al., 2006), le cycle de reproduction (Wolschin et al., 2004). Dans la plupart de ces études, ces variations sont interprétées comme étant le résultat de la régulation par l'hôte en fonction de leurs besoins physiologiques, qui peuvent dépendre de leur stade de

développement, leur sexe et de facteurs environnementaux. Un des exemples emblématiques de cette capacité de régulation des hôtes, est le cas du charançon du riz *Sitophilus oryza*. Chez cet insecte, l'endosymbiote *Sodalis pierantinnus* participe à la production de composants essentiels à la formation de la cuticule. Il a été montré qu'il proliférait après la métamorphose de son hôte et était éliminé par l'hôte lorsque sa cuticule était définitivement formée (Vigneron et al., 2014). Cette dynamique semble parfaitement adaptée aux besoins de l'hôte.

Dans les systèmes symbiotiques dans lesquels il y a cohabitation de plusieurs symbiotes obligatoires, le contrôle de la densité des symbiotes devient complexe. L'hôte doit en effet réguler chacun de ses partenaires dans des conditions où il existe probablement une compétition entre symbiotes pour l'espace et les ressources (Goto et al., 2006; Perreau et al., 2021). Il a cependant été montré que les insectes pouvaient contrôler leur densité de symbiotes dans des systèmes pluri-symbiotiques. C'est le cas chez l'aleurode, *Diaphorina citri* chez lequel on retrouve deux symbiotes obligatoires *Carsonella ruddii* et *Proffella armatura* dont la dynamique au cours de la vie de l'hôte est parallèle malgré une plus grande proportion de *Proffella* (Dossi et al., 2014). Par ailleurs, il semble que l'hôte puisse exercer une régulation discriminante entre ses différents symbiotes selon ses besoins. Par exemple, chez la mouche *G. morsitans* on observe des dynamiques symbiotiques différentes chez deux symbiotes obligatoires *Sodalis* et *Wigglesworthia*, et des variations entre les mâles et les femelles de cette espèce pour lesquels les besoins nutritionnels diffèrent (Rio et al., 2006). Cependant, dans la plupart des cas, les processus impliqués dans le contrôle des symbiotes sont encore inconnus.

Chez les pucerons l'association symbiotique obligatoire avec *Buchnera aphidicola* comble les déficits en AA essentiels et en vitamines manquant à leur alimentation (Buchner, 1953a; Douglas, 1998a; Gosalbes et al., 2008b). Ces bactéries sont localisées dans des bactériocytes dont la dynamique intra-hôte, a été décrite chez l'espèce *Acyrtosiphon pisum*. La dynamique de *Buchnera* se caractérise par une forte augmentation de sa densité de la naissance du puceron jusqu'au dernier stade larvaire puis par une diminution lente à l'âge adulte qui s'accélère après la dernière reproduction (Lu et al., 2014; Nishikori et al., 2009; Simonet et al., 2016a). Cette dynamique semble corrélée aux besoins nutritionnels du puceron lors de son développement (Simonet et al., 2016a). Les processus régulant cette dynamique symbiotique ont eux aussi été décrits par (Simonet et al., 2018a) chez *A. pisum*, chez qui il a été observé une dégénérescence des bactériocytes suivie de leur mort cellulaire.

La découverte de l'existence de plusieurs bactéries obligatoires acquises indépendamment chez diverses espèces de pucerons appartenant au genre *Cinara* (Manzano-

Marín et al., 2017; A. S. Meseguer et al., 2017), interroge sur la capacité de régulation de la densité de symbiotes dans ce système. De plus, comme décrit par Meseguer et al., (2017) et Manzano-Marín et al. 2017 les *Cinara* ne possèdent pas toutes les mêmes co-symbiotes, ceux-ci appartiennent à différentes lignées bactériennes. Leur localisation dans le bactériome varie aussi selon les espèces : ils peuvent être hébergés dans des bactériocytes spécifiques ou bien à la périphérie des cellules contenant *Buchnera* (Manzano-Marín et al., 2017; Manzano-Marín et al., 2020b). Pour la plupart des co-symbiotes obligatoires étudiés jusqu'à présent, un rôle dans la supplémentation nutritionnelle a été inféré par des analyses génomiques. Dans la plupart des cas, les gènes impliqués dans la synthèse de la riboflavine et la biotine, absents du génome de *Buchnera aphidicola* chez ces espèces, ont été identifiés et annotés chez les co-symbiotes. *Buchnera* continue de produire tous les Acides Aminés essentiels (sauf chez *Cinara cedri*, (Gosalbes et al., 2008b)) et les co-symbiotes sont impliqués dans la production de vitamines B. Mais des études plus complètes des génomes, chez un de ces symbiotes, *Erwinia haradae* associé à un clade particulier de *Cinara*, ont aussi montré que celui-ci avait une capacité métabolique supplémentaire : la synthèse de thiamine (Manzano-Marín et al., 2020b). La dépendance vis-à-vis de ce symbiote au cours du développement du puceron n'est donc peut-être pas concomitante de celle existante vis-à-vis de *Serratia*.

Ici nous avons mesuré la densité en symbiotes au cours de leur développement chez trois espèces de *Cinara* associées à *Buchnera aphidicola* comme symbiote primaire et à différents co-symbiotes acquis indépendamment : *Serratia symbiotica* récemment acquise et partiellement extracellulaire chez l'espèce *C. tujafilina*, *Serratia symbiotica* anciennement acquise et intracellulaire chez l'espèce *C. palaestinensis* et *Erwinia haradae* anciennement acquise et intracellulaire chez l'espèce *C. cuneomaculata*. Ainsi, nous voulions tester différentes hypothèses et leurs prédictions quant à l'influence de ces différentes caractéristiques du co-symbiote sur la variation de la dynamique symbiotique et donc sur la capacité de régulation du puceron. 1) Les deux bactéries ne peuvent pas être régulées indépendamment par le puceron ou les besoins pourvus par chacun des symbiotes (apport en AA pour *Buchnera*, apport en vitamines B pour le co-symbiote) sont synchrones au cours du développement des pucerons, on s'attend alors à des dynamiques très similaires entre *Buchnera* et le co-symbiote quelles que soient l'espèce hôte et la souche bactérienne nouvellement acquise. 2) La régulation est plus efficace chez les symbiotes les plus anciennement acquis (ayant donc une longue histoire co-évolutive avec leurs hôtes) et intracellulaires. Dans ce cas on s'attend à avoir des dynamiques variables entre symbiotes selon l'espèces hôte et, avec une moins bonne régulation

de *Serratia symbiotica* chez *C. tujafilina*. 3) Les pucerons peuvent réguler leurs co-symbiotes au cours de leur développement en fonction de leurs besoins. Dans ce cas, comme sous l'hypothèse 2, on s'attend à avoir des dynamiques variables des deux symbiotes selon les espèces hôtes et au cours du développement, pas nécessairement liée à l'âge de la symbiose. Malheureusement, les quantifications de symbiotes chez l'espèce *C. cuneomaculata* (associée à *Erwinia*) n'ayant pas encore été finalisées, celles-ci ne seront pas présentées dans la suite.

II. Matériels et méthodes

Matériel biologique

Nous avons utilisé trois espèces de *Cinara* qui abritent des co-symbiotes acquis de façon indépendante : 1) *Cinara tujafilina* dont la *Serratia* est récemment acquise, cette *Serratia* a un génome proche des *Serratia* facultatives et se retrouve à la périphérie des bactériocytes de *Buchnera*, 2) *Cinara palaestinensis* dont la *Serratia* est ancestralement acquise dans un clade de *Cinara* (Meseguer et al. 2017, voir chapitres 1 et 2), 3) *Cinara cuneomaculata* associé au co-symbiote *Erwinia haradae* qui a une longue histoire co-évolutive avec un clade de *Cinara* (Manzano-Marín et al., 2020b). Pour chacune de ces espèces, des individus ont été récoltés sur le terrain, puis mis en élevage à partir de colonies distinctes (Table 1). Chacun des élevages a été mis en place à partir d'un seul individu parthénogénétique afin d'obtenir des lignées clonales. Pour chacune des espèces trois lignées distinctes (c'est-à-dire des individus issus de trois colonies différentes) ont été utilisées (Table 1).

Les élevages sont réalisés sur les plantes hôtes spécifiques des espèces de *Cinara*, sur des arbres de petites taille: *Thuja orientalis* (30-40cm) pour *C. tujafilina*, *Pinus halepensis* (15-30cm) pour *C. palaestinensis* et *Larix decidua* pour *Cinara cuneomaculata*, achetés régulièrement en pépinière. Les jeunes plants sont placés à l'intérieur d'une cage de plexiglass et les cages sont disposées dans une chambre climatique maintenue à température constante 21°C et un cycle journalier de 16h jour et 8h nuit, pour maintenir la reproduction asexuée (Figure 1).

Table 1 : Localisation d'échantillonnage des différentes lignées de *Cinara* qui ont été mises en élevage.

Espèce	Lignée	Localité	Date de collecte	GPS	Plante hôte
<i>C. tujafilina</i>	1	Montpellier (Occitanie, Hérault)	26/11/2020	43.614365 ; 3.861511	<i>Thuja orientalis</i>
<i>C. tujafilina</i>	2	Aquitaine	13/11/2020	43.945405 ; -0.522269	<i>Thuja orientalis</i>
<i>C. tujafilina</i>	3	Aquitaine	13/11/2020	43.945405 ; -0.522269	<i>Thuja orientalis</i>
<i>C. palaestinensis</i>	1	Montpellier (Occitanie, Hérault)	03/07/2020	43.682302 ; 3.874883	<i>Pinus halepensis</i>
<i>C. palaestinensis</i>	2	Montpellier (Occitanie, Hérault)	03/07/2020	43.682302 ; 3.874883	<i>Pinus halepensis</i>
<i>C. palaestinensis</i>	3	Montpellier (Occitanie, Hérault)	03/07/2020	43.682302 ; 3.874883	<i>Pinus halepensis</i>
<i>C. cuneomaculata</i>	1	Le Vigan (Occitanie, Gard)	20/10/2020	44.090565 ; 3.545675	<i>Larix decidua</i>
<i>C. cuneomaculata</i>	2	Le Vigan (Occitanie, Gard)	17/11/2020	44.090565 ; 3.545675	<i>Larix decidua</i>
<i>C. cuneomaculata</i>	3	Le Vigan (Occitanie, Gard)	17/11/2020	44.090565 ; 3.545675	<i>Larix decidua</i>



Figure 1 : Dispositif expérimental d'élevage des pucerons *C. tujafilina* sur *Thuya orientalis*.

Echantillonnage des stades de développement

Dans une étude préliminaire nous avons déterminé le temps de développement à 21°C de chacune des espèces, ainsi que le nombre de stades larvaires et les âges approximatifs auxquels ont lieu les mues (Table 2, Figure S1, S2 et S3) (Durak and Borowiak-Sobkowiak, 2013). Cela nous a permis de calibrer notre expérimentation en déduisant les différents âges auxquels prélever des individus afin de quantifier les symbiotes à chacun des stades larvaires et durant toute la vie du puceron.

Table 2 : Résumé des paramètres de développement de *C.tujafilina*, *C.palaestinensis* et *C. cuneomaculata* à 21°C. décrivant le nombre de stades larvaires, le nombre de jours post naissance auquel les mues ont lieu, ainsi que la durée de vie des pucerons.

Espèce	Nb stades	Mue 1 (jours)	Mue 2 (jours)	Mue 3 (jours)	Adulte (jours)	Durée de vie (jours)
<i>C.tujafilina</i>	3	2-3	6-7	NA	11-12	39-42
<i>C.palaestinensis</i>	4	4-6	8-9	10-11	12-13	38-41
<i>C. cuneomaculata</i>	4	3-4	5-6	8-9	11-12	36-39

Pour obtenir des individus du même âge, des femelles adultes sont récupérées à l'aide d'un pinceau et placées sur un plant vide et laissées 24h pour leur permettre de produire des larves. Une fois les adultes retirés, les larves sont laissées sur le jeune plant afin de leur permettre de grandir. Par la suite, afin de quantifier les symbiotes, nous avons prélevé des individus à différents âges pour chacune des espèces. Pour *C. tujafilina* les individus ont été prélevés à 1, 3, 7, 10, 13, 24 et 38 jours, pour *C. palaestinensis* à 1, 3, 6, 9, 11, 14, 24 et 38 jours et pour *C. cuneomaculata* à 1, 4, 6, 9, 11, 14, 24 et 38. Cela nous a permis de récupérer des pucerons de chaque stade larvaire ainsi que des adultes jeunes et âgés. Pour chaque classe d'âge, 20 individus sont récupérés au total et conservés dans des tubes d'alcool 96°.

Pour quantifier les symbiotes au sein des pucerons, nous avons utilisé des PCR quantitatives en temps réel (qPCR) (voir encadré ci-dessus). Celle-ci nous ont permis de comptabiliser le nombre de copies de génome de chaque bactérie à partir du gène monocopie *gyrB* et le nombre de copies du génome du puceron à partir du gène monocopie *ef1-alpha*. Pour cela des qPCR ont été réalisées sur des extraits d'ADN des individus échantillonnés tout au long du développement de chaque lignée. Pour obtenir des quantifications absolues du nombre de copies génomiques, nous avons utilisé des gammes étalons réalisées par clonage bactérien.

Protocole de quantification

Le principe de la PCR en temps réel repose sur la possibilité de suivre la quantité d'ADN présente dans la réaction à tout instant. Des sondes fluorescentes se fixent soit sur l'ADN double brin grâce à la technologie SYBR pour laquelle la quantité intercalée dans la double hélice est proportionnelle à la quantité et à la longueur des doubles hélices d'ADN présentes dans la solution. Ces sondes ne fluorescent qu'une fois fixées à l'ADN. Un seuil de fluorescence est établi par le programme de l'appareil de PCR en temps réel. Une fois que la quantité d'ADN permet aux sondes fluorescentes de dépasser ce seuil alors on obtient un nombre de cycles PCR appelé « Ct » pour « Threshold Cycle » soit « cycle seuil ». C'est cette valeur qui est à la base des calculs pour quantifier l'ADN de façon absolue ou relative. Il est important de connaître l'efficacité E de la PCR. Pour cela, on effectue une PCR en temps réel sur des échantillons de dilution croissante pour obtenir une courbe étalon correspondant au fragment d'intérêt. Plus souvent, une PCR en temps réel sur une série de dilution avec une quantité d'ADN initiale connue permet de calculer l'efficacité de la réaction et ainsi en déduire la quantité d'ADN dans les échantillons testés.

Extraction

Nous avons extrait l'ADN de chaque individu de chaque classe d'âge (Oliver et al., 2006) en suivant le protocole d'extraction du kit BioBasic Inc, EZ-10 96 Well Plate Genomic DNA Isolation Kit (For Animal). Celles-ci sont faites sur des individus entiers et sans broyage. Pour chaque individu, on récupère 90µL de solution d'élution d'ADN qui est alors stocké à -20°C.

Conception des amorces

Afin de pouvoir quantifier par qPCR les symbiotes bactériens *Buchnera*, *Serratia* et *Erwinia* chez nos trois espèces, des amorces ont été conçues à partir des séquences de référence des génomes des symbiotes avec le logiciel geneious v11.1.5 (<https://www.geneious.com>). Les amorces ont été produites pour amplifier un fragment du gène monocopie GyrB (DNA gyrase subunit B) présent chez chacun des symbiotes ainsi que pour un fragment du gène monocopie du puceron hôte Ef1alpha (Elongation Factor 1-alpha) (Table 3). Ce dernier nous permet de normaliser nos résultats. Les amorces ont été construites de manière à respecter les conditions suivantes : une taille comprise entre 18 et 24 bp, un Tm de 60±1 degrés Celsius, une extrémité 3' contenant au moins 3 bases discriminantes parmi les 4 premières bases, contenir au moins 8

bases discriminantes au total et une position permettant d'obtenir des amplicons entre 150 et 300 pb. Une fois les amorces construites, celle-ci sont vérifiées à l'aide du logiciel Perlprimer 1.1 afin de vérifier qu'elles ne produisent pas de dimère d'amorce. Le logiciel BLAST est utilisé afin de déterminer la spécificité des amorces c'est-à-dire de vérifier si les amorces ne s'accrochent pas sur d'autres parties du génome ou d'autres bactéries.

Courbe standard : clonage

Afin de produire des gammes étalons pour la qPCR, il nous a fallu synthétiser les fragments de gènes en quantité élevée et connue grâce à un clonage bactérien. Le clonage consiste à insérer des fragments d'ADN amplifiés par PCR dans un vecteur qui lui-même sera introduit dans une bactérie. Un vecteur ne reçoit qu'un seul fragment d'ADN amplifié et une bactérie ne reçoit qu'un vecteur. Les bactéries mises en culture vont se multiplier et donner naissance à des colonies. Toutes les bactéries d'une même colonie possèdent donc le même insert. Dans notre cas, le clonage a été effectué pour chacun des fragments de gènes décrit précédemment (Table 3).

Nous avons donc dans un premier temps effectué une PCR de chacun de nos fragments de gènes avec les amorces précédemment décrites (Table 3). Les fragments amplifiés ont alors été purifiés (kit nucleospin gel and PCR clean-up) et la quantité d'ADN a été mesurée via Qubit pour vérifier la présence d'une quantité d'ADN suffisante pour le clonage. Le clonage a été effectué avec le kit PROMEGA pGEM®-t. Les amplicons ont été insérés par une réaction de ligation dans les vecteurs pGEM®-T easy (kit PROMEGA pGEM®-t). Une transformation de bactérie hautement compétente OneShot® TOP10 (Invitrogen) a ensuite été effectuée pour insérer le vecteur contenant l'amplicon. Cette insertion confère alors à la bactérie une résistance à l'ampicilline. Les bactéries transformées ont été mises en culture dans des boîtes de pétrie contenant du milieu LB (bouillon lysogène) avec agar, 100 mg/mL d'ampicilline, X-gal (5-bromo-4chloro-3-indolyl- β -D-galactopyranoside) et IPTG ; seules les colonies contenant le vecteur se développent. La présence des réactifs (X-gal et IPTG) sur le milieu de culture permet de distinguer les colonies possédant l'insert dans le vecteur (colonies blanches) de celles qui ne l'ont pas (colonies bleues). Les colonies contenant l'insert souhaité ont ensuite été mises en culture liquide pour augmenter la quantité d'ADN.

Table 3 : Liste des amorces utilisées pour les qPCR, ainsi que la longueur des amplicons associés.

Nom forward ; Nom reverse	Espèces	Cibles	Séquences forward ; Séquences reverse	Longueur de l'amplicon (pb)	Efficacité
Ef1a_Ctuj_F757 ; Af1a_Ctuj_R949	<i>C. tujafilina</i>	Ef1-alpha	TTGGAACAGTCCCAGTTGGTC ; TTGTGTCACCAGCAACGAAAC	213	2
gB_Bu_Ct_F814 ; gB_Bu_Ct_R971	<i>C. tujafilina</i>	GyrB ; Buchnera	GTACCTCAAAAAGATGGTGGAAAGT ; TGATAGCTGTTAGTCCTTCACGAAT	151	1.88
gB_Se_Ct_F523 ; gB_Se_Ct_R767	<i>C. tujafilina</i>	GyrB ; Serratia	TCAGACCTTCACTAACGCCAC ; CTACTTCCACGCCGATGTCAT	245	2
Ef1a_Cpala_F339 ; Ef1a_Cpala_R493	<i>C. palaestinensis</i>	Ef1-alpha	TGCTTTTCGTCCCAATCTCA ; GACTGGGTGGCAGGATAG	155	2
gB_Bu_Cp_F991 ; gB_Bu_Cp_R974	<i>C. palaestinensis</i>	GyrB ; Buchnera	TGCTATGCAATGGAACGAAACC ; AGTGCTGTTAATCCTTCTCTTGTA	204	1.97
gB_Se_Cp_F1,685 ; gB_Se_Cp_R1,813	<i>C. palaestinensis</i>	GyrB ; Serratia	TCGCAATAGATAGTACATACCTGC ; CAATCGGTTAAGCAGTGACGC	152	1.94
Ef1a_Cpala_F339 ; Ef1a_Cpala_R493	<i>C. cuneomaculata</i>	Ef1-alpha	TGCTTTTCGTCCCAATCTCA ; GACTGGGTGGCAGGATAG	155	NA
GyrB_Bu_Ccune_1,121F ; GyrB_Bu_Ccune_1,278R	<i>C. cuneomaculata</i>	GyrB ; Buchnera	GATCCTCCTGCTGAATCACCT ; GCAAGATCTAGAGAAGCGGCA	158	NA
GyrB_Er_Ccune_86F ; GyrB_Er_Ccune_272R	<i>C. cuneomaculata</i>	GyrB ; Erwinia	ATACAGATGACGGCACAGGTC ; GCAGCAGACACTCCTTCTTCT	187	NA

Une méthode de lyse alcaline a été utilisée pour extraire les plasmides et un screening PCR effectué avec les amorces T7 et SP6 (kit PROMEGA pGEM®-t) pour vérifier la présence de l'insert et sa correspondance avec la taille du fragment d'intérêt via une électrophorèse sur gel d'agarose à 1.5 %. Les amplifications de plasmide contenant la taille d'insertion correcte ont ensuite été purifiées (kit qiagen minipreps) pour obtenir des Minipreps qui ont été dosées (par Qubit) et séquencées pour vérifier que l'insert soit bien le fragment de gène d'intérêt. Finalement, des dilutions sérielles au dixième des Minipreps ont été faites, celle-ci seront utilisées pour réaliser les courbes standards pour chacun des gènes.

Le nombre de copies de l'amplicon du gène *gyrB* des deux bactéries endosymbiotiques et le nombre de copies de l'amplicon du gène *efl-alpha* des pucerons dans les Minipreps ont été calculées avec l'équation suivante :

$$N = \frac{\text{quantité d'ADN (g} \cdot \mu\text{L}^{-1}) \times \text{volume Minipreps (}\mu\text{L)}}{\text{masse molaire (vecteur + insert) (g} \cdot \text{mol}^{-1})} \times 6.022 \times 10^{23} \text{mol}^{-1}$$

N correspond au nombre de copies, la quantité d'ADN correspondant est celle mesurée dans les Minipreps en g/μL, $6.022 \times 10^{23} \text{ mol}^{-1}$ est le nombre d'Avogadro (nombre de molécules dans une mol) et la masse molaire du vecteur et de l'insert est calculée à l'aide de la formule :

$(An \times 313.2) + (Tn \times 304.2) + (Cn \times 289.2) + (Gn \times 329.2) + (Tn \times 313.2) + (An \times 304.2) + (Gn \times 289.2) + (Cn \times 329.2)$ (<https://www.thermofisher.com/fr/fr/home/references/ambion-tech-support/rna-tools-and-calculators/dna-and-rna-molecular-weights-and-conversions.html>) avec An= nombre d'adénine, Tn= nombre de thiamine, Cn= nombre de cytosine et Gn= nombre de guanine.

PCR quantitative

Les réactions qPCR en temps réel ont été réalisées sur une machine LightCycler 480 (Roche), avec une étape initiale de dénaturation à 95 C° pendant 10 minutes et 45 cycles de 95 C° pendant 5s, 60 C° pendant 15s et 72C° pendant 10s (amplification). Pour chaque réaction une mesure des températures de fusion a été réalisée pour vérifier la présence d'un seul amplicon. Chaque réaction de 10μL comprenait : 0.5μL de chaque amorce à 10μM, 5μL de SYBRGREEN à 2X, 2μL d'eau et 2μL d'ADN. Nous avons effectué trois réplicats techniques pour chaque réaction (pour chacun des individus et chacun des trois gènes). Des mix avec de l'eau stérile à la place de l'ADN ont été utilisés comme contrôles négatifs dans les essais de qPCR. Au total, 4320 réactions qPCR ont été faites pour *C. palaestinensis* (soit 3 lignées x 8

âges x 20 individus x 3 replicats x 3 gènes) et 2835 pour *C.tujafilina* (soit 3 lignées x 7 âges x 15 individus x 3 replicats x 3 gènes) l'ensemble sur des plaques PCR 384 puits.

Nous avons calculé une valeur Ct pour chaque individu en calculant la valeur Ct moyenne des trois réplicats techniques puis, si celles-ci étaient cohérentes entre elles. La densité des symbiotes a été obtenue comme mesure du nombre de copies du gène cible sélectionné en utilisant les valeurs Ct obtenues pour chaque échantillon par rapport à la courbe standard produite pour chaque gène cible.

Pour cela nous avons d'abord estimé l'efficacité de l'amplification des amorces (reporté Tab 3) avec l'équation de la courbe des standards :

$$Ct = \frac{\text{Log}(N_C)}{\text{Log}(E)} - \frac{\text{Log}(N_0)}{\text{Log}(E)}$$

Soit :

$$E = \left(\frac{N_C}{N_0}\right)^{\frac{1}{Ct}}$$

E étant l'efficacité de la PCR. N_C est le nombre de produit PCR présent au Ct (Cycle Threshold) et N_0 est le nombre de copie de du vecteur + amplicon dans la solution de départ.

Nous avons ensuite utilisé l'équation de la courbe standard pour déterminer le nombre de copies absolues de nos échantillons. L'abondance (ou le titre de symbiote) est définie ici comme le rapport entre le nombre de gènes à copie unique du symbiote et le nombre de gènes à copie unique de puceron, et constitue donc une estimation du nombre de génomes de symbiote par rapport au nombre de cellules hôtes dans un échantillon. Le calcul de ce ratio nous permet d'obtenir une abondance normalisée et donc de se détacher de potentiels biais pouvant venir de la taille de l'individu ou de la qualité de l'extraction d'ADN. L'ensemble du traitement des données sortie de l'appareil qPCR a été automatisé par script R studio (R Core Team, 2022).

Analyses

Une fois le log N_0 calculé pour chaque gène, nous avons estimé des « densités » de bactéries, en calculant les ratio $(\log N_0 \text{ Buchnera})/(\log N_0 \text{ pucerons})$ et $(\log N_0 \text{ Serratia})/(\log N_0 \text{ pucerons})$ pour chaque échantillon. Ainsi les estimations du N_0 des deux bactéries sont normalisées avec les N_0 du puceron, ceci afin de bien mesurer un effet « stade de

développement » et non un effet de la variation de tailles et du nombre de cellules du puceron extrait.

Microscopie à hybridation in situ par fluorescence

Nous avons étudié la localisation et la morphologie des symbiotes chez des individus des trois espèces. Ceux-ci ont d'abord été fixés lors de la collecte sur le terrain dans un fixateur de Carnoy modifié (6 chloroforme : 3 éthanol absolu : 1 acide acétique glacial) et laissés pendant la nuit, en suivant le protocole de (Koga et al., 2009). Les individus ont ensuite été blanchis dans une solution de 6% de H₂O₂ diluée dans de l'éthanol absolu et ont été laissés dans cette solution pendant 2 semaines (en changeant la solution tous les 3 jours), puis lavés deux fois avec de l'éthanol absolu. Ils ont ensuite été disséqués dans de l'éthanol absolu pour extraire les embryons. L'hybridation a été réalisée pendant une nuit à 30 °C dans un tampon d'hybridation standard (20mM Tris-HCl [pH 8,0], 0,9M NaCl, 0,01% SDS et 30% formamide), puis lavée (20mM Tris-HCl [pH 8,0], 5 mM EDTA, 0,1M NaCl et 0,01% SDS) avant la préparation des lames (Manzano-Marín et al., 2020b). Les embryons d'un maximum de 10 individus ont été observés sous un microscope confocal. Nous avons utilisé des sondes spécifiques compétitives pour les symbiotes *Buchnera* et *Serratia* dans *Cinara* (Table 4). Les images au format TIFF ont été exportées à l'aide de la suite d'applications X de Leica. Les images exportées ont été importées dans Inkscape v0.92.4 pour construire les figures publiées.

Table 4 : Liste des amorces utilisées pour la microscopie à hybridation in situ par fluorescence.

Espèce hôte	Bactérie	Sonde	Séquence	Référence
<i>C. tujaefilina</i>	<i>Buchnera</i>	BLach-2x6-FAM	5'-2x6-FAM- CCCGTTYGCCGCTCGCCGTCA-3'	(Manzano-Marín et al., 2017)
<i>C. tujaefilina</i>	<i>Serratia</i>	Ssym-Cy3	5'-Cy3-CTCGTCACCCAGAGAGCA-3'	(Manzano-Marín et al., 2017)
<i>C.palaestinensis</i>	<i>Buchnera</i>	BLach-FITC	5'-FITC- CCCGTTCGCCGCTCGCCGGCA-FITC-3'	(Manzano-Marín et al., 2017)
<i>C.palaestinensis</i>	<i>Serratia</i>	SeJR5Cinara-Cy3	5'-Cy3-CCGTYCGCCGCTCGTCACC- Cy3-3'	Cette etude
<i>C. cuneomaculata</i>	<i>Buchnera</i>	BLach-FITC	5'FITC-CCCGTTYGCCGCTCGCCGTCA- FITC-3'	(Manzano-Marín et al., 2017)
<i>C. Cuneomaculata</i>	<i>Erwinia</i>	ErCinara-Cy3	5'Cy3-CCCGTTCGCCACTCGTCGBCA-3'	(Manzano-Marín et al., 2020b)

III. Résultats

Quantification de la densité en symbiote dans les espèces de Cinara

Les résultats obtenus montrent tout d'abord l'existence d'une grande variabilité interindividuelle de la quantité de symbiotes que ce soit en quantité absolue ou relative au nombre de cellules hôtes. Pour *C. palaestinensis* nos résultats montrent une dynamique de *Serratia* et de *Buchnera* similaire, on peut remarquer une plus grande proportion de *Serratia* entre 6 et 14 jours de développement ce qui correspond au stade de développement Larve 3 et 4 ainsi que jeunes adultes (Figures 3 et 4).

Les résultats obtenus pour *C. tujafilina* montrent une dynamique de *Serratia* et de *Buchnera* différente, avec une dynamique stable de *Buchnera* alors que la densité de *Serratia* augmente tout au long de la vie de l'hôte. On peut remarquer une plus grande densité de *Serratia* dès le troisième jour de développement soit au stade larvaire 2 avec un écart de densité de symbiotes qui augmente jusqu'à l'âge adulte (Figures 5 et 6).

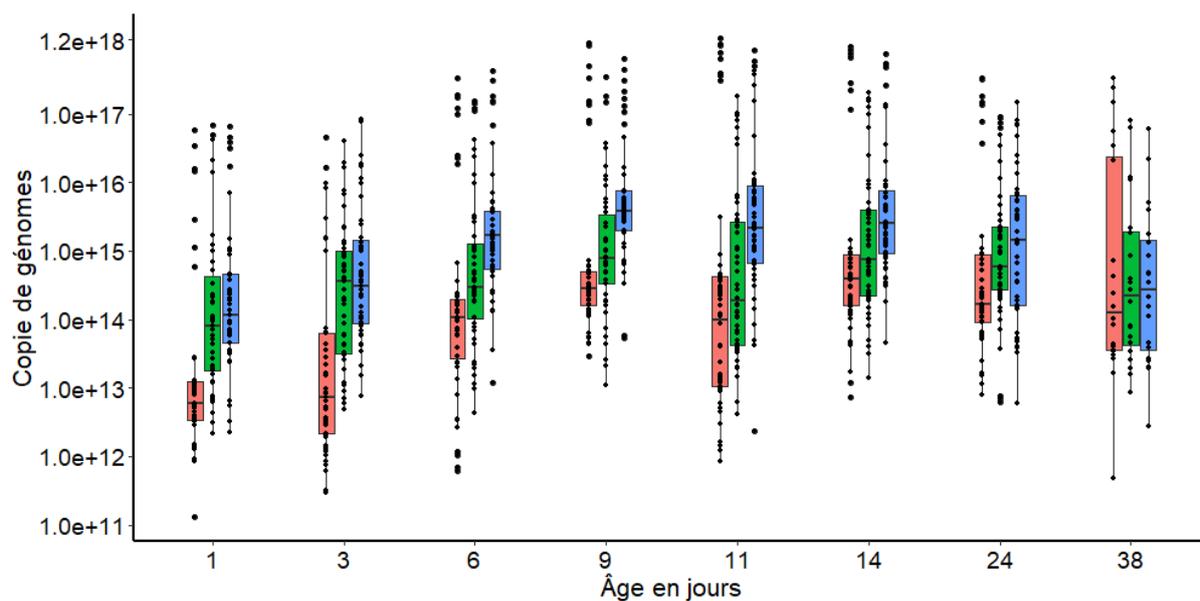


Figure 3 : Boxplot de la densité absolue en nombre de copie de gène bactérien et hôte, de *Serratia* (bleu), *Buchnera* (vert) et hôte (rouge) (quantifié via qPCR des gènes monocopies *gyrB* pour les symbiotes et *ef1-alpha* pour l'hôte) en fonction de l'âge de *C. palaestinensis*. Chaque point correspond à un individu.

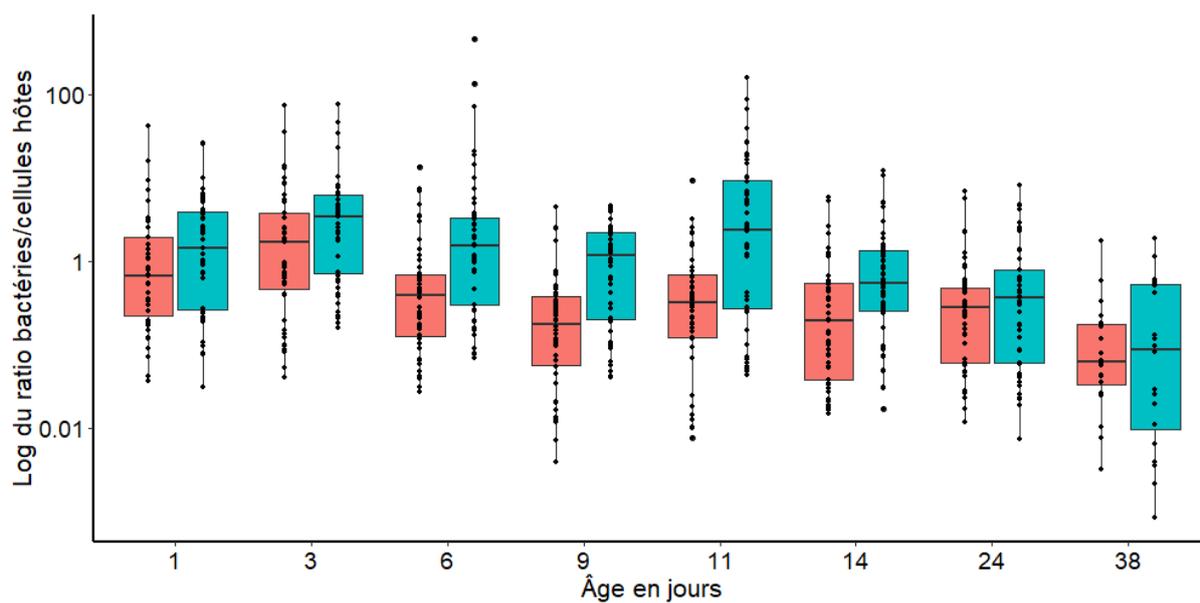


Figure 4 : Boxplot de la densité (en nombre de copie de gène bactérien par copie de gène hôte) de *Serratia* (bleu) et *Buchnera* (rouge) (quantifié via qPCR des gènes monocopies *gyrB* pour les symbiotes et *ef1-alpha* pour l'hôte) en fonction de l'âge de *C. palaestinensis*. Chaque point correspond à un individu.

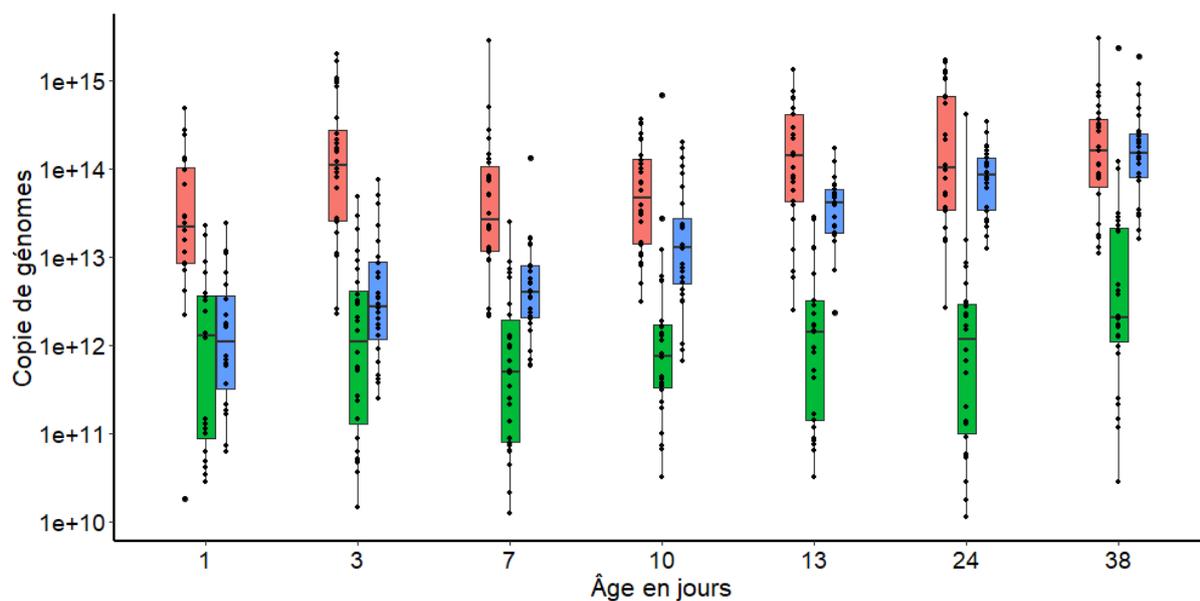


Figure 5 : Boxplot de la densité absolue en nombre de copie de gène bactérien et hôte, de *Serratia* (bleu), *Buchnera* (vert) et hôte (rouge) (quantifié via qPCR des gènes monocopies *gyrB* pour les symbiotes et *ef1-alpha* pour l'hôte) en fonction de l'âge de *C. tujaifilina*. Chaque point correspond à un individu.

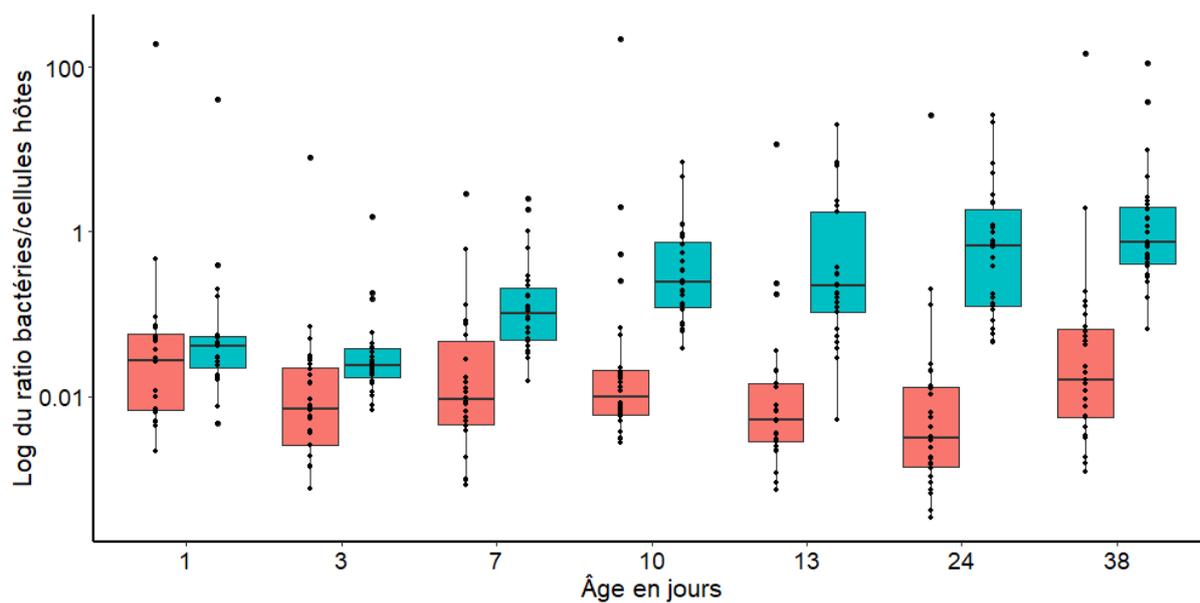
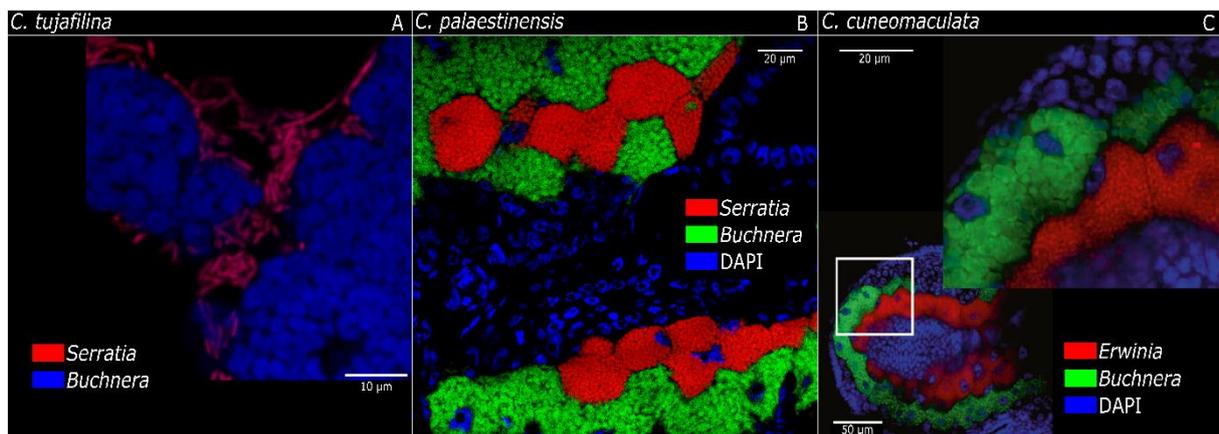


Figure 6 : Boxplot de la densité (en nombre de copie de gène bactérien par copie de gènes hôte) de *Serratia* (bleu) et *Buchnera* (rouge) (quantifié via qPCR des gènes monocopies *gyrB* pour les symbiotes et *ef1-alpha* pour l'hôte) en fonction de l'âge de *C. tujaifilina*. Chaque point correspond à un individu.

Localisation des co-symbiotes dans les espèces de *Cinara*

Les microscopies par FISH ont révélé que *Serratia symbiotica* chez *C. palaestinensis* sont dans des bactériocytes indépendants des bactériocytes de *Buchnera* (Figure 7). De précédentes observations sur *C. tujaefilina* ont révélé la présence de *Serratia* localisées hors des bactériocytes à la périphérie des bactériocytes de *Buchnera*, et celles-ci ont la particularité d'être bacilliformes (Manzano-Marín et al., 2017) (Figure 7). Quant à *C. cuneomaculata*, les FISH ont révélé la présence de *Erwinia* localisée dans des bactériocytes indépendants des bactériocytes de *Buchnera* (Manzano-Marín et al., 2020b) (Figure 7). De plus, les *Serratia* de *C. palaestinensis* et *Erwinia* de *C. cuneomaculata* présentent également une forme de cellule coccoïde plus petite que les cellules de *Buchnera*, comme décrit précédemment dans les endosymbiotes d'autres espèces de pucerons avec un génome de symbiote fortement réduit (Manzano-Marín et al., 2017).



IV. Discussion

Dans le système di-symbiotique des *Cinara*, il semble y avoir une régulation de *Buchnera* dans les deux espèces que nous avons étudié avec une densité de symbiotes relativement stable au cours du développement. Dans le cas de *C. palaestinensis*, on observe une diminution en fin de vie rappelant la dynamique de *Buchnera* observée chez le puceron du pois (Simonet et al., 2016a). Pour ce qui est des co-symbiotes *Serratia symbiotica* présents chez la majorité des *Cinara*, nous avons montré l'existence de dynamiques symbiotiques différentes chez les deux espèces étudiées. Pour la première, *C. palaestinensis*, abritant une *Serratia* anciennement acquise et intracellulaire (voir chapitre II), la dynamique observée est parallèle à celle de *Buchnera* suggérant une régulation fine de la bactérie. Le même phénomène a été observé au cours du développement pour le système di-symbiotique de *Diaphorina citri* (Hemiptera, Liviidae) abritant *Carsonella* et *Proffella* comme symbiotes obligatoires (Dossi et al., 2014). Les fluctuations similaires de la densité des deux symbiotes suggèrent un ajustement continu de leur nombre au fur et à mesure du développement de l'hôte selon les besoins métaboliques. Dans ce cas, *Buchnera* et *Serratia* agissent comme des partenaires syntrophiques et sont responsables de la biosynthèse des composés requis pour soutenir la croissance et le développement de l'hôte (Sloan and Moran, 2012; Wilkinson et al., 2007).

En revanche, pour la deuxième espèce, *C. tujafilina*, abritant une *Serratia* récemment acquise et extracellulaire, la dynamique observée est différente de celle de *Buchnera* avec une augmentation de la population de symbiotes tout au long de la vie de l'hôte suggérant alors une faible régulation par l'hôte. La croissance continue de *Serratia* pourrait être attribuée à leur tolérance/échappement au système immunitaire de l'hôte, comme cela a été suggéré pour d'autres complexes symbiote-hôte (Login and Heddi, 2013; Ratzka et al., 2012; Rio et al., 2006). Mais, elle pourrait également être due à des niveaux d'interactions différents avec leur hôte en raison de l'histoire évolutive de leur association avec l'hôte et des interactions entre symbiotes. En effet, *Buchnera*, le symbiote le plus ancien, dispose de taux de croissance plus faibles que *Serratia* acquis récemment qui lui se développe à un rythme rapide. Ces résultats sont donc en accord avec notre hypothèse 2 puisque nous observons une régulation moins efficace de la lignée de *Serratia* récemment acquise. Néanmoins, l'absence de résultats pour la lignée de co-symbiote *Erwinia*, ne permet pas de conclure quant à l'existence d'une dynamique symbiotique finement régulée chez les co-symbiotes anciens tel qu'observé chez *C. palaestinensis*. De plus, l'effet de la présence de co-symbiotes appartenant à une lignée différente reste à élucider.

Les premiers résultats suggèrent tout de même que le puceron contrôle moins bien un nouveau partenaire qui est généralement à la périphérie des bactériocytes. Cela paraît cohérent avec ce que l'on sait des mécanismes de régulation de la densité de *Buchnera* par le puceron du pois. Chez celui-ci, la dynamique intra-hôte a été décrite comme un processus activé progressivement tout au long de l'âge adulte du puceron et se caractérise par : 1) une accumulation cytoplasmique de vacuoles non autophagiques, 2) une séquence de réponses au stress cellulaire avec formation d'autophagosomes dans les espaces inter-vacuolaires et activation de dérivés réactifs de l'oxygène et 3) une dégradation de *Buchnera* par le système lysosomal. On peut alors se demander si ces processus de régulation sont les mêmes pour *Serratia*, ce qui expliquerait la différence de régulation que nous observons entre nos deux lignées. Ainsi, même si il a été montré dans d'autres systèmes que la régulation symbiotique peut être faite via les défenses immunitaires innées de l'hôte (Login et al., 2011b; Login and Heddi, 2013; Ratzka et al., 2012; Weiss et al., 2012), la compartimentation dans des bactériocytes joue probablement un rôle important pour la régulation des co-symbiotes des *Cinara*, comme observé dans la plupart des systèmes symbiotiques (Chomicki et al., 2020). Celle-ci permet par exemple de réguler les services rendus par les symbiotes en fonction des besoins de l'hôte par la régulation de l'expression des gènes du symbiote ou la provision en substrats nécessaires aux réactions métaboliques réalisées par les symbiotes (Macdonald et al., 2011; Pers and Hansen, 2021).

Cependant afin de confirmer nos résultats, il sera nécessaire d'effectuer des mesures absolues du nombre de bactéries (par exemple par cytométrie de flux et suivi par FISH). En effet, de précédentes études ont montré que les bactéries endosymbiotiques chez les pucerons sont fortement polyploïdes (Komaki and Ishikawa, 1999) et que le nombre de copies de génome varie au cours du développement de l'hôte (Komaki and Ishikawa, 2000). Il nous est donc difficile d'évaluer le nombre de bactéries au cours du temps via qPCR, or, les mécanismes de régulation de symbiote décrit à ce jour agissent sur le nombre de bactérie et du bactériocyte et non sur la ploïdie.

Afin de vérifier l'existence d'une régulation en fonction des besoins du puceron, de futures recherches consisteront à élever des pucerons sur milieu artificiel suppléé et/ou privé des nutriments apportés par les symbiotes, afin de tester si le puceron ajuste sa quantité de symbiotes en fonction de sa nutrition. Ceci a été démontré pour *Buchnera* par des expérimentations sur le puceron du pois en réponse à une manipulation du contenu en acides aminés du régime alimentaire (Colella et al., 2018; Viñuelas et al., 2011), et en réponse à une

manipulation des apports en azote (Wilkinson et al., 2007). Dans d'autres modèles tels que la mouche tsétsé, des expérimentations similaires ont également démontré la présence d'une régulation des symbiotes en fonction de l'apport en thiamine (Snyder et al., 2012, 2010). En parallèle, une meilleure connaissance des besoins au cours du développement pourrait également amener à mieux comprendre ces dynamiques. Notamment par des analyses de la composition de l'alimentation des pucerons via stylectomie qui permet de récupérer le phloème ingéré par les pucerons afin d'analyser son contenu (Gaupels et al., 2008a, 2008b).

Néanmoins, tester si in fine le contrôle de la densité des symbiotes améliore réellement la valeur sélective de l'insecte hôte est difficile compte tenu de la difficulté de prédire la densité de symbiotes optimale en fonction des besoins de l'insecte et surtout du manque de connaissances quant aux besoins de l'hôte. Le genre *Cinara* utilisé dans cette étude semble cependant être un modèle prometteur pour mener à bien des tests et des analyses comparatives de la régulation des symbiotes puisque ce groupe dispose d'une grande diversité de co-symbiotes aux âges d'acquisitions différents, aux fonctions potentiellement variables ainsi qu'aux structures différentes.

Chapitre 4
The protector within: Comparative genomics of APSE phages across aphids reveals rampant recombination and diverse toxin arsenals

Contexte de l'étude

Les données de séquençage disponibles ont également révélé la présence de génomes de bactériophages chez certaines espèces de pucerons du genre *Cinara*. Les données sur la diversité de ces virus chez les bactéries endosymbiotiques et leur rôle dans leur histoire évolutive étant rares, nous avons également effectué une étude phylogénomique sur ces phages. Les bactériophages sont des virus qui infectent et se répliquent dans les bactéries, ils présentent des génomes très dynamiques (Dion et al., 2011). Les APSEs (pour *Acyrtosiphon pisum* secondary endosymbiont) sont un type de phage qui ont été décrits dans un premier temps chez les endosymbiotes de *A. pisum* (van der Wilk et al., 1999). Par la suite, ce phage a été trouvé associé à la bactérie endosymbiotique *Hamiltonella defensa* présente chez plusieurs espèces d'hémiptères. Dans le cas des pucerons, il est admis que les gènes de toxines portés par ce phage confèrent un phénotype protecteur contre les guêpes parasitoïdes (Brandt et al., 2017; Oliver et al., 2009). L'analyse microscopique et le séquençage de plusieurs génomes de ce phage ont montré qu'il en existe plusieurs types (7 décrits aujourd'hui) (Degnan and Moran, 2008; Moran et al., 2005). Parmi ces différents types d'APSE, des études ont montré l'existence de différences dans le niveau de protection selon l'association d'espèces / génotypes de pucerons et de combinaisons *Hamiltonella*-APSE (Asplen et al., 2014; Hopper et al., 2018; Lenhart and White, 2017; Leybourne et al., 2020).

Afin d'élargir nos connaissances sur la diversité des phages APSEs chez les espèces de pucerons, nous avons assemblé et annoté les génomes complets de 16 phages APSE qui infectent les endosymbiotes de *Hamiltonella* dans 10 espèces d'insectes, dont plusieurs pucerons et *Bemisia tabaci* (Manzano-Marin). Nous avons effectué des analyses phylogénétiques sur *Hamiltonella* et ses phages APSE associés afin de comprendre la diversité et la trajectoire évolutive de ce phage défensif. Par le biais de la génomique comparative, nous avons étudié le réarrangement du génome des APSE (génome de 35kb environ comportant 34 à 42 gènes codants) et regardé plus spécifiquement l'évolution de la cassette de toxines

The Protector within: Comparative Genomics of APSE Phages across Aphids Reveals Rampant Recombination and Diverse Toxin Arsenals

Jeff Rouil^{1,*}, Emmanuelle Jouselin¹, Armelle Coeur d'acier¹, Corinne Cruaud², and Alejandro Manzano-Marín^{3,*},[†]

¹UMR 1062 Centre de Biologie pour la Gestion des Populations, INRAE, CIRAD, IRD, Montpellier SupAgro, Université de Montpellier, France

²Genoscope, Institut de Biologie François-Jacob, Commissariat à l'Energie Atomique (CEA), Université Paris-Saclay, Évry, France

³Centre for Microbiology and Environmental Systems Science, University of Vienna, Austria

*These authors contributed equally to this work.

[†]Corresponding author: E-mail: alejandro.manzano.marin@gmail.com.

Accepted: 1 May 2020

Data deposition: The annotated genome sequences from this project have been deposited at the European Nucleotide Archive under the accessions PRJEB15504, PRJEB31183, PRJEB37342, PRJEB37357, and PRJEB37362–PRJEB37365. The low-coverage *Hamiltonella* drafts are available in [supplementary file S4](#), [Supplementary Material](#) online.

Abstract

Phages can fundamentally alter the physiology and metabolism of their hosts. Although these phages are ubiquitous in the bacterial world, they have seldom been described among endosymbiotic bacteria. One notable exception is the APSE phage that is found associated with the gammaproteobacterial *Hamiltonella defensa*, hosted by several insect species. This secondary facultative endosymbiont is not necessary for the survival of its hosts but can infect certain individuals or even whole populations. Its infection in aphids is often associated with protection against parasitoid wasps. This protective phenotype has actually been linked to the infection of the symbiont strain with an APSE, which carries a toxin cassette that varies among so-called “types.” In the present work, we seek to expand our understanding of the diversity of APSE phages as well as the relations of their *Hamiltonella* hosts. For this, we assembled and annotated the full genomes of 16 APSE phages infecting *Hamiltonella* symbionts across ten insect species. Molecular and phylogenetic analyses suggest that recombination has occurred repeatedly among lineages. Comparative genomics of the phage genomes revealed two variable regions that are useful for phage typing. Additionally, we find that mobile elements could play a role in the acquisition of new genes in the toxin cassette. Altogether, we provide an unprecedented view of APSE diversity and their genome evolution across aphids. This genomic investigation will provide a valuable resource for the design and interpretation of experiments aiming at understanding the protective phenotype these phages confer to their insect hosts.

Key words: APSE phage, *Hamiltonella defensa*, endosymbiont, toxin, parasitoid defense.

Introduction

Bacteriophages, the viruses that infect and replicate in bacteria, exhibit genomes that are highly dynamic (reviewed by Dion et al. [2020]). They undergo rampant recombination and constant rearrangements. They are also known to play a prominent role in horizontal gene transfers between bacteria (reviewed by Touchon et al. [2017]). As such, they can be a source of innovation for the bacteria that are infected as well

as for the eukaryotic organisms that carry these bacteria. Aphids (Hemiptera: Aphididae) can host a wide variety of facultative endosymbiotic bacteria (Guo et al. 2017), in addition to their obligate nutritional endosymbiont *Buchnera aphidicola*. These bacteria are not necessary for reproduction nor survival but can endow their host with a variety of beneficial effects, ranging from survival after heat stress to defense against pathogenic fungi and parasitoid wasps (reviewed by

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Table 1

Previously Characterized APSE Toxins

Insect Host	<i>Hamiltonella</i> Strain	APSE Type	Toxin	Reference
<i>Acyrtosiphon pisum</i>	NA	APSE-1	Shigga-like	van der Wilk et al. (1999)
<i>Acyrtosiphon pisum</i>	5AT	APSE-2	CdtB	Moran et al. (2005); Degnan and Moran (2008)
<i>Acyrtosiphon pisum</i>	NY26	APSE-2	CdtB	Brandt et al. (2017)
<i>Acyrtosiphon pisum</i>	82B	APSE-2	CdtB	Moran et al. (2005); Degnan and Moran (2008)
<i>Acyrtosiphon pisum</i>	ZA17	APSE-2 ^a	CdtB	Martinez et al. (2014)
<i>Acyrtosiphon pisum</i>	WA4	APSE-2	CdtB	Martinez et al. (2014)
<i>Acyrtosiphon pisum</i>	A1A, A2F, A2H	APSE-3	YD-repeat	Degnan and Moran (2008)
<i>Acyrtosiphon pisum</i>	AS3	APSE-3	YD-repeat	Oliver et al. (2009)
<i>Acyrtosiphon pisum</i>	AS5	APSE-3	YD-repeat	Oliver et al. (2009)
<i>Acyrtosiphon pisum</i>	R7	APSE-3	YD-repeat	Oliver et al. (2009)
<i>Aphis fabae</i>	H76	APSE-3	YD-repeat	Dennis et al. (2017)
<i>Aphis craccivora</i>	5ATac	APSE-4 ^b	Shigga-like	Degnan and Moran (2008)
<i>Uroleucon rudbeckiae</i>	NA	APSE-5 ^b	Shigga-like	Degnan and Moran (2008)
<i>Chaitophorus</i> sp.	N4	APSE-6	CdtB	Degnan and Moran (2008)
<i>Aphis fabae</i>	H402	APSE-6	CdtB	Dennis et al. (2017)
<i>Bemisia tabaci</i>	NA	APSE-7	CdtB	Degnan and Moran (2008)

Note.—Summary of the previous works characterizing APSE phage toxins. NA, not available.

^aLater referred to as APSE-8 (Brandt et al. 2017).

^bVariants of APSE-1, see Degnan and Moran (2008).

Oliver et al. [2010]). Although bacteriophages are ubiquitous wherever bacteria exist, their discovery is scarce among the endosymbiotic bacteria of insects. One such case is the APSE (for *Acyrtosiphon pisum* secondary endosymbiont) phage: a lambdoid phage with an isometric head and a short tail (resembling species in the family *Podoviridae*) which was originally isolated from an *Acyrtosiphon pisum* aphid infected by an unidentified secondary endosymbiont (van der Wilk et al. 1999). This phage has been found so far associated with the gammaproteobacterial symbiont species *Hamiltonella defensa*: an endosymbiont lineage that infects several hemipteran species. In the case of aphids, it is clear that toxin genes carried by this phage confer a protective phenotype against parasitoid wasps by disabling wasp development (Oliver et al. 2009; Brandt et al. 2017).

Microscopic analysis and genome sequencing of this phage, first termed APSE-1 (for bacteriophage 1 from APSE), revealed a circularly permuted and terminally redundant double-stranded DNA molecule of 36,524 base pairs (bp) (van der Wilk et al. 1999). APSE-1 encodes for a protein product showing low sequence similarity with the shiga-like toxin B subunit of several bacteriophages. Further whole-genome sequencing of a second APSE genome from *A. pisum* (termed APSE-2) as well as specific regions of other APSE phages revealed variation across the so-called “types” (Moran et al. 2005; Degnan and Moran 2008). These “types” are characterized by carrying a unique set of genes including (putative) toxins from three protein families: shiga-like toxin, cytolethal distending toxin (CdtB), and YD-repeat toxin. So far, seven types of APSE have been described (table 1). They are associated with *Hamiltonella* strains from six aphid species (from two subfamilies) and the whitefly *Bemisia tabaci*.

Differences in the level of protection have been observed in different associations of aphids species/genotypes and *Hamiltonella*–APSE combinations. For example, experimental work has shown that *Hamiltonella* endosymbionts can confer varying levels of resistance against different parasitoid wasp species, and even no defense at all (Asplen et al. 2014; Lenhart and White 2017; Hopper et al. 2018; Leybourne et al. 2020). It has been shown that *Hamiltonella* carrying APSE-3 confer a strong defensive phenotype, whereas the ones carrying APSE-2 generally confer moderate defense against parasitoids (Oliver et al. 2009; Martinez et al. 2014; Brandt et al. 2017). This protection is indeed largely dependent on the APSE “type” (Degnan and Moran 2008; Martinez et al. 2014), and experiments have conclusively demonstrated that APSE-3, carried by *Hamiltonella defensa* strain AS3, can confer its protective phenotype when transplanted to the naturally occurring APSE-free and nonprotective strain A2C (Brandt et al. 2017). The beneficial fitness effect of APSE-bearing *Hamiltonella* has been shown to be conditional to the presence of APSE and the environmental pressure of parasitoid infection. In one study, the authors found that, although apparently no cost to infection with *Hamiltonella* could be detected in aphids not exposed to parasitism, a significant decline in the frequency of *Hamiltonella* was observed (Oliver et al. 2008). In another study, the loss of APSE led to an increase in intra-aphid *Hamiltonella* abundance and was associated to “severe” deleterious effects on aphid fitness (Weldon et al. 2013).

In the current work, we sought to explore and expand our knowledge on the diversity of APSE phages across aphid species. For this purpose, we assembled and annotated full genomes for 16 APSE phages that infect *Hamiltonella*

endosymbionts across ten insect species, including several aphids (from five subfamilies) and the whitefly *B. tabaci*. We performed phylogenetic analyses on *Hamiltonella* and its associated APSE phages in order to understand the diversity and evolutionary trajectory of this defensive phage. Through comparative genomics, we investigated genome rearrangement and more specifically look at the evolution of the toxin cassette. We found evidence suggesting that recombination takes place among APSE types in aphids and that most variation in gene content is observed in two main regions of their genomes. Finally, analysis of the toxin cassettes revealed that mobile elements might be involved in at least some of the variation observed in this symbiotically relevant genomic region.

Materials and Methods

Aphid Collection, DNA Extraction, and Sequencing

Complete genome sequences of APSE phages were retrieved from the NCBI database for APSE-1 (also known as *Hamiltonella virus APSE1*) and *Hamiltonella* strain 5AT. We then used the genome data available for five *Hamiltonella* strains from the aphid *A. pisum* and two from the whitefly *B. tabaci* to extract integrated APSE phages. Additionally, we gathered sequencing data from eight different aphid species (from four genera belonging to four subfamilies) collected between the years 2008 and 2016: some of which were previously known to host *Hamiltonella* (Meseguer et al. 2017). Specimens were kept in 70% ethanol at 6 °C. For whole-genome sequencing, we prepared DNA samples enriched with bacteria following a slightly modified version of the protocol by Charles and Ishikawa (1999) as described in Jouselin et al. (2016). For this filtration protocol, ~15 aphids from one colony were pooled together. Extracted DNA was used to prepare custom paired-end libraries in Genoscope as in Manzano-Marín et al. (2018). These libraries were sequenced using either 151- or 251-bp paired-end reads chemistry on a HiSeq2500 Illumina sequencer. For full details on specimen collection, species identification, and accession numbers for the samples from which APSE phages were extracted, see [supplementary table S1, Supplementary Material](#) online.

Hamiltonella and APSE Genome Assembly and Annotation

First, we scanned the aforementioned *Hamiltonella* assembled genomes using the sequence of the phage attachment site (also known as attP) and the last 62 bp with a 90% identity threshold in UGENE v1.29.0 (Okonechnikov et al. 2012). For noncircularized genomes, we first scanned the contigs using BlastN (Altschul 1997) for these sequences in order to identify the contig or scaffold where the putative APSE phage resided in. Illumina sequences from the eight newly sampled aphid species were first right-tail clipped (requiring a minimum quality threshold of 20 and a minimum length of

75 bp) using FASTX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/, last accessed March 4, 2020). Additionally, PRINSEQ v0.20.4 (Schmieder and Edwards 2011) was used to remove reads containing undefined nucleotides as well as those left without a pair after the filtering and clipping process. The resulting reads were assembled using SPAdes v3.11.1 (Bankevich et al. 2012) with the `–only-assembler` option and k-mer sizes of 33, 55, 77, 99, and 127. From the resulting contigs, those that were shorter than 200 bp were dropped. The remaining contigs were binned using results from a BlastX (Altschul 1997) search (best hit per contig) against a database consisting of the Pea aphid's proteome and a selection of aphid's symbiotic bacteria proteomes ([supplementary table S2, Supplementary Material](#) online). When no genome was available for a specific bacterial lineage, closely related bacteria were used. The assigned contigs were manually screened using the BlastX web server (searching against the nr database) to ensure correct assignment. This binning process confirmed the presence of *Buchnera*, *Hamiltonella*, and its corresponding APSE phage. APSE was confidently assigned to *Hamiltonella* by checking for chimeric APSE–*Hamiltonella* chromosome contig ends in the corresponding insertion sites and/or by checking that no other facultative symbionts were sequenced in the sample. The resulting contigs were then used as reference for read mapping and individual genome assembly using SPAdes, as described above. *Hamiltonella* draft assemblies included APSE-assigned reads. APSE individual assemblies were visually screened for inconsistencies using Tablet (Milne et al. 2013).

The resulting APSE genomes underwent a draft annotation using Prokka 1.14.4 (Seemann 2014). This was followed by noncoding RNA prediction using infernal v1.1.2 (Nawrocki and Eddy 2013) (against the Rfam v14.1 database, Kalvari, Argasinska, et al. 2018; Kalvari, Nawrocki, et al. 2018) and tRNAscan-SE v2.0.5 (Chan et al. 2019). We then performed manual curation of the annotations on UGENE (Okonechnikov et al. 2012) through online BlastX searches of the intergenic regions, open reading frame (ORF) finding feature in UGENE, and through DELTA-BLAST (Boratyn et al. 2012) searches of the predicted ORFs against NCBI's nr database and the InterProScan v5 web server (Jones et al. 2014; Mitchell et al. 2019). SignalP v5.0 (Almagro Armenteros et al. 2019) and Phobius v1.01 (Kall et al. 2005) were used to predict signal peptides. ORFs were considered to be putative functional proteins (and thus not pseudogenes) if seemingly essential domains for the function were found or if the ORFs displayed truncations but retained identifiable domains. Short ORFs (≤ 300) that were conserved but consistently pseudogenized in all but one APSE genome were annotated as miscellaneous features. The origin of replication was determined using originX (Worning et al. 2006). Once the first APSE genome was curated, we used these proteins as a “genus”-specific database in Prokka to annotate the subsequent genomes and iteratively added novel proteins to the

database. Codon usage was calculated separately for closed high-quality *Hamiltonella*–APSE pairs using the Cousin web server (Bourret et al. 2019).

Finally, phage-like regions were searched in *Arsenophonus* spp. genomes using the web server for PHASTER (Zhou et al. 2011; Arndt et al. 2016). The retrieved phage-like regions were then manually inspected in UGENE, and identity of genes versus those of APSE was assessed using the online web server of BLAST. The EMBOSS suite program primer-search (Rice et al. 2000) was used to predict possible amplification targets in the *Arsenophonus* genomes using the primers reported in Duron (2014).

Phylogenetic and Recombination Analyses

In order to reconstruct the phylogeny of *Hamiltonella* endosymbionts, we used seven gene sequences (*accD*, *dnaA*, *gyrB*, *hrpA*, *murE*, *ptsI*, and *recI*) following Degnan and Moran (2008). Genes sequences were gathered from the Supplementary data for Manzano-Marín et al. (2020) (<https://doi.org/10.5281/zenodo.2566355>, last accessed March 4, 2020) then used to identify orthologs in newly sequenced *Hamiltonella* strains using BlastX. Due to the low coverage for *Hamiltonella* draft genomes for two samples (3702 and 3692), we were unable to recover sequences for all seven genes used for phylogenetic inference, and thus these specimens were excluded from the *Hamiltonella* phylogenetic analysis. For each gene, sequence alignments were conducted using MUSCLE v.3.8.31 (Edgar 2004) and visually checked using SeaView v4 (Gouy et al. 2010). We then removed divergent and ambiguously aligned blocks using Gblocks v0.91b (Talavera and Castresana 2007). A partitioned scheme was selected using PartitionFinder v2.1.1 (Lanfear et al. 2017) with one data block defined for each codon position (first, second, and third) in each gene. The best model of evolution of each partition was selected among those available in MrBayes, using the Bayesian Information Criterion metric under a *greedy* algorithm. Finally, we concatenated the resulting alignments and ran a Bayesian phylogenetic inference using the GTR+I+G model in MrBayes v3.2.7 (Ronquist et al. 2012) running two independent analyses with four chains each for 1,500,000 generations and checked for convergence.

For performing both phylogenetic inferences and analyzing the genetic differences across APSE, we first ran an orthologous protein clustering analysis using OrthoMCL v2.0.9 (Li 2003; Chen et al. 2007) on the full predicted proteome for the phage genomes (supplementary table S3, Supplementary Material online). We then extracted the single-copy core proteins for phylogenetic reconstruction (30 protein groups). Sequences were aligned using MAFFT v7.450 (maxiterate 1,000 localpair) (Katoh and Standley 2013). Divergent and ambiguously aligned blocks were removed using Gblocks. Substitution model was selected using ModelTest-NG v0.1.6

(Darriba et al. 2020) using the Akaike information criterion criterium. Bayesian phylogenetic reconstruction was conducted with MrBayes as described above by running chains for 300,000 generations and checking for convergence. Given prior evidence for intragenic recombination in APSE proteins (Degnan and Moran 2007), we constructed a second data set removing all genes where recombination has putatively occurred. Recombination was tested for running PhiPack (Bruen et al. 2006) on the aligned proteins and results of these analyses, for the putative nonrecombining proteins based on Φ_{ω} , can be found in supplementary table S4, Supplementary Material online. Bayesian phylogenetic reconstruction was ran as described above.

To infer the relationships of the 14 toxin-cassette- and lytic-region-proteins (lysozyme and holins) in APSE genomes, we searched for similar sequences using the online BlastP web server versus the NCBI's nr database and selected the top 50 nonredundant hits. These were filtered for a minimum of 70% query coverage and a 35% sequence identity. We aligned each of them using MAFFT and manually removed divergent and ambiguously aligned blocks. For each gene, model selection was done with ModelTest-NG. Bayesian phylogenetic inference was run for 300,000 generations in MrBayes as described above.

All resulting trees were visualized and exported with FigTree v1.4.1 (<http://tree.bio.ed.ac.uk/software/figtree/>, last accessed March 4, 2020). All files used for phylogenetic analyses as well as PHASTER phage search results and comparative genomics files used in this study are available in <https://doi.org/10.5281/zenodo.3764739> (last accessed March 5, 2020).

Results and Discussion

APSE Phage Genomes

We successfully de novo assembled genomes for eight complete APSE phages from eight different aphid species and extracted an additional eight from both closed and draft *Hamiltonella* endosymbionts from the pea aphid *A. pisum* and the whitefly *B. tabaci* (table 2). Additionally, we recovered draft genomes for all eight *Hamiltonella* strains hosting the newly sequenced APSE phages. All recovered phage genomes have a genome size between 33,476 and 41,343 bp. They have a very conserved G+C content ranging between 42.81% and 45.04% and code for an average of 38 coding sequences (CDSs) with 3 to no pseudogenes (table 2). Most APSE phages preserve an intact tRNA-Lys-(UC)UU, which has been lost in 5D and ZA17, both of which still keep an internally degraded tRNA pseudogene. An analysis of 17 bacterial hosts and 37 phage genomes found that the presence of tRNAs in these phages tended to correspond to codons that were highly used by the phage's protein coding genes but were rare in the host's (Baillly-Bechet et al. 2007). However, we found virtually no difference in the lysin codon usage

Table 2
Genome Assembly and Annotation Statistics

Insect Host	<i>Hamiltonella</i>			APSE			
	Strain	Coverage	Genome Size	Coverage	G + C Content	CDSs(ψ)	ncRNAs
<i>Bemisia tabaci</i>	MEAM1 ^a	NA	38,988	NA	43.36	39(2)	4
<i>Bemisia tabaci</i>	MED-Q1 ^a	NA	38,949	NA	43.34	39(2)	4
<i>Cinara confinis</i>	2801 ^b	178 \times	38,949	1,638 \times	43.94	38(2)	3
<i>Cinara</i> sp. 3046	3046 ^b	5 \times	37,531	38 \times	43.96	38(2)	3
<i>Cinara pinimaritimae</i>	2836 ^b	27 \times	35,191	597 \times	43.84	38(0)	3
<i>Cinara cuneomaculata</i>	2628 ^b	777 \times	36,094	6,498 \times	43.88	37(1)	3
<i>Drepanosiphum platanoidis</i>	3702 ^b	3 \times	36,931	111 \times	43.86	38(2)	3
<i>Eriosoma grossulariae</i>	3692 ^b	3 \times	38,420	38 \times	42.90	42(3)	3
<i>Acyrtosiphon pisum</i>	5AT ^a	28.7 \times	39,867	NA	42.91	41(2)	3
<i>Acyrtosiphon pisum</i>	NY26 ^a	643 \times	39,887	NA	42.86	41(2)	3
<i>Acyrtosiphon pisum</i>	5D ^a	NA	39,146	NA	42.81	41(2)	3
<i>Acyrtosiphon pisum</i>	ZA17 ^a	567 \times	39,145	NA	42.81	41(2)	3
<i>Acyrtosiphon pisum</i>	APSE1 ^c	NA	36,524	NA	43.89	37(2)	3
<i>Acyrtosiphon pisum</i>	MI47 ^a	NA	36,522	NA	43.88	37(2)	3
<i>Periphyllus testudinaceus</i>	2671 ^b	8 \times	41,343	57 \times	44.37	38(0)	3
<i>Acyrtosiphon pisum</i>	AS3 ^a	535 \times	38,992	NA	45.04	36(0)	3
<i>Cinara watanabei</i>	3293 ^b	194 \times	33,476	25,300 \times	44.46	34(2)	3

Note.—Assembly and genome statistics for complete APSE genomes and their *Hamiltonella* hosts. NA, not available.

^aExtracted from publicly available *Hamiltonella* genomes.

^bSequenced in this study.

^cPhage isolated from an unidentified secondary symbiont of *Acyrtosiphon pisum* (van der Wilk et al. 1999).

between pairs of *Hamiltonella*–APSE genomes (supplementary table S5, Supplementary Material online). Nonetheless, we found that the *Hamiltonella* strain 5AT only codes for a tRNA-Lys-CUU, following the pseudogenization of the conserved tRNA-Lys-UUU, which recognizes the most commonly used codon for lysine. Conversely, its associated APSE preserves its own tRNA-Lys-UUU. This suggests that the presence of a tRNA-Lys-UUU in the APSE phage could drive the loss of this tRNA in the bacterial host, which could in turn provide a selective advantage to carrying this phage.

All APSE have two noncoding RNAs (ncRNAs) that code for small bacterial RNAs (sRNA): a c4 antisense family sRNA and an *IsrK* Hfq binding family sRNA. In the case of the former, a c4 sRNA has been identified in P1 and P7 phages from *Escherichia coli* as a regulator of the *ant* gene, an antirepressor, through the binding to complementary regions in the upstream region of this gene (Citron and Schuster 1990). Through the analysis of the putative so-called *a'* and *b'* antisense regions of the c4 sRNA, we found that they consistently matched the 3'-end region of the distant DNA polymerase gene and the upstream noncoding region of the adjacent BRO family, N-terminal domain-containing protein. Three BRO family proteins (BRO-A, BRO-C, and BRO-D) from the *Bombyx mori nucleopolyhedrovirus* (BmNPV) have been suggested to influence host DNA replication and/or transcription, following the finding that they are able to bind DNA from the host with a stronger affinity for single-stranded than for double-stranded DNA (Zemskov et al. 2000). Regarding

IsrK, its expression pattern in *Salmonella typhimurium* suggested that this sRNA might be involved in the regulation of virulence (Padalon-Brauch et al. 2008). Lastly, both APSE phages harbored by *Hamiltonella* infecting *B. tabaci* MEAM1 and MED-Q1 biotypes code for an additional sRNA-Xcc1 family sRNA upstream of a protein of unknown function (ORF6N domain-containing protein) also unique to these APSE. In *Xanthomonas campestris* pv. *vesicatoria*, sRNA-Xcc1 is under the positive control of two important virulence regulators, suggesting that they might play a role in pathogenesis (Chen et al. 2011). All this evidence points toward an important regulatory function of these sRNAs in APSE and highlight them as targets to study their downstream regulatory effects in both APSE and *Hamiltonella* gene expression.

APSE Gene Content Variation and Toxin Cassettes

APSE genomes were highly conserved in both gene content, order, and sequence identity (supplementary fig. S1, table S3, and file S3, Supplementary Material online). The most variable genes, in terms of sequence identity, across APSE were those of a BRO family, N-terminal domain-containing protein (orthogroup APSEcp_004), the DNA polymerase (APSEcp_008), the lysozyme (APSElp_003), the small subunit of the terminase (APSEcp_020), and the tail proteins (needle, spike, and fiber assembly; APSEcp_028 and APSEcp_033–34). Regarding the DNA polymerase gene, we observed differential loss of the intein domain (a protein intron). The lysozyme

gene was by far the one that showed the most sequence divergence. A phylogenetic analysis suggested two well supported lysozyme phylogroups (so-called “P13” and “F,” supplementary fig. S2, Supplementary Material online), consistent with previous findings (Degnan and Moran 2008). From our analyses, “P13” is restricted to APSE-7 and APSE-1 (including the APSE-1 variants APSE-4 and APSE-5).

One marked difference in CDS content between APSE harbored by *Hamiltonella* symbionts of aphids and those harbored by the whitefly *B. tabaci* was in the protein flanked by the BRO family, N-terminal domain-containing protein and the putative P-loop NTPase. Although all APSE from aphid-infecting *Hamiltonella* code for a putative phage regulatory protein (Rha family), those infecting the whitefly code for an AntA/AntB antirepressor family protein. To our knowledge, the function of both of these proteins is not well understood, and thus the significance of this difference remains to be studied.

Altogether, we identified two main regions of genomic variation in APSE phages and propose an APSE classification based on the composition of these two variable sequence stretches based on Degnan and Moran (2008), and keeping the main type numbering (X), with a slight revision (X.Yz) based on our findings (fig. 1): The first variable region (X) is the toxin cassette, with small variations within the types (Y), and the second one is found around the vicinity of the DNA polymerase gene (z). In regards to the toxin cassettes, we found five different “types” or “gene sets” represented in the newly assembled APSE genomes. We found seven phages that could be classified within APSE-7. These were associated with *Hamiltonella* from two whiteflies and five species of aphids (from two subfamilies). We found that although the CdtB family protein was conserved across APSE-7, its companion hypothetical protein was not. Both proteins in the cassette showed identifiable signal peptides, suggesting their export. The CdtB family protein is also present in APSE-2; however, the structure of the cassette is unlike that of APSE-7. It codes for an additional two putative toxin genes, and in the case of APSE-2.2 subtypes, two flanking hypothetical proteins, all with signal peptides. In the case of APSE-2.1a, we found that the hypothetical proteins flanking CdtB were missing. Nonetheless, we found a different hypothetical protein (containing a signal peptide) flanked by insertion sequence pseudogenes. This suggests that, at least in this case, the new member of the toxin cassette was likely mobilized by the action of transposable elements. As in a previous study by Degnan and Moran (2008), we found that all APSE-2 and -7 types had both holin family genes (λ and HP1) followed by a different lysozyme gene phylotype, so-called “F” and “P13,” respectively. The APSE-1 type was represented by two nearly identical APSE genomes, both associated with *Hamiltonella* strains from *A. pisum*. Both presented the same four hypothetical proteins, including two previously identified as putative subunits of a shiga-like toxin (van der

Wilk et al. 1999; Degnan and Moran 2008). APSE-3 types were found to be associated with *Hamiltonella* hosted by two distantly related aphid species, *A. pisum* and *Periphyllus testudinaceus*. They showed a very conserved cassette, encoding for a hypothetical protein, with no identifiable signal peptide, and an RHS-repeat putative toxin (previously referred to as “YD-repeat-containing” by Degnan and Moran [2008]) with a signal peptide. Finally, the cassette found in 3293 (*Hamiltonella* from *Cinara watanabei*) represents a completely novel type, which we classified as APSE-8. This toxin cassette encodes for a single putative virulence-associated protein with no recognizable signal peptide. Previous studies have classified the APSE-2 variant found in *Hamiltonella* strain ZA17 as APSE-8 (Brandt et al. 2017; Doremus and Oliver 2017; Chevignon et al. 2018; Patel et al. 2019). However, we found that the only difference in its toxin cassette was the pseudogenization of the first putative toxin gene within this region when compared with APSE-2 5AT and NY26. Therefore, we have considered it to be a subtype of APSE-2, in agreement with Martinez et al. (2014).

Regarding the second variable region, the main differences laid in the presence/absence of the Kila-N domain-containing protein and its companion hypothetical protein, a putative transcriptional regulator upstream of the DNA polymerase gene. How this changes impact the phage fitness and phenotype remains to be investigated.

Although most genes in the cassette had only very distant or no affiliations to genes in the databases, both the RHS-repeat and CdtB toxins had many close hits in the NCBI’s nr database (see Materials and Methods). In both cases, they formed a well-supported monophyletic group (fig. 2). As aforementioned, the RHS-repeat toxin is, to the best of our knowledge, restricted to APSE-3 types. On the contrary, the CdtB protein is actually present in several APSE phages, namely APSE-7 and APSE-2. The CdtB proteins encoded in the two different APSE types do not form distinct clades, rather the proteins of APSE-2 are nested within CdtB from APSE-7 (fig. 2B). Due to previous evidence for intragenic recombination (Degnan and Moran 2007), we ran a recombination test using PhiPack (Bruen et al. 2006) and found no significant evidence for recombination (Φ_{ω} , P value = 7.02e-02). However, NSS- and Max χ^2 -tests did find significant recombination among CdtB genes (P value = 4.00e-03 and P value = 2.00e-03, respectively), and thus, the relationships among APSE-encoded CdtB proteins should be interpreted with caution.

In both cases, these proteins were closely related to CdtB encoded by *Arsenophonus* spp. hosted by two whitefly species: *Aleurodicus floccissimus* and *B. tabaci*. Additionally, another four proteins in the toxin cassettes had high similarity and a close phylogenetic relation to proteins in *Arsenophonus* spp. (orthogroups APSEtp_002, APSEtp_007, APSEtp_013, and APSEtp_014). Previous analyses have identified APSE-like DNA polymerase genes in *Arsenophonus* symbionts of whiteflies, aphids, louse flies, bat flies, and psyllids (Hansen

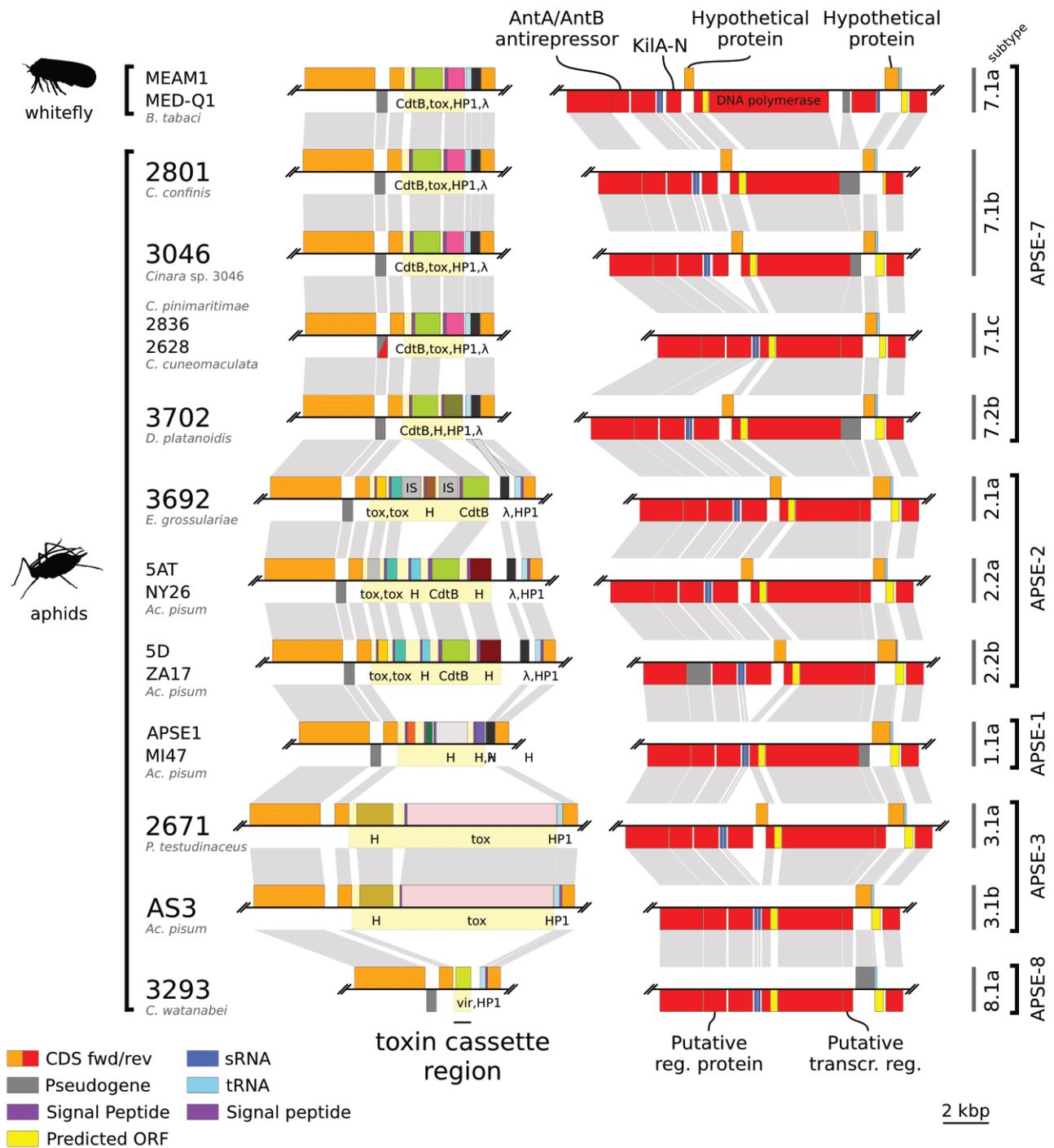


Fig. 1—Variable regions in APSE genomes. Genome synteny plots of the two variable regions across APSE phages. On the bottom left, a color key for the features shown in the plot. On the right, proposed classification of APSE in types and subtypes based on the variable regions composition.

et al. 2007; Duron 2014). The diagnostic polymerase chain reactions used in these studies targeted the so-called P45 (DNA polymerase) and P3 (virulence-associated protein E family protein). Analysis of three available high-quality *Arsenophonus* genomes (*Arsenophonus nasoniae* strain FIN [CP038613.1CP038621.1], *Arsenophonus* endosymbiont of

Nilaparvata lugens strain Hangzhou [JRLH01000001.1], and *Arsenophonus* endosymbiont of *Aleurodicus floccissimus* [OUNDO1000000.1]), revealed that these reported hits actually corresponded to prophage regions in their genomes (PHASTER annotations available in <https://doi.org/10.5281/zenodo.3764739>, last accessed March 5, 2020). On closer

Downloaded from <https://academic.oup.com/gbe/article/12/6/878/5835279> by guest on 26 November 2021

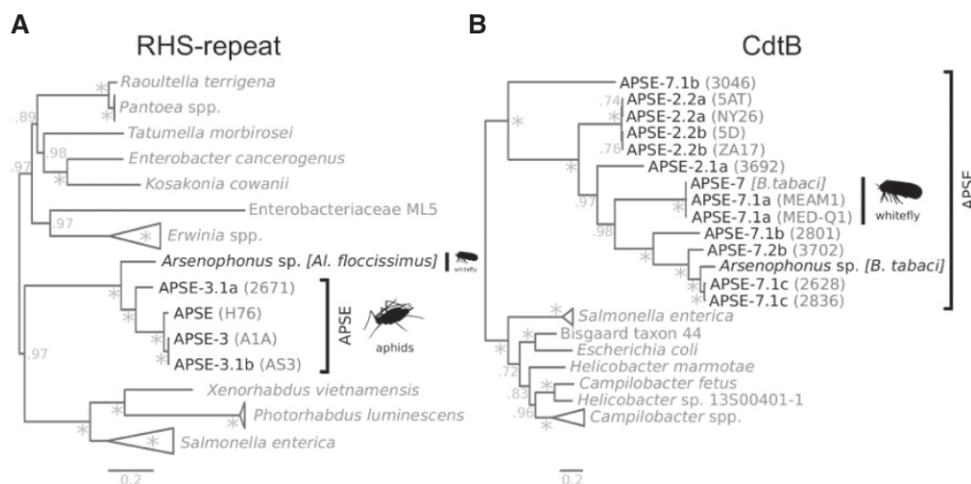


Fig. 2—APSE toxin gene phylogenies. Bayesian phylogeny of (A) RHS-repeat toxins and (B) CdtB proteins. Numbers at nodes represent Bayesian posterior probabilities. “*” = 1. Phylogenies were midpoint rooted.

inspection, we found that only one region in the chromosome of *Arsenophonus nasoniae* strain FIN contained both of these proteins. This region, as did others, preserved high sequence identity to one or a few APSE proteins. However, the regions were not in fact APSE phages: that is, they consistently showed a different gene order, insertion of proteins not characteristic of fully sequenced APSE phages, and/or pseudogenized proteins that are otherwise conserved across APSE analyzed in this study (including the DNA polymerase). Although there are currently no full genome sequences for the strains analyzed by Hansen et al. (2007) or Duron (2014), this genomic scan, along with the presence of putative toxins in *Arsenophonus* spp. closely related to those present in APSE, strongly suggests that *Arsenophonus* bacteria have indeed been historically infected by close relatives of the *Hamiltonella*–associated APSE phages, and that these phages have left an imprint in the bacterial genomes. The retention of RHS-repeat- and CdtB-APSE-like toxin genes in at least two different strains of *Arsenophonus* suggests that these proteins could in fact endow the *Arsenophonus* symbiont with a protective phenotype. This could in turn provide a positive “APSE-free” fitness effect of carrying the endosymbiont when confronted with the environmental pressure of parasitoid infection. So far *Arsenophonus*, while quite widespread and very diversified in insects (Nováková et al. 2009), has not been credited with conferring protection against parasitoids of its insect hosts.

Phylogenetic History of *Hamiltonella* and APSE

Phylogenetic analyses supported previous observations (Degnan and Moran 2007) that closely related *Hamiltonella* strains were harbored by distantly related aphid species (i.e., aphids belonging to distinct subfamilies) and hosted different APSE types (fig. 3A). A noteworthy case is that of *Hamiltonella* strains ZA17 and 5D: They belong to two

different clades but host nearly identical APSE phages. Conversely, the closely related *Hamiltonella* strain 5D and MI47 host rather different APSE types. APSE-7 is notably present in both *Hamiltonella* clade C and in the distantly related lineages of this symbiont hosted by the whitefly *B. tabaci*. This provides strong evidence and supports previous findings, based on single-gene phylogenies (Degnan and Moran 2007, 2008), of horizontal transfer of these viral entities across *Hamiltonella* strains and lineages. The horizontal transfer of phages across lineages of endosymbiotic bacteria has been suggested for *Wolbachia*, where horizontal transfers of the WO phage seem to occur across both related and divergent *Wolbachia* (Masui et al. 2000; Bordenstein 2004). This horizontal transfer could be facilitated by the coexistence of two divergent *Hamiltonella* lineages within the same host, similarly to what has been observed for *Wolbachia* (Sasaki et al. 2005; Raychoudhury and Werren 2012). Indeed, multiple symbiont species and strains of the same species can coexist within the same population of an aphid species (Sandström et al. 2001; Tsuchida et al. 2002; Haynes et al. 2003; Russell et al. 2013; Meseguer et al. 2017), offering an opportunity for coinfection and transfer or recombination of their phages. Through the reconstruction of a phylogenetic network using the concatenated single-copy shared proteins of APSE, we observe evidence for recombination across APSE types (fig. 3B). This is a feature that has previously been observed for these phages (Degnan and Moran 2007) and is a general feature of phage genomes (reviewed by Dion et al. [2020]). In fact, only 18 out of 30 single-copy shared orthologous groups of proteins showed no significant evidence of intra-genic recombination based on a Φ_{ω} test (supplementary table S4, Supplementary Material online). From these 18, nine showed significant evidence of recombination as judged by two additional tests implemented in PhiPack: NSS and Max χ^2 . A concatenated nucleotide alignment of the nine putative

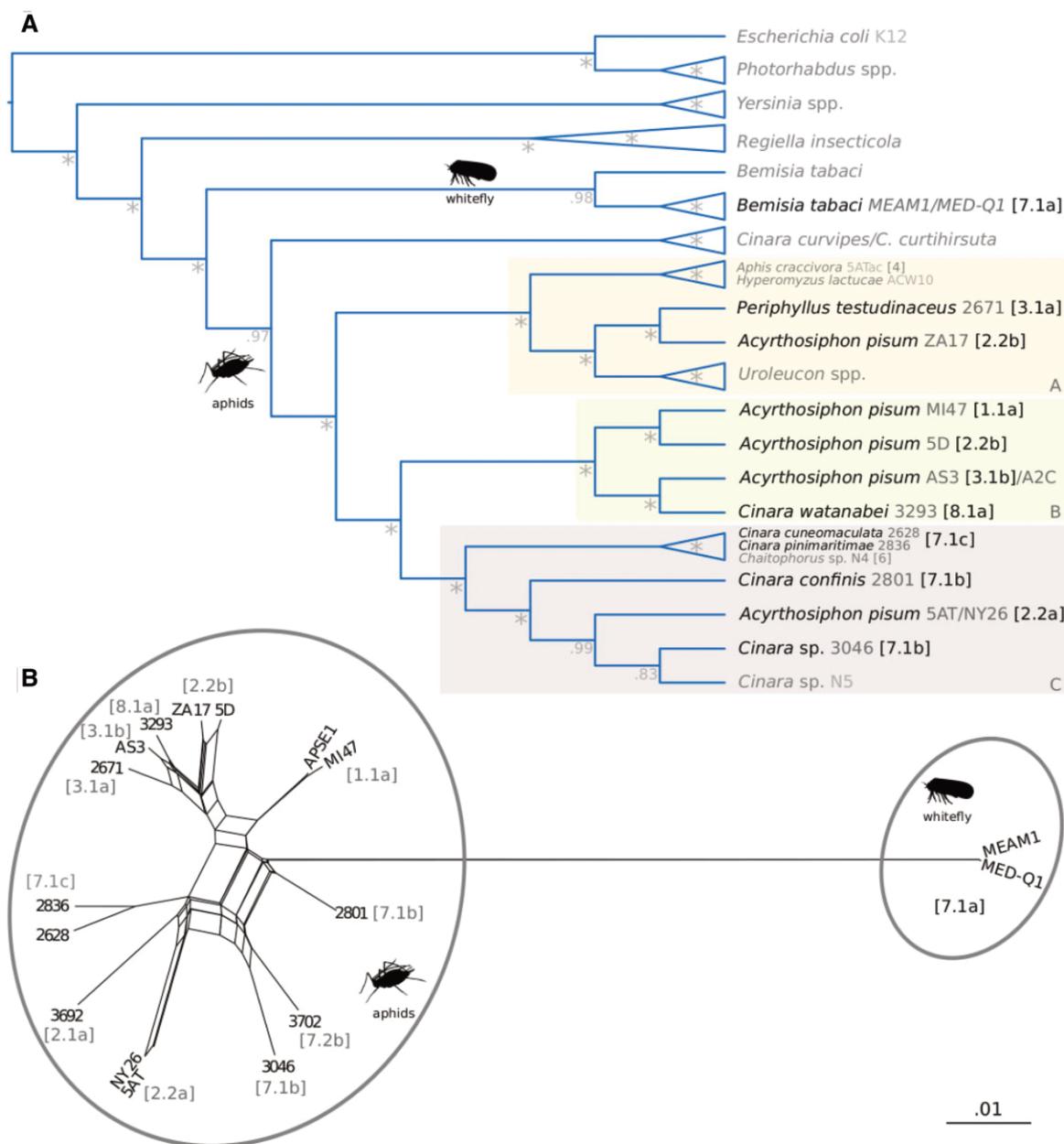


Fig. 3—Phylogenetic relations of *Hamiltonella* and APSE phages. (A) Dendrogram of phylogenetic relationships among *Hamiltonella* endosymbionts based on single-copy shared proteins. Numbers at nodes represent Bayesian posterior probabilities. “*” = 1. (B) Phylogenetic network of single-copy shared proteins as calculated by SplitsTree.

nonrecombinant genes revealed significant intergenomic recombination ($\Phi_{\omega}P$ value = 0.00e + 00). These results suggest genome-wide recombination for APSE phages, which is expected to complicate the inference of phylogenetic relationships. Regardless, Bayesian phylogenies were reconstructed for the three sets of single-copy shared proteins and can be found in [supplementary figure S3](#), [Supplementary Material](#) online. Given the rampant recombination of their genomes, it is therefore hard to conclude on the phylogenetic history of these viruses.

One common feature of all phylogenetic analyses was that the APSE phages infecting *B. tabaci*-associated *Hamiltonella* strains were consistently highly divergent from those of aphid-associated ones. Additionally, the SplitsTree network reconstruction suggests little recombination (relative to that observed among aphid-associated APSE) between these two groups of APSE infecting different hosts (aphids and *B. tabaci*). This is congruent with the *Hamiltonella* symbiont phylogeny which positions those infecting the whitefly *B. tabaci* as a separate sister clade to that made up of all

currently available aphid-hosted ones (fig. 3A). Previous experimental studies of *Hamiltonella* infections in *Sitobion avenae* aphids have shown that strains of this symbiont establish infections more easily when transferred from the same host species (Łukasik et al. 2015). Additionally, these infections were found to be more stable. The same study also found that infection success was higher when the transferred symbiont strain was more closely related to the one that was originally present in the host. All this strongly suggests that these two lineages infecting different host groups can follow separate evolutionary histories, possibly driven by both ecological factors and other barriers limiting the horizontal transfer of the two distinct symbiont lineages.

Conclusion

Facultative symbionts in insects can confer conditional fitness benefits to their hosts (Polin et al. 2015; Cass et al. 2016; Doremus et al. 2018), and *Hamiltonella* and its APSE-conferred defense against parasitoid wasps is one of the best studied cases (Oliver and Higashi 2019). The results presented in this study support previous studies reporting a variable toxin cassette across APSE as well as intragenic recombination. Additionally, the whole-genome sequencing, annotation, and phylogenetic analyses of three new APSE types conducted here have revealed important features of these protective entities: 1) their genomes are highly conserved in terms of gene content with most of the variation localized to the toxin cassette; 2) alterations can occur within a cassette, with pseudogenization events affecting the putative toxin genes (i.e., APSE-2.2a) or IS-mediated insertion of novel genes; 3) APSE harbored by whitefly-associated *Hamiltonella* seem to have undergone a separate evolutionary history, in line with their hosts; and 4) *Arsenophonus* spp. might have historically had association with APSE-like phages, with some of these having left an imprint in their genomes. This study provides the first genome-wide overview of the diversity of five APSE phage types hosted by *Hamiltonella* endosymbionts, improving our understanding of these phages. The features highlighted in this study will be of great value in the design and interpretation of future experiments of APSE-mediated *Hamiltonella* parasitoid defense.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was supported by a PhD grant from the INRAE ECOFA department (to J.R.), the Marie-Curie AgreeSkills+ fellowship program cofunded by the EU's Seventh Framework Programme (FP7-609398 to A.M.-M.), the Marie

Sklódowska-Curie fellowship program *H2020-MSCA-IF-2016* (MicroPhan: 746189 to E.J.), and the *France Génomique* National Infrastructure, funded as part of the *Investissement d'Avenir* program managed by the Agence Nationale pour la Recherche (ANR-10-INBS-09 to C.C.). This publication has been written with the support of the AgreeSkills+ fellowship program which has received funding from the EU's Seventh Framework Programme (Grant Agreement No. FP7-609398; AgreeSkills+ contract). We are grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) for providing help and/or computing and/or storage resources. The authors are grateful to the CBGP-HPC computational platform. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Literature Cited

- Almagro Armenteros JJ, et al. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* 37(4):420–423.
- Altschul S. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25(17):3389–3402.
- Arndt D, et al. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44(W1):W16–W21.
- Asplen MK, et al. 2014. Specialisation of bacterial endosymbionts that protect aphids from parasitoids. *Ecol Entomol.* 39(6):736–739.
- Bailly-Bechet M, Vergassola M, Rocha E. 2007. Causes for the intriguing presence of tRNAs in phages. *Genome Res.* 17(10):1486–1495.
- Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19(5):455–477.
- Boratyn GM, et al. 2012. Domain enhanced lookup time accelerated BLAST. *Biol Direct* 7(1):12.
- Bordenstein SR. 2004. Bacteriophage flux in endosymbionts (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. *Mol Biol Evol.* 21(10):1981–1991.
- Bourret J, Alizon S, Bravo IG. 2019. COUSIN (COdon Usage Similarity INdex): a normalized measure of codon usage preferences. *Genome Biol Evol.* 11(12):3523–3528.
- Brandt JW, Chevignon G, Oliver KM, Strand MR. 2017. Culture of an aphid heritable symbiont demonstrates its direct role in defence against parasitoids. *Proc R Soc B Biol Sci.* 284(1866):20171925.
- Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* 172(4):2665–2681.
- Cass BN, et al. 2016. Conditional fitness benefits of the Rickettsia bacterial symbiont in an insect pest. *Oecologia* 180(1):169–179.
- Chan PP, Lin BY, Mak AJ, Lowe TM. 2019. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *bioRxiv*:614032.
- Charles H, Ishikawa H. 1999. Physical and genetic map of the genome of *Buchnera*, the primary endosymbiont of the pea aphid *Acyrtosiphon pisum*. *J Mol Evol.* 48(2):142–150.
- Chen F, Mackey AJ, Vermunt JK, Roos DS. 2007. Assessing performance of orthology detection strategies applied to eukaryotic genomes. *PLoS One* 2(4):e383.
- Chen XL, et al. 2011. sRNA-Xcc1, an integron-encoded transposon- and plasmid-transferred trans-acting sRNA, is under the positive control of

- the key virulence regulators HrpG and HrpX of *Xanthomonas campestris* pathovar *campestris*. *RNA Biol.* 8(6):947–953.
- Chevignon G, Boyd BM, Brandt JW, Oliver KM, Strand MR. 2018. Culture-facilitated comparative genomics of the facultative symbiont *Hamiltonella defensa*. *Genome Biol Evol.* 10(3):786–802.
- Citron M, Schuster H. 1990. The c4 repressors of bacteriophages P1 and P7 are antisense RNAs. *Cell* 62(3):591–598.
- Darriba D, et al. 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol Biol Evol.* 37(1):291–294.
- Degnan PH, Moran NA. 2007. Evolutionary genetics of a defensive facultative symbiont of insects: exchange of toxin-encoding bacteriophage. *Mol Ecol.* 17(3):916–929.
- Degnan PH, Moran NA. 2008. Diverse phage-encoded toxins in a protective insect endosymbiont. *Appl Environ Microbiol.* 74(21):6782–6791.
- Dennis AB, Patel V, Oliver KM, Vorburger C. 2017. Parasitoid gene expression changes after adaptation to symbiont-protected hosts. *Evolution* 71(11):2599–2617.
- Dion MB, Oechslein F, Moineau S. 2020. Phage diversity, genomics and phylogeny. *Nat Rev Microbiol.* 18(3):125–138.
- Doremus MR, Oliver KM. 2017. Aphid heritable symbiont exploits defensive mutualism. *Appl Environ Microbiol.* 83(8):e03276–16.
- Doremus MR, et al. 2018. Breakdown of a defensive symbiosis, but not endogenous defences, at elevated temperatures. *Mol Ecol.* 27(8):2138–2151.
- Duron O. 2014. *Arsenophonus* insect symbionts are commonly infected with APSE, a bacteriophage involved in protective symbiosis. *FEMS Microbiol Ecol.* 90(1):184–194.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32(5):1792–1797.
- Gouy M, Guindon S, Gascuel O. 2010. SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol.* 27(2):221–224.
- Guo J, et al. 2017. Nine facultative endosymbionts in aphids. A review. *J Asia-Pac Entomol.* 20(3):794–801.
- Hansen AK, Jeong G, Paine TD, Stouthamer R. 2007. Frequency of secondary symbiont infection in an invasive psyllid relates to parasitism pressure on a geographic scale in California. *Appl Environ Microbiol.* 73(23):7531–7535.
- Haynes S, et al. 2003. Diversity of bacteria associated with natural aphid populations. *Appl Environ Microbiol.* 69(12):7216–7223.
- Hopper KR, et al. 2018. The defensive aphid symbiont *Hamiltonella defensa* affects host quality differently for *Aphelinus glycinis* versus *Aphelinus atriplicis*. *Biol Control* 116:3–9.
- Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9):1236–1240.
- Jousselin E, et al. 2016. Assessment of a 16S rRNA amplicon Illumina sequencing procedure for studying the microbiome of a symbiont-rich aphid genus. *Mol Ecol Resour.* 16(3):628–640.
- Kall L, Krogh A, Sonnhammer E. 2005. An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics* 21(Suppl 1):i251–i257.
- Kalvari I, Argasinska J, et al. 2018. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* 46(D1):D335–D342.
- Kalvari I, Nawrocki EP, et al. 2018. Non-coding RNA analysis using the Rfam database. *Curr Protoc Bioinf.* 62(1):e51.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.
- Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol.* 34(3):772–773.
- Lenhart PA, White JA. 2017. A defensive endosymbiont fails to protect aphids against the parasitoid community present in the field. *Ecol Entomol.* 42(5):680–684.
- Leybourne DJ, Bos JIB, Valentine TA, Karley AJ. 2020. The price of protection: a defensive endosymbiont impairs nymph growth in the bird cherry-yat aphid, *Rhopalosiphum padi*. *Insect Sci.* 27(1):69–85.
- Li L. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13(9):2178–2189.
- Łukasik P, et al. 2015. Horizontal transfer of facultative endosymbionts is limited by host relatedness. *Evolution* 69(10):2757–2766.
- Manzano-Marín A, et al. 2018. A freeloader? The highly eroded yet large genome of the *Serratia symbiotica* symbiont of *Cinara strobilifera*. *Genome Biol Evol.* 10(9):2178–2189.
- Manzano-Marín A, et al. 2020. Serial horizontal transfer of vitamin-biosynthetic genes enables the establishment of new nutritional symbionts in aphids' di-symbiotic systems. *ISME J.* 14(1):259–273.
- Martinez AJ, Weldon SR, Oliver KM. 2014. Effects of parasitism on aphid nutritional and protective symbioses. *Mol Ecol.* 23(6):1594–1607.
- Masui S, Kamoda S, Sasaki T, Ishikawa H. 2000. Distribution and evolution of bacteriophage WO in *Wolbachia*, the endosymbiont causing sexual alterations in arthropods. *J Mol Evol.* 51(5):491–497.
- Meseguer AS, et al. 2017. *Buchnera* has changed flatmate but the repeated replacement of co-obligate symbionts is not associated with the ecological expansions of their aphid hosts. *Mol Ecol.* 26(8):2363–2378.
- Milne I, et al. 2013. Using Tablet for visual exploration of second-generation sequencing data. *Briefings Bioinf.* 14(2):193–202.
- Mitchell AL, et al. 2019. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* 47(D1):D351–D360.
- Moran NA, Degnan PH, Santos SR, Dunbar HE, Ochman H. 2005. The players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. *Proc Natl Acad Sci U S A.* 102(47):16919–16926.
- Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29(22):2933–2935.
- Nováková E, Hypša V, Moran NA. 2009. *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. *BMC Microbiol.* 9(1):143.
- Okonechnikov K, Golosova O, Fursov M. 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28(8):1166–1167.
- Oliver KM, Campos J, Moran NA, Hunter MS. 2008. Population dynamics of defensive symbionts in aphids. *Proc R Soc B Biol Sci.* 275(1632):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(1):247–266.
- Oliver KM, Degnan PH, Hunter MS, Moran NA. 2009. Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science* 325(5943):992–994.
- Oliver KM, Higashi CH. 2019. Variations on a protective theme: *Hamiltonella defensa* infections in aphids variably impact parasitoid success. *Curr Opin Insect Sci.* 32:1–7.
- Padalon-Brauch G, et al. 2008. Small RNAs encoded within genetic islands of *Salmonella typhimurium* show host-induced expression and role in virulence. *Nucleic Acids Res.* 36(6):1913–1927.
- Patel V, et al. 2019. Cultivation-assisted genome of *Candidatus Fukatsuiia symbiotica*; the enigmatic 'X-type' symbiont of aphids. *Genome Biol Evol.* 11(12):3510–3522.
- Polin S, Le Gallic JF, Simon JC, Tsuchida T, Outreman Y. 2015. Conditional reduction of predation risk associated with a facultative symbiont in an insect. *PLoS One* 10(11):e0143728.
- Raychoudhury R, Werren JH. 2012. Host genotype changes bidirectional to unidirectional cytoplasmic incompatibility in *Nasonia longicornis*. *Heredity* 108(2):105–114.

- Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16(6):276–277.
- Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61(3):539–542.
- Russell JA, et al. 2013. Uncovering symbiont-driven genetic diversity across North American pea aphids. *Mol Ecol.* 22(7):2045–2059.
- Sandström JP, Russell JA, White JP, Moran NA. 2001. Independent origins and horizontal transfer of bacterial symbionts of aphids. *Mol Ecol.* 10(1):217–228.
- Sasaki T, Massaki N, Kubo T. 2005. *Wolbachia* variant that induces two distinct reproductive phenotypes in different hosts. *Heredity* 95(5):389–393.
- Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27(6):863–864.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14):2068–2069.
- Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 56(4):564–577.
- Touchon M, Moura de Sousa JA, Rocha EP. 2017. Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. *Curr Opin Microbiol.* 38:66–73.
- Tsuchida T, Koga R, Shibao H, Matsumoto T, Fukatsu T. 2002. Diversity and geographic distribution of secondary endosymbiotic bacteria in natural populations of the pea aphid, *Acyrtosiphon pisum*. *Mol Ecol.* 11(10):2123–2135.
- van der Wilk F, Dullemans AM, Verbeek M, van den Heuvel JF. 1999. Isolation and characterization of APSE-1, a bacteriophage infecting the secondary endosymbiont of *Acyrtosiphon pisum*. *Virology* 262(1):104–113.
- Weldon SR, Strand MR, Oliver KM. 2013. Phage loss and the breakdown of a defensive symbiosis in aphids. *Proc R Soc B Biol Sci.* 280(1751):20122103.
- Worning P, Jensen LJ, Hallin PF, Stærfeldt HH, Ussery DW. 2006. Origin of replication in circular prokaryotic chromosomes. *Environ Microbiol.* 8(2):353–361.
- Zemskov EA, Kang W, Maeda S. 2000. Evidence for nucleic acid binding ability and nucleosome association of *Bombyx mori* nucleopolyhedrovirus BRO proteins. *J Virol.* 74(15):6784–6789.
- Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. *Nucleic Acids Res.* 39(Suppl):W347–W352.

Associate editor: Esperanza Martinez-Romero

Discussion générale

Les systèmes multi-symbiotiques ont émergé à de multiples reprises au sein des arthropodes et leur récente étude a permis de mettre en évidence une grande diversité et complexité de ces systèmes. Dans le genre *Cinara* plusieurs co-symbiotes ont été décrits (Lamelas et al., 2011a; Manzano-Marín et al., 2020a; Meseguer et al., 2017) et ces études ont mis en évidence la prédominance de la bactérie *Serratia symbiotica* comme co-symbiote associé au symbiote primaire *Buchnera aphidicola*. Cependant, aucun scénario n'existait pour retracer l'histoire de cette association. C'est pourquoi, dans ce travail de thèse nous avons essayé de reconstruire l'histoire évolutive de l'association de *Serratia* avec *Buchnera* et le puceron hôte. A travers nos travaux nous avons voulu décrire comment évolue une symbiose à partenaires multiples et ce à l'échelle de leurs génomes mais aussi de leur localisation et régulation au sein de l'hôte.

Une histoire évolutive complexe

De multiples acquisitions

Serratia symbiotica est présente comme co-symbiote obligatoire dans plusieurs sous-familles d'Aphididae (Manzano-marín, 2022; Monnin et al., 2020) et plus particulièrement dans le genre *Cinara* (Lachninae) (Meseguer et al., 2017), dans lesquelles elle forme alors un système di-symbiotique avec le symbiote primaire *Buchnera aphidicola*. Au fil des différentes études sur ces systèmes symbiotiques particuliers, différentes souches de *Serratia* ont été mises en évidence chez les *Cinara*. En décrivant ces différentes lignées de *Serratia*, nous avons observé que plusieurs d'entre elles avaient des caractéristiques différentes. Tout d'abord leurs génomes varient en taille du simple au double (1.1Mb à 2.5Mb) et elles ont des taux de G-C variant de l'attendu de 50% à des taux relativement faibles de 30%. Mais ces symbiotes varient également au niveau de leurs caractéristiques cellulaires avec des localisations intra ou extra-cellulaires, des formes coccoïdes ou baciliformes et des tailles variables. (Manzano-Marín, Simon, et Latorre 2016; Meseguer et al. 2017; Lamelas et al. 2011; Manzano-Marín et Latorre 2016; chapitre 1 et 2). Ces résultats reflètent des associations obligatoires de plus ou moins long terme avec leurs pucerons hôtes (Moran et al., 2008b). Les analyses phylogénomiques utilisant les génomes mitochondriaux des pucerons et les génomes entiers de *Buchnera* et *Serratia* ont permis de confirmer que *Buchnera* s'est diversifié parallèlement avec les *Cinara* dans certaines lignées ; mais aussi pour la première fois que plusieurs acquisitions indépendantes plus ou

moins récentes de lignées distinctes de *Serratia symbiotica* ont eu lieu au sein des *Cinara*. Ainsi, l'histoire évolutive de l'association des *Cinara* et de leur co-symbiote semble bien plus complexe et dynamique que ce que l'on pensait. En effet, plusieurs souches de bactéries différentes présentes comme co-symbiotes avait déjà été observées par (Meseguer et al 2017) et nous venons de montrer que même les hôtes disposant de la même bactérie co-symbiotique peuvent l'avoir obtenue via des acquisitions indépendantes d'âges différents. Néanmoins, le nombre exacte d'acquisitions de *Serratia* reste difficile à déterminer et nécessitera dans l'avenir un échantillonnage plus important dans certains groupes de *Cinara* pour lesquels l'origine de l'acquisition reste incertaine. Par ailleurs, les données existantes dans une autre sous-famille de pucerons, les Chaitophorinae, chez lesquels trois espèces au moins sont associées à *Serratia symbiotica* de manière obligatoire (Manzano-marin, 2022; Monnin et al., 2020; Renoz et al., 2022) ou à l'échelle des Lachninae (Meseguer et al., 2017), suggèrent que l'histoire de *Serratia symbiotica*, comme nouveau partenaire obligatoire complétant *Buchnera*, est emprunte d'acquisitions et de pertes successives, Cette complexité de l'histoire évolutive de l'association avec des co-symbiotes n'est pas exceptionnelle puisque des résultats similaires d'instabilité des systèmes di-symbiotiques ont été observés dans plusieurs clades d'insectes et notamment chez les hémiptères (Dial et al., 2022; Husnik and McCutcheon, 2016; Koga et al., 2013; Rosenblueth et al., 2012; Szabó et al., 2017; Elena R. Toenshoff et al., 2012; Urban and Cryan, 2012).

Et de la co-spéciation

Une des acquisitions indépendantes de *Serratia* que nous avons décrites a eu lieu dans un clade (clade A) pour lequel nous avons montré une diversification parallèle de l'hôte et de ses deux symbiotes. Ce patron de co-spéciation soutenu par les analyses phylogénomiques, la synténie et l'observation de taille de génome et de taux de GC homogènes et faibles semble provenir d'une acquisition relativement ancienne de *Serratia symbiotica*, il y a 25 Ma. L'existence de cette co-spéciation dans une symbiose à trois partenaires n'est pas exceptionnelle chez les *Cinara* puisque Manzano-Marín et al. (2020) avait déjà décrit une co-spéciation du même type avec le co-symbiote *Erwinia* et *Buchnera*. C'est aussi le cas chez d'autres hémiptères comme les Cicadelles (Vasquez and Bennett, 2022). Cependant, nous avons décrit pour la première fois un système dans lequel la vitesse d'évolution de chacun des trois partenaires est corrélée et où l'ensemble des gènes des deux bactéries évoluent sous une forte pression de sélection purifiante, à la même vitesse. Cette découverte suggère que chacun des symbiotes est soumis à des évènements démographiques similaires au cours de la vie de

leur hôte et lors de leur transmission. Cette hypothèse est corroborée pour une espèce de ce clade pour laquelle nos analyses de quantification bactérienne ont montré une dynamique populationnelle similaire entre *Buchnera* et *Serratia* au cours du développement du puceron. De prochaines expériences seront cependant nécessaires afin de vérifier si les évènements démographiques subis par les deux symbiotes sont les mêmes lors de leur transmission à la descendance.

Pourquoi une telle dynamique et pourquoi Serratia symbiotica ?

La raison de la présence simultanée chez les *Cinara* de systèmes di-symbiotiques stables se diversifiant par co-spéciation et de multiples remplacements de co-symbiotes reste à découvrir. En effet, plusieurs hypothèses peuvent être faites impliquant le rôle nutritionnel des symbiotes, la compétition intra-hôte entre les bactéries ou encore un avantage évolutif autre que nutritionnel apporté par le symbiote pour son hôte.

Toutes les symbioses multipartenaires décrites chez les pucerons se sont révélées permettre la complémentation des pertes de fonctions de *Buchnera*. C'est le cas chez les *Cinara* et d'autres espèces étudiées précédemment ou parallèlement à cette thèse (Lamelas et al., 2011a; Manzano-marín, 2022; Manzano-Marín et al., 2018, 2016; Manzano-Marín and Latorre, 2016; Meseguer et al., 2017; Renoz et al., 2022) et nous l'avons également mis en évidence pour l'ensemble des espèces du clade A (chapitre 2). Pour l'ensemble de ces espèces, *Serratia symbiotica* permet principalement la synthèse de vitamine B (Biotine et Riboflavine), ces gènes ayant été perdus par *Buchnera*. Les *Cinara* ne sont pas une exception puisque la plupart des symbioses à partenaires multiples décrites à ce jour sont elles aussi nutritionnelles (Buysse et al., 2021; Dial et al., 2022; Douglas, 2016; Duron and Gottlieb, 2020). Mais quels sont les facteurs qui favorisent la fixation de *Serratia symbiotica* dans les populations de pucerons hôtes (et parfois son remplacement dans certaines lignées) ?

Un avantage évolutif pour l'hôte

Une hypothèse serait que cette nouvelle bactérie apporte de nouvelles compétences à ses hôtes et la dynamique observée serait donc le résultat de la sélection sur l'hôte. Cette compétence pourrait être nutritionnelle et l'acquisition d'un nouveau co-symbiote par l'hôte lui procurerait un avantage évolutif pour s'alimenter sur certains substrats. Meseguer et al (2017) ont exploré cette hypothèse en analysant s'il existait une corrélation entre l'identité taxonomique du co-symbiote et le genre de plante hôte du puceron ou l'organe attaqué sur la

plante. Cette analyse n'a pas montré de lien significatif entre co-symbiote et genre de plante hôte. Les analyses des génomes d'*Erwinia haradae* associées à un clade de *Cinara*, ont par contre montré que celles-ci pouvaient synthétiser la Thiamine (Manzano-Marín et al., 2020a), une compétence métabolique absente de la plupart des génomes de *Buchnera*. En effet, seules les *Buchnera* associés à deux autres genres de pucerons se nourrissant sur conifères (*Neophyllaphis* et *Mindarus*) peuvent synthétiser la Thiamine (Manzano-Marín et al. 2022). Nos analyses génomiques au sein du clade A (Chapitre 2) montrent élégamment que *Serratia symbiota* peut synthétiser cette vitamine. Par contraste, *Serratia symbiotica*, associée à *Tuberolachnus salignus* (se nourrissant sur chêne) n'a pas cette compétence (Manzano-Marín et al., 2016). On peut donc émettre l'hypothèse que cette « nouvelle » compétence (apportée par *Serratia* dans un clade *Cinara* et par *Erwinia* dans un autre) a peut-être permis la fixation de ces bactéries chez les espèces de pucerons se nourrissant sur conifères.

Pour vérifier l'hypothèse selon laquelle les nouveaux symbiotes peuvent augmenter les capacités métaboliques de leurs hôtes (et pas seulement compléter celles apportées par *Buchnera*), il serait nécessaire d'annoter tous les génomes de *Serratia symbiotica* disponibles chez les pucerons (ainsi que ceux des autres endosymbiotes obligatoires) afin de prendre en compte les capacités métaboliques de chaque co-symbiote. Dans la plupart des études sur les systèmes di-symbiotiques chez les pucerons, les génomes des co-symbiotes sont partiellement annotés (Monnin et al., 2020; Renoz et al., 2022; Yorimoto et al., 2022) pour vérifier la présence des gènes comblant les déficiences de *Buchnera*, mais les nouvelles capacités des co-symbiotes sont insuffisamment explorées. D'autre part, la composition exacte de l'alimentation des pucerons dans ce genre n'a pas été étudiée, il pourrait exister des différences dans la composition de la sève élaborée que le puceron ingère. Il n'existe pas non plus d'études comparative de la composition de la sève des conifères et des angiospermes, la seule étude récente dans ce domaine compare les contenus en sucres et non la présence d'acides aminés et de vitamines (Gallinger and Gross, 2018). Des études complémentaires consisteraient donc à comparer de façon précise : capacités métaboliques du co-symbiote et composition du phloème.

Les co-symbiotes pourraient également apporter un avantage évolutif non nutritionnel, leur fixation dans les populations et leur dynamique à long terme pourraient être le résultat de la sélection sur des traits autres que la nutrition. De précédents travaux sur des *Serratia symbiotica* facultatives ont mis en évidence que leur présence augmentait la résistance aux chocs thermiques de l'hôte (Burke et al., 2010; Montllor et al., 2002). Ainsi on peut supposer que les pucerons porteurs de cette bactérie soient avantagés dans certaines conditions

environnementales et que *Serratia symbiotica* puisse être fixée dans les populations correspondantes. Lors de la perte de certaines gènes essentiels chez le symbiote primaire *Buchnera*, *Serratia symbiotica* pourrait alors combler les déficits de *Buchnera*. Il est difficile cependant de comprendre pourquoi cette *Serratia symbiotica*, dans certains cas, serait remplacée par une nouvelle souche de la même espèce de bactérie (comme suggérée par nos analyses chapitre 1) ou un nouveau symbiote. Les avantages conférés par les symbiotes à leurs hôtes varient (Cass et al., 2016; Doremus et al., 2018; Polin et al., 2015) et la dynamique observée pourrait être le résultat de sélection sur diverses caractéristiques. En ce qui concerne *Hamiltonella* (chapitre 4), cette bactérie procure une défense contre les guêpes parasitoïdes conférée par les APSE (bacteriophages) (Lynn-Bell et al., 2019; Oliver and Higashi, 2019). L'analyse de ces APSE a révélé que leurs génomes sont hautement conservés, la plupart des variations étant localisées dans la cassette de la toxine protectrice. Il n'existe qu'une espèce de *Cinara* chez laquelle *Hamiltonella* est potentiellement impliquée dans une symbiose obligatoire (Manzano-Marín et al., 2020a), mais les phages associés n'ont pas été décrits. De manière générale les phages sont encore peu décrits parmi les bactéries endosymbiotiques obligatoires. Il serait donc intéressant d'effectuer une recherche de phages plus exhaustive au sein des différents génomes de symbiotes présents chez les *Cinara*.

Ou une compétition entre endosymbiotes bactériens

Une hypothèse alternative expliquant l'histoire évolutive de cette association, consisterait non plus à l'apport d'un avantage évolutif du symbiote pour l'hôte, mais plutôt des remplacements via des mécanismes de sélection à l'échelle des populations bactériennes. En effet, plusieurs études ont mis en évidence chez les pucerons la présence de bactéries facultatives plus ou moins diverses selon l'espèce hôte (Manzano-marín, 2022; Zytynska and Weisser, 2016). Ainsi, plusieurs partenaires alternatifs sont potentiellement présents dans le « microbiote » de l'hôte. Compte tenu de l'ubiquité des fonctions perdues par *Buchnera* chez les bactéries libres ou facultatives, n'importe quelle bactérie présente avec « les bonnes » capacités métaboliques (Biotine et Riboflavine) pourrait remplacer le co-symbiote. De plus, l'existence de compétition intra-bactériocyte entre différentes lignées bactériennes a déjà été mise en évidence par l'étude de Perreau et al., (2021) pour différentes souches de *Buchnera*. Ainsi, les remplacements successifs de co-symbiotes chez les *Cinara* pourrait provenir d'une compétition entre les bactéries présentes chez l'hôte. La perte de fonctions essentielles chez *Buchnera*, ouvrirait une niche écologique pouvant être remplie par n'importe quel symbiote présent et capable d'occuper cette niche. Si l'on accepte cette hypothèse, les patrons de co-

spéciation durant de longues périodes pourraient en partie être le résultat d'une absence de compétition. La rareté de partenaires alternatifs (c'est-à-dire l'absence de bactéries secondaires alternative chez certains *Cinara* et/ou leur incapacité compétitive à remplacer l'ancien co-symbiote) expliquerait la dynamique évolutive observée. Cette hypothèse s'accorde avec la présence majoritaire de *Serratia symbiotica* comme co-symbiote des pucerons puisque celle-ci est retrouvée comme symbiote facultatif dans un grand nombre d'espèces de pucerons (Burke et al., 2009; Pons et al., 2021; Zytynska and Weisser, 2016).

Régulations de l'interaction hôte co-symbiote

La régulation des symbiotes obligatoires par l'hôte paraît primordiale pour lui permettre d'une part de minimiser les coûts de la symbiose et d'autre part de maximiser les bénéfices qu'il en retire et notamment la production des différents nutriments selon ses besoins au cours de son développement. Cette régulation, précédemment mise en évidence pour *Buchnera* chez le puceron du pois par [Simonet et al., \(2018, 2016\)](#) a aussi été observée dans nos analyses chez les *Cinara*. Cependant, cette régulation semble ne pas être présente ou du moins être différente selon l'espèce hôte. Les reconstructions phylogénomiques, la description des génomes et l'observation des bactéries et de leur localisation, semble indiquer que cette différence de régulation est corrélée à des différences entre les lignées de *Serratia*. En effet, les *Serratia* étudiées proviennent d'acquisitions indépendantes plus ou moins anciennes et disposent de génomes dont la taille et le taux de GC diffèrent. De plus, leur localisation pour l'une intracellulaire et pour l'autre extracellulaire suggère qu'il existe une importance particulière du passage à un mode de vie intracellulaire pour permettre la régulation du symbiote par l'hôte, ce qui est en accord avec les études montrant que la compartimentation des symbiotes est importante pour leur régulation (Chomicki et al., 2020).

Afin d'approfondir les connaissances sur la régulation des co-symbiotes par l'hôte, de futures analyses utilisant des élevages de puceron sur des milieux artificiels déficients ou complétés permettront de vérifier si l'hôte peut contrôler sa densité de co-symbiotes selon ses besoins. En effet, une telle régulation a déjà été montrée pour *Buchnera* (Colella et al., 2018; Viñuelas et al., 2011; Wilkinson et al., 2007) et dans d'autres modèles Hémiptères (Snyder et al., 2012, 2010). Nous pourrions ainsi manipuler le milieu en vitamines (produites par le co-symbiote) ou en acides aminés (produits par *Buchnera*) et vérifier si les pucerons régulent leurs deux symbiotes en fonction de cette manipulation.

Réduction des génomes symbiotiques

A travers nos résultats nous confirmons l'existence d'une convergence des génomes symbiotiques vers une forme réduite stable où seuls les gènes essentiels subsistent. En effet, hormis les souches de *C. cembrae* et *C. tujafilina* dérivant très vraisemblablement d'une acquisition très récente, la plupart des lignées de *Serratia* séquencées disposent malgré leurs acquisitions indépendantes de génomes réduits. Un des points marquants de nos analyses est le manque voire l'absence de certains intermédiaires entre les bactéries facultatives faiblement réduites et les co-symbiotes obligatoires montrant une longue histoire co-évolutive avec leurs hôtes et ayant un génome fortement réduit. De la même manière les études précédentes sur les Lachninae mais aussi sur d'autres groupes d'Hémiptères disposant de co-symbiotes ne décrivent pas ou peu de stades intermédiaires dans cette transition de mode de vie facultatif à obligatoire. Ensemble, ces données semblent indiquer que le processus d'érosion du génome lors de cette transition se produit rapidement dans la phase initiale après l'acquisition du co-symbiote, ce qui expliquerait la difficulté que nous avons à le caractériser même à l'échelle des gènes. En effet, nos mesures des taux d'évolution du clade A des *Cinara* ne semblent pas montrer d'accélération et peu de pertes de gènes entre les différentes espèces, ce qui impliquerait que la réduction initiale du génome de *Serratia symbiotica* s'est produit avant la diversification de ce clade A. Cependant, les mécanismes impliqués dans cette réduction rapide sont inconnus. Les scénarii expliquant cette réduction (voir introduction) suggèrent que ce processus est entièrement dû à la dérive génétique (absence de sélection sur certains gènes et effets démographiques) (Toft and Andersson, 2010). Mais il est également parfois suggéré que l'érosion génomique pourrait provenir d'une sélection par l'hôte sur les symbiotes plus petits, permettant une réduction des coûts de la symbiose et une plus grande stabilité (Ankrah et al., 2018), mais aussi d'une sélection intra-hôte avec une compétition entre les symbiotes (Perreau et al., 2021). Une autre question qui reste en suspens est jusqu'à quelle limite continue la réduction des génomes symbiotiques ? En effet, si la première phase de réduction semble être rapide, nous avons aussi mis en évidence la perte de gènes des symbiotes entre les différentes espèces du clade A des *Cinara* ce qui indique que les génomes continuent leur érosion à un rythme plus lent. Cette perte de gènes continue pourrait être un des facteurs impliqués dans les remplacements de symbiote en rendant certains co-symbiotes incapables de compléter *Buchnera* à long terme.

Conclusion

L'ensemble des travaux effectués pendant cette thèse ont permis d'approfondir les connaissances et la compréhension générale de la symbiose multipartenaire chez les pucerons. Premièrement, nous avons confirmé la grande dynamique de ces systèmes symbiotiques dans lesquels de multiples remplacements se produisent, mais aussi des périodes de co-spéciation pendant plusieurs dizaines de millions d'années. Nous avons aussi montré qu'à chaque acquisition indépendante décrite il existe une convergence vers des génomes bactériens réduits. Cependant, les mécanismes à l'origine de cette instabilité restent difficiles à déterminer et paraissent surprenants en comparaison avec l'extrême stabilité de l'association avec le symbiote primaire *Buchnera* dans l'ensemble des pucerons. Deuxièmement, nous avons pu montrer pour la première fois l'existence d'une évolution parallèle entre chacun des partenaires d'une symbiose multipartenaire en mesurant leur taux d'évolution au cours de leur co-spéciation. Nous avons également proposé comme raison de cette évolution parallèle la présence de processus démographiques équivalents chez chacun des symbiotes. Finalement, ce travail confirme le rôle important de ces nouveaux symbiotes dans la complémentation du symbiote primaire *Buchnera* confortant l'idée que les mécanismes par lesquels les bactéries libres s'établissent comme partenaires obligatoires dans des endosymbioses multi-partenaires sont les mêmes chez les espèces *Cinara* avec des rôles très similaires à travers les lignées. En outre, ce travail souligne que le genre *Cinara* est un modèle d'étude d'intérêt pour l'étude des mécanismes impliqués dans la mise en place et l'évolution des symbioses à plusieurs partenaires de par la grande diversité de systèmes présents dans des espèces phylogénétiquement proches.

Références bibliographiques

- Akman Gündüz, E., Douglas, A.E., 2009. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. *Proc. R. Soc. B Biol. Sci.* 276, 987–991. <https://doi.org/10.1098/rspb.2008.1476>
- Aksoy, S., 1995. *Wigglesworthia* gen. nov. and *Wigglesworthia glossinidia* sp. nov., Taxa Consisting of the Mycetocyte-Associated, Primary Endosymbionts of Tsetse Flies. *Int. J. Syst. Bacteriol.* 45, 848–851. <https://doi.org/10.1099/00207713-45-4-848>
- Alegado, R.A., Brown, L.W., Cao, S., Dermenjian, R.K., Zuzow, R., Fairclough, S.R., Clardy, J., King, N., 2012. A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. *eLife* 1, e00013. <https://doi.org/10.7554/eLife.00013>
- Altschul, S., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
- Andersson, J.O., Andersson, S.G., 1999. Genome degradation is an ongoing process in *Rickettsia*. *Mol. Biol. Evol.* 16, 1178–1191. <https://doi.org/10.1093/oxfordjournals.molbev.a026208>
- Anisimova, M., Bielawski, J.P., Yang, Z., 2001. Accuracy and Power of the Likelihood Ratio Test in Detecting Adaptive Molecular Evolution. *Mol. Biol. Evol.* 18, 1585–1592. <https://doi.org/10.1093/oxfordjournals.molbev.a003945>
- Ankrah, N.Y.D., Chouaia, B., Douglas, A.E., 2018. The Cost of Metabolic Interactions in Symbioses between Insects and Bacteria with Reduced Genomes. *mBio* 9, e01433-18. <https://doi.org/10.1128/mBio.01433-18>
- Arab, D.A., Bourguignon, T., Wang, Z., Ho, S.Y.W., Lo, N., 2020. Evolutionary rates are correlated between cockroach symbionts and mitochondrial genomes. *Biol. Lett.* 16, 20190702. <https://doi.org/10.1098/rsbl.2019.0702>
- Arab, D.A., Lo, N., 2020. Evolutionary rates are correlated between *Buchnera* endosymbionts and mitochondrial genomes of their aphid hosts (preprint). *Evolutionary Biology*. <https://doi.org/10.1101/2020.11.15.383851>
- Asplen, M.K., Bano, N., Brady, C.M., Desneux, N., Hopper, K.R., Malouines, C., Oliver, K.M., White, J.A., Heimpel, G.E., 2014. Specialisation of bacterial endosymbionts that protect aphids from parasitoids: Defensive symbiosis in the cowpea aphid. *Ecol. Entomol.* 39, 736–739. <https://doi.org/10.1111/een.12153>
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* 19, 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Baumann, L., Baumann, P., 1994. Growth Kinetics of the Endosymbiont *Buchnera aphidicola* in the Aphid *Schizaphis graminum*. *Appl. Environ. Microbiol.* 60, 3440–3443. <https://doi.org/10.1128/aem.60.9.3440-3443.1994>
- Baumann, P., 2005a. Biology of Bacteriocyte-Associated Endosymbionts of Plant Sap-Sucking Insects. *Annu. Rev. Microbiol.* 59, 155–189. <https://doi.org/10.1146/annurev.micro.59.030804.121041>
- Baumann, P., 2005b. Biology of bacteriocyte associated endosymbionts of plant sap-sucking insects. *Annu. Rev. Microbiol.* 59, 155–189. <https://doi.org/10.1146/annurev.micro.59.030804.121041>

- Baumann, P., Lai, C.Y., Rouhbakhsh, D., Moran, N.A., Clark, M.A., 1995. Genetics, physiology, and evolutionary relationships of the genus *Buchnera*: intracellular symbionts of aphids. *Annu. Rev. Microbiol.* 491 55-94.
- Bennett, G.M., Moran, N.A., 2015. Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proc. Natl. Acad. Sci.* 112, 10169–10176. <https://doi.org/10.1073/pnas.1421388112>
- Bennett, G.M., Moran, N.A., 2013. Small, Smaller, Smallest: The Origins and Evolution of Ancient Dual Symbioses in a Phloem-Feeding Insect. *Genome Biol. Evol.* 5, 1675–1688. <https://doi.org/10.1093/gbe/evt118>
- Bermingham, J., Rabatel, A., Calevro, F., Viñuelas, J., Febvay, G., Charles, H., Douglas, A., Wilkinson, T., 2009. Impact of Host Developmental Age on the Transcriptome of the Symbiotic Bacterium *Buchnera aphidicola* in the Pea Aphid (*Acyrtosiphon pisum*). *Appl. Environ. Microbiol.* 75, 7294–7297. <https://doi.org/10.1128/AEM.01472-09>
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsche, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* 69, 313–319. <https://doi.org/10.1016/j.ympev.2012.08.023>
- Binetruy, F., Buysse, M., Lejarre, Q., Barosi, R., Villa, M., Rahola, N., Paupy, C., Ayala, D., Duron, O., 2020. Microbial community structure reveals instability of nutritional symbiosis during the evolutionary radiation of *Amblyomma* ticks. *Mol. Ecol.* 29, 1016–1029. <https://doi.org/10.1111/mec.15373>
- Blow, F., Bueno, E., Clark, N., Zhu, D.T., Chung, S.H., Güllert, S., Schmitz, R.A., Douglas, A.E., 2020. B-vitamin nutrition in the pea aphid-*Buchnera* symbiosis. *J. Insect Physiol.* 126, 104092. <https://doi.org/10.1016/j.jinsphys.2020.104092>
- Boscaro, V., Kolisko, M., Felletti, M., Vannini, C., Lynn, D.H., Keeling, P.J., 2017. Parallel genome reduction in symbionts descended from closely related free-living bacteria. *Nat. Ecol. Evol.* 1, 1160–1167. <https://doi.org/10.1038/s41559-017-0237-0>
- Boucher, D.H., James, S., Keeler, K.H., 1982. The Ecology of Mutualism. *Annu. Rev. Ecol. Syst.* 13, 315–347. <https://doi.org/10.1146/annurev.es.13.110182.001531>
- Bourguignon, T., Kinjo, Y., Villa-Martín, P., Coleman, N.V., Tang, Q., Arab, D.A., Wang, Z., Tokuda, G., Hongoh, Y., Ohkuma, M., Ho, S.Y.W., Pigolotti, S., Lo, N., 2020. Increased Mutation Rate Is Linked to Genome Reduction in Prokaryotes. *Curr. Biol.* 30, 3848-3855.e4. <https://doi.org/10.1016/j.cub.2020.07.034>
- Braendle, C., Miura, T., Bickel, R., Shingleton, A.W., Kambhampati, S., Stern, D.L., 2003. Developmental Origin and Evolution of Bacteriocytes in the Aphid–*Buchnera* Symbiosis. *PLoS Biol.* 1, e21. <https://doi.org/10.1371/journal.pbio.0000021>
- Brandt, J.W., Chevignon, G., Oliver, K.M., Strand, M.R., 2017. Culture of an aphid heritable symbiont demonstrates its direct role in defence against parasitoids. *Proc. R. Soc. B Biol. Sci.* 284, 20171925. <https://doi.org/10.1098/rspb.2017.1925>
- Bressan, A., Terlizzi, F., Credi, R., 2012. Independent Origins of Vectored Plant Pathogenic Bacteria from Arthropod-Associated *Arsenophonus* Endosymbionts. *Microb. Ecol.* 63, 628–638. <https://doi.org/10.1007/s00248-011-9933-5>
- Bright, M., Bulgheresi, S., 2010. A complex journey: transmission of microbial symbionts. *Nat. Rev. Microbiol.* 8, 218–230. <https://doi.org/10.1038/nrmicro2262>

- Brune, A., 2014. Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* 12, 168–180. <https://doi.org/10.1038/nrmicro3182>
- Buchner, P., 1953a. Endosymbiose der Tiere mit Pflanzlichen Mikroorganismen, Endosymbiose der Tiere mit Pflanzlichen Mikroorganismen. <https://doi.org/10.1007/978-3-0348-6958-4>
- Buchner, P., 1953b. Endosymbiose der Tiere mit pflanzlichen Mikroorganismen.
- Burke, G., Fiehn, O., Moran, N., 2010. Effects of facultative symbionts and heat stress on the metabolome of pea aphids. *ISME J.* 4, 242–252. <https://doi.org/10.1038/ismej.2009.114>
- Burke, G.R., Normark, B.B., Favret, C., Moran, N.A., 2009. Evolution and Diversity of Facultative Symbionts from the Aphid Subfamily Lachninae. *Appl. Environ. Microbiol.* 75, 5328–5335. <https://doi.org/10.1128/AEM.00717-09>
- Buysse, M., Floriano, A.M., Gottlieb, Y., Nardi, T., Comandatore, F., Olivieri, E., Giannetto, A., Palomar, A.M., Makepeace, B.L., Bazzocchi, C., Cafiso, A., Sasser, D., Duron, O., 2021. A dual endosymbiosis supports nutritional adaptation to hematophagy in the invasive tick *Hyalomma marginatum*. *eLife* 10, e72747. <https://doi.org/10.7554/eLife.72747>
- Caspi-Fluger, A., Inbar, M., Mozes-Daube, N., Katzir, N., Portnoy, V., Belausov, E., Hunter, M.S., Zchori-Fein, E., 2012. Horizontal transmission of the insect symbiont *Rickettsia* is plant-mediated. *Proc. R. Soc. B Biol. Sci.* 279, 1791–1796. <https://doi.org/10.1098/rspb.2011.2095>
- Cass, B.N., Himler, A.G., Bondy, E.C., Bergen, J.E., Fung, S.K., Kelly, S.E., Hunter, M.S., 2016. Conditional fitness benefits of the *Rickettsia* bacterial symbiont in an insect pest. *Oecologia* 180, 169–179. <https://doi.org/10.1007/s00442-015-3436-x>
- Chen, C.-Y., Lai, C.-Y., Kuo, M.-H., 2009. Temperature effect on the growth of Buchnera endosymbiont in *Aphis craccivora* (Hemiptera: Aphididae). *Symbiosis* 49, 53–59. <https://doi.org/10.1007/s13199-009-0011-4>
- Chen, D.-Q., Montllor, C.B., Purcell, A.H., 2000. Fitness effects of two facultative endosymbiotic bacteria on the pea aphid, *Acyrtosiphon pisum*, and the blue alfalfa aphid, *A. kondoi*. *Entomol. Exp. Appl.* 95, 315–323. <https://doi.org/10.1046/j.1570-7458.2000.00670.x>
- Chen, R., Favret, C., Jiang, L., Wang, Z., Qiao, G., 2016. An aphid lineage maintains a bark-feeding niche while switching to and diversifying on conifers. *Cladistics* 32, 555–572. <https://doi.org/10.1111/cla.12141>
- Chen, X., Li, S., Aksoy, S., 1999. Concordant Evolution of a Symbiont with Its Host Insect Species: Molecular Phylogeny of Genus *Glossina* and Its Bacteriome-Associated Endosymbiont, *Wigglesworthia glossinidia*. *J. Mol. Evol.* 48, 49–58. <https://doi.org/10.1007/PL00006444>
- Chevenet, F., Doyon, J.-P., Scornavacca, C., Jacox, E., Jousset, E., Berry, V., 2016. SylvX: a viewer for phylogenetic tree reconciliations. *Bioinformatics* 32, 608–610. <https://doi.org/10.1093/bioinformatics/btv625>
- Chomicki, G., Weber, M., Antonelli, A., Bascompte, J., Kiers, E.T., 2019. The Impact of Mutualisms on Species Richness. *Trends Ecol. Evol.* 34, 698–711. <https://doi.org/10.1016/j.tree.2019.03.003>
- Chomicki, G., Werner, G.D.A., West, S.A., Kiers, E.T., 2020. Compartmentalization drives the evolution of symbiotic cooperation. *Philos. Trans. R. Soc. B Biol. Sci.* 375, 20190602. <https://doi.org/10.1098/rstb.2019.0602>

- Chong, R.A., Park, H., Moran, N.A., 2019. Genome Evolution of the Obligate Endosymbiont *Buchnera aphidicola*. *Mol. Biol. Evol.* 36, 1481–1489. <https://doi.org/10.1093/molbev/msz082>
- Christian, N., Whitaker, B.K., Clay, K., 2015. Microbiomes: unifying animal and plant systems through the lens of community ecology theory. *Front. Microbiol.* 6. <https://doi.org/10.3389/fmicb.2015.00869>
- Clark, M.A., Moran, N.A., Baumann, P., 1999. Sequence evolution in bacterial endosymbionts having extreme base compositions. *Mol. Biol. Evol.* 16, 1586–1598. <https://doi.org/10.1093/oxfordjournals.molbev.a026071>
- Clayton, A.L., Jackson, D.G., Weiss, R.B., Dale, C., 2016. Adaptation by Deletogenic Replication Slippage in a Nascent Symbiont. *Mol. Biol. Evol.* 33, 1957–1966. <https://doi.org/10.1093/molbev/msw071>
- Colella, S., Parisot, N., Simonet, P., Gaget, K., Duport, G., Baa-Puyoulet, P., Rahbé, Y., Charles, H., Febvay, G., Callaerts, P., Calevro, F., 2018. Bacteriocyte Reprogramming to Cope With Nutritional Stress in a Phloem Sap Feeding Hemipteran, the Pea Aphid *Acyrtosiphon pisum*. *Front. Physiol.* 9, 1498. <https://doi.org/10.3389/fphys.2018.01498>
- Dalet, C., Maudlin, I., 1999. *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. *Int. J. Syst. Bacteriol.* 9.
- Danneels, B., Pinto-Carbó, M., Carlier, A., 2018. Patterns of Nucleotide Deletion and Insertion Inferred from Bacterial Pseudogenes. *Genome Biol. Evol.* 10, 1792–1802. <https://doi.org/10.1093/gbe/evy140>
- Darling, A.C.E., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Res.* 14, 1394–1403. <https://doi.org/10.1101/gr.2289704>
- Dayel, M.J., Alegado, R.A., Fairclough, S.R., Levin, T.C., Nichols, S.A., McDonald, K., King, N., 2011. Cell differentiation and morphogenesis in the colony-forming choanoflagellate *Salpingoeca rosetta*. *Dev. Biol.* 357, 73–82. <https://doi.org/10.1016/j.ydbio.2011.06.003>
- Dedeine, F., Vavre, F., Fleury, F., Loppin, B., Hochberg, M.E., Boulétreau, M., 2001. Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. *Proc. Natl. Acad. Sci.* 98, 6247–6252. <https://doi.org/10.1073/pnas.101304298>
- Degnan, P.H., Lazarus, A.B., Brock, C.D., Wernegreen, J.J., 2004. Host–Symbiont Stability and Fast Evolutionary Rates in an Ant–Bacterium Association: Cospeciation of *Camponotus* Species and Their Endosymbionts, *Candidatus Blochmannia*. *Syst. Biol.* 53, 95–110. <https://doi.org/10.1080/10635150490264842>
- Degnan, P.H., Lazarus, A.B., Wernegreen, J.J., 2005. Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects. *Genome Res.* 15, 1023–1033. <https://doi.org/10.1101/gr.3771305>
- Degnan, P.H., Leonardo, T.E., Cass, B.N., Hurwitz, B., Stern, D., Gibbs, R.A., Richards, S., Moran, N.A., 2009. Dynamics of genome evolution in facultative symbionts of aphids. *Environ. Microbiol.* <https://doi.org/10.1111/j.1462-2920.2009.02085.x>
- Degnan, P.H., Moran, N.A., 2008. Diverse Phage-Encoded Toxins in a Protective Insect Endosymbiont. *Appl. Environ. Microbiol.* 74, 6782–6791. <https://doi.org/10.1128/AEM.01285-08>

- Dial, D.T., Weglarz, K.M., Aremu, A.O., Havill, N.P., Pearson, T.A., Burke, G.R., von Dohlen, C.D., 2022. Transitional genomes and nutritional role reversals identified for dual symbionts of adelgids (Aphidoidea: Adelgidae). *ISME J.* 16, 642–654. <https://doi.org/10.1038/s41396-021-01102-w>
- Dion, E., Zélé, F., Simon, J.-C., Outreman, Y., 2011. Rapid evolution of parasitoids when faced with the symbiont-mediated resistance of their hosts: Symbiont effects on host-parasitoid evolution. *J. Evol. Biol.* 24, 741–750. <https://doi.org/10.1111/j.1420-9101.2010.02207.x>
- Dohlen, C.D.V., Moran, N.A., 2000. Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation. *Biol. J. Linn. Soc.* 71, 689–717. <https://doi.org/10.1111/j.1095-8312.2000.tb01286.x>
- Doré, J., Simrén, M., Buttle, L., Guarner, F., 2013. Hot topics in gut microbiota. *United Eur. Gastroenterol. J.* 1, 311–318. <https://doi.org/10.1177/2050640613502477>
- Doremus, M.R., Smith, A.H., Kim, K.L., Holder, A.J., Russell, J.A., Oliver, K.M., 2018. Breakdown of a defensive symbiosis, but not endogenous defences, at elevated temperatures. *Mol. Ecol.* 27, 2138–2151. <https://doi.org/10.1111/mec.14399>
- Dossi, F.C.A., da Silva, E.P., Cônsoli, F.L., 2014. Population Dynamics and Growth Rates of Endosymbionts During *Diaphorina citri* (Hemiptera, Liviidae) Ontogeny. *Microb. Ecol.* 68, 881–889. <https://doi.org/10.1007/s00248-014-0463-9>
- Douglas, A.E., 2016. How multi-partner endosymbioses function. *Nat. Rev. Microbiol.* 14, 731–743. <https://doi.org/10.1038/nrmicro.2016.151>
- Douglas, A.E., 2009. The microbial dimension in insect nutritional ecology. *Funct. Ecol.* 23, 38–47. <https://doi.org/10.1111/j.1365-2435.2008.01442.x>
- Douglas, A.E., 1998a. Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria *Buchnera*. *Annu. Rev. Entomol.* 43, 17–37. <https://doi.org/10.1146/annurev.ento.43.1.17>
- Douglas, A.E., 1998b. Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria *Buchnera*. *Annu. Rev. Entomol.* 43, 17–37. <https://doi.org/10.1146/annurev.ento.43.1.17>
- Douglas, A.E., 1992. Requirement of pea aphids (*Acyrtosiphon pisum*) for their symbiotic bacteria. *Entomol. Exp. Appl.* 65, 195–198. <https://doi.org/10.1111/j.1570-7458.1992.tb01643.x>
- Douglas, A.E., Werren, J.H., 2016. Holes in the Hologenome: Why Host-Microbe Symbioses Are Not Holobionts. *mBio* 7, e02099-15. <https://doi.org/10.1128/mBio.02099-15>
- Doyon, J.-P., Ranwez, V., Daubin, V., Berry, V., 2011. Models, algorithms and programs for phylogeny reconciliation. *Brief. Bioinform.* 12, 392–400. <https://doi.org/10.1093/bib/bbr045>
- Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. <https://doi.org/10.1093/molbev/mss075>
- Durak, R., Borowiak-Sobkowiak, B., 2013. Influence of temperature on the biological parameters of the anholocyclic species *Cinara tujafilina* (Hemiptera: Aphidoidea). *Open Life Sci.* 8, 570–577. <https://doi.org/10.2478/s11535-013-0161-x>
- Duron, O., Gottlieb, Y., 2020. Convergence of Nutritional Symbioses in Obligate Blood Feeders. *Trends Parasitol.* 36, 816–825. <https://doi.org/10.1016/j.pt.2020.07.007>

- Duron, O., Hurst, G.D.D., Hornett, E.A., Josling, J.A., Engelstädter, J., 2008. High incidence of the maternally inherited bacterium *Cardinium* in spiders. *Mol. Ecol.* 17, 1427–1437. <https://doi.org/10.1111/j.1365-294X.2008.03689.x>
- Duron, O., Morel, O., Noël, V., Buysse, M., Binetruy, F., Lancelot, R., Loire, E., Ménard, C., Bouchez, O., Vavre, F., Vial, L., 2018. Tick-Bacteria Mutualism Depends on B Vitamin Synthesis Pathways. *Curr. Biol.* 28, 1896-1902.e5. <https://doi.org/10.1016/j.cub.2018.04.038>
- Elston, K.M., Leonard, S.P., Geng, P., Bialik, S.B., Robinson, E., Barrick, J.E., 2022. Engineering insects from the endosymbiont out. *Trends Microbiol.* 30, 79–96. <https://doi.org/10.1016/j.tim.2021.05.004>
- Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16, 157. <https://doi.org/10.1186/s13059-015-0721-2>
- Engl, T., Schmidt, T.H.P., Kanyile, S.N., Klebsch, D., 2020. Metabolic Cost of a Nutritional Symbiont Manifests in Delayed Reproduction in a Grain Pest Beetle. *Insects* 11, 717. <https://doi.org/10.3390/insects11100717>
- Favret, C., Havill, N.P., Miller, G.L., Sano, M., Victor, B., 2015. Catalog of the adelgids of the world (Hemiptera, Adelgidae). *ZooKeys* 534, 35–54. <https://doi.org/10.3897/zookeys.534.6456>
- Fenn, K., Blaxter, M., 2007. Coexist, Cooperate and Thrive: *Wolbachia* as Long-Term Symbionts of Filarial Nematodes, in: Hoerauf, A., Rao, R.U. (Eds.), *Issues in Infectious Diseases*. KARGER, Basel, pp. 66–76. <https://doi.org/10.1159/000104233>
- Ferrari, J., Vavre, F., 2011. Bacterial symbionts in insects or the story of communities affecting communities. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 1389–1400. <https://doi.org/10.1098/rstb.2010.0226>
- Fialho, R.F., Stevens, L., 2000. Male-killing *Wolbachia* in a flour beetle. *Proc. R. Soc. Lond. B Biol. Sci.* 267, 1469–1473. <https://doi.org/10.1098/rspb.2000.1166>
- Fleury, F., Vavre, F., Ris, N., Fouillet, P., Boulétreau, M., 2000. Physiological cost induced by the maternally-transmitted endosymbiont *Wolbachia* in the *Drosophila* parasitoid *Leptopilina heterotoma*. *Parasitology* 121, 493–500. <https://doi.org/10.1017/S0031182099006599>
- Flores, H.A., O'Neill, S.L., 2018. Controlling vector-borne diseases by releasing modified mosquitoes. *Nat. Rev. Microbiol.* 16, 508–518. <https://doi.org/10.1038/s41579-018-0025-0>
- Frago, E., Dicke, M., Godfray, H.C.J., 2012. Insect symbionts as hidden players in insect–plant interactions. *Trends Ecol. Evol.* 27, 705–711. <https://doi.org/10.1016/j.tree.2012.08.013>
- Frago, E., Zytynska, S.E., Fatouros, N.E., 2020. Microbial symbionts of herbivorous species across the insect tree, in: *Advances in Insect Physiology*. Elsevier, pp. 111–159. <https://doi.org/10.1016/bs.aiip.2020.04.002>
- Fukatsu, T., Ishikawa, H., 1998. Differential immunohistochemical visualization of the primary and secondary intracellular symbiotic bacteria of aphids. *Appl. Entomol. Zool.* 33, 321–326. <https://doi.org/10.1303/aez.33.321>
- Funk, D.J., Helbling, L., Wernegreen, J.J., Moran, N.A., 2000. Intraspecific phylogenetic congruence among multiple symbiont genomes. *Proc. R. Soc. Lond. B Biol. Sci.* 267, 2517–2521. <https://doi.org/10.1098/rspb.2000.1314>

- Gabriel, W., Lynch, M., Bürger, R., 1993. muller's ratchet and mutational meltdowns. *Evolution* 47, 1744–1757. <https://doi.org/10.1111/j.1558-5646.1993.tb01266.x>
- Gallinger, J., Gross, J., 2018. Unraveling the Host Plant Alternation of *Cacopsylla pruni* – Adults but Not Nymphs Can Survive on Conifers Due to Phloem/Xylem Composition. *Front. Plant Sci.* 9, 484. <https://doi.org/10.3389/fpls.2018.00484>
- Garber, A.I., Kupper, M., Laetsch, D.R., Weldon, S.R., Ladinsky, M.S., Bjorkman, P.J., McCutcheon, J.P., 2021. The evolution of interdependence in a four-way mealybug symbiosis 29. <https://doi.org/10.1093/gbe/evab123>
- Gaupels, F., Buhtz, A., Knauer, T., Deshmukh, S., Waller, F., van Bel, A.J.E., Kogel, K.-H., Kehr, J., 2008a. Adaptation of aphid stylectomy for analyses of proteins and mRNAs in barley phloem sap. *J. Exp. Bot.* 59, 3297–3306. <https://doi.org/10.1093/jxb/ern181>
- Gaupels, F., Knauer, T., van Bel, A.J.E., 2008b. A combinatory approach for analysis of protein sets in barley sieve-tube samples using EDTA-facilitated exudation and aphid stylectomy. *J. Plant Physiol.* 165, 95–103. <https://doi.org/10.1016/j.jplph.2007.07.023>
- Gil, R., Sabater-Muñoz, B., Latorre, A., Silva, F.J., Moya, A., 2002. Extreme genome reduction in *Buchnera* spp.: Toward the minimal genome needed for symbiotic life. *Proc. Natl. Acad. Sci.* 99, 4454–4458. <https://doi.org/10.1073/pnas.062067299>
- Gilbert, S.F., Sapp, J., Tauber, A.I., 2012. A Symbiotic View of Life: We Have Never Been Individuals. *Q. Rev. Biol.* 87, 325–341. <https://doi.org/10.1086/668166>
- Giorgini, M., Bernardo, U., Monti, M.M., Nappo, A.G., Gebiola, M., 2010. *Rickettsia* Symbionts Cause Parthenogenetic Reproduction in the Parasitoid Wasp *Pnigalio soemius* (Hymenoptera: Eulophidae). *Appl. Environ. Microbiol.* 76, 2589–2599. <https://doi.org/10.1128/AEM.03154-09>
- Giron, D., Dedeine, F., Dubreuil, G., Huguet, E., Mouton, L., Outreman, Y., Vavre, F., Simon, J.-C., 2017. Influence of Microbial Symbionts on Plant–Insect Interactions, in: *Advances in Botanical Research*. Elsevier, pp. 225–257. <https://doi.org/10.1016/bs.abr.2016.09.007>
- Gómez-Valero, L., Silva, F.J., Christophe Simon, J., Latorre, A., 2007. Genome reduction of the aphid endosymbiont *Buchnera aphidicola* in a recent evolutionary time scale. *Gene* 389, 87–95. <https://doi.org/10.1016/j.gene.2006.10.001>
- Gosalbes, M.J., Lamelas, A., Moya, A., Latorre, A., 2008a. The Striking Case of Tryptophan Provision in the Cedar Aphid *Cinara cedri*. *J. Bacteriol.* 190, 6026–6029. <https://doi.org/10.1128/JB.00525-08>
- Gosalbes, M.J., Lamelas, A., Moya, A., Latorre, A., 2008b. The striking case of tryptophan provision in the cedar aphid *Cinara cedri*. *J. Bacteriol.* 190, 6026–6029. <https://doi.org/10.1128/JB.00525-08>
- Goto, S., Anbutsu, H., Fukatsu, T., 2006. Asymmetrical Interactions between *Wolbachia* and *Spiroplasma* Endosymbionts Coexisting in the Same Insect Host. *Appl. Environ. Microbiol.* 72, 4805–4810. <https://doi.org/10.1128/AEM.00416-06>
- Gough, E., Shaikh, H., Manges, A.R., 2011. Systematic Review of Intestinal Microbiota Transplantation (Fecal Bacteriotherapy) for Recurrent *Clostridium difficile* Infection. *Clin. Infect. Dis.* 53, 994–1002. <https://doi.org/10.1093/cid/cir632>

- Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol. Biol. Evol.* 27, 221–224. <https://doi.org/10.1093/molbev/msp259>
- Haine, E.R., 2008. Symbiont-mediated protection. *Proc. R. Soc. B Biol. Sci.* 275, 353–361. <https://doi.org/10.1098/rspb.2007.1211>
- Hansen, A.K., Moran, N.A., 2014. The impact of microbial symbionts on host plant utilization by herbivorous insects. *Mol. Ecol.* 23, 1473–1496. <https://doi.org/10.1111/mec.12421>
- Hansen, A.K., Moran, N.A., 2011. Aphid genome expression reveals host–symbiont cooperation in the production of amino acids. *Proc. Natl. Acad. Sci.* 108, 2849–2854. <https://doi.org/10.1073/pnas.1013465108>
- Heddi, A., Charles, H., Khatchadourian, C., Bonnot, G., Nardon, P., 1998. Molecular Characterization of the Principal Symbiotic Bacteria of the Weevil *Sitophilus oryzae*: A Peculiar G + C Content of an Endocytobiotic DNA. *J. Mol. Evol.* 47, 52–61. <https://doi.org/10.1007/PL00006362>
- Heddi, A., Grenier, A.-M., Khatchadourian, C., Charles, H., Nardon, P., 1999. Four intracellular genomes direct weevil biology: Nuclear, mitochondrial, principal endosymbiont, and *Wolbachia*. *Proc. Natl. Acad. Sci.* 96, 6814–6819. <https://doi.org/10.1073/pnas.96.12.6814>
- Henry, L.M., Peccoud, J., Simon, J.-C., Hadfield, J.D., Maiden, M.J.C., Ferrari, J., Godfray, H.C.J., 2013. Horizontally Transmitted Symbionts and Host Colonization of Ecological Niches. *Curr. Biol.* 23, 1713–1717. <https://doi.org/10.1016/j.cub.2013.07.029>
- Herre, E.A., Mejía, L.C., Kylo, D.A., Rojas, E., Maynard, Z., Butler, A., Bael, S.A.V., 2007. Ecological Implications of Anti-Pathogen Effects of Tropical Fungal Endophytes and Mycorrhizae. *Ecology* 88, 550–558.
- Hilgenboecker, K., Hammerstein, P., Schlattmann, P., Telschow, A., Werren, J.H., 2008. How many species are infected with *Wolbachia*? – a statistical analysis of current data: *Wolbachia* infection rates. *FEMS Microbiol. Lett.* 281, 215–220. <https://doi.org/10.1111/j.1574-6968.2008.01110.x>
- Hooper, L.V., Wong, M.H., Thelin, A., Hansson, L., Falk, P.G., Gordon, J.I., 2001. Molecular Analysis of Commensal Host-Microbial Relationships in the Intestine. *Sci. New Ser.* 291, 881–884.
- Hopper, K.R., Kuhn, K.L., Lanier, K., Rhoades, J.H., Oliver, K.M., White, J.A., Asplen, M.K., Heimpel, G.E., 2018. The defensive aphid symbiont *Hamiltonella defensa* affects host quality differently for *Aphelinus glycinis* versus *Aphelinus atriplicis*. *Biol. Control* 116, 3–9. <https://doi.org/10.1016/j.biocontrol.2017.05.008>
- Hosokawa, T., Fukatsu, T., 2020. Relevance of microbial symbiosis to insect behavior. *Curr. Opin. Insect Sci.* 39, 91–100. <https://doi.org/10.1016/j.cois.2020.03.004>
- Hosokawa, T., Nikoh, N., Koga, R., Satô, M., Tanahashi, M., Meng, X.Y., Fukatsu, T., 2012. Reductive genome evolution, host-symbiont co-speciation and uterine transmission of endosymbiotic bacteria in bat flies. *ISME J.* 6, 577–587. <https://doi.org/10.1038/ismej.2011.125>
- Husník, F., Chrudimský, T., Hypša, V., 2011. Multiple origins of endosymbiosis within the Enterobacteriaceae (γ -Proteobacteria): convergence of complex phylogenetic approaches. *BMC Biol.* 9, 87. <https://doi.org/10.1186/1741-7007-9-87>
- Husník, F., McCutcheon, J.P., 2016. Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *Proc. Natl. Acad. Sci.* 113. <https://doi.org/10.1073/pnas.1603910113>

- Husnik, F., Nikoh, N., Koga, R., Ross, L., Duncan, R.P., Fujie, M., Tanaka, M., Satoh, N., Bachtrog, D., Wilson, A.C.C., von Dohlen, C.D., Fukatsu, T., McCutcheon, J.P., 2013. Horizontal Gene Transfer from Diverse Bacteria to an Insect Genome Enables a Tripartite Nested Mealybug Symbiosis. *Cell* 153, 1567–1578. <https://doi.org/10.1016/j.cell.2013.05.040>
- Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119. <https://doi.org/10.1186/1471-2105-11-119>
- Jamin, A.R., Vorburger, C., 2019. Estimating costs of aphid resistance to parasitoids conferred by a protective strain of the bacterial endosymbiont *Regiella insecticola*. *Entomol. Exp. Appl.* 167, 252–260. <https://doi.org/10.1111/eea.12749>
- Janson, E.M., Stireman, J.O., Singer, M.S., Abbot, P., 2008. phytophagous insect-microbe mutualisms and adaptative evolutionary diversification. *Evolution* 62, 997–1012. <https://doi.org/10.1111/j.1558-5646.2008.00348.x>
- Jousselin, E., Clamens, A.-L., Galan, M., Bernard, M., Maman, S., Gschloessl, B., Duport, G., Meseguer, A.S., Calevro, F., Coeur d’acier, A., 2016. Assessment of a 16S rRNA amplicon Illumina sequencing procedure for studying the microbiome of a symbiont-rich aphid genus. *Mol. Ecol. Resour.* 16, 628–640. <https://doi.org/10.1111/1755-0998.12478>
- Jousselin, E., Cœur d’Acier, A., Vanlerberghe-Masutti, F., Duron, O., 2013. Evolution and diversity of *Arsenophonus* endosymbionts in aphids. *Mol. Ecol.* 22, 260–270. <https://doi.org/10.1111/mec.12092>
- Jousselin, E., Desdevises, Y., Coeur d’acier, A., 2009. Fine-scale cospeciation between *Brachycaudus* and *Buchnera aphidicola*: bacterial genome helps define species and evolutionary relationships in aphids. *Proc. R. Soc. B Biol. Sci.* 276, 187–196. <https://doi.org/10.1098/rspb.2008.0679>
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. <https://doi.org/10.1038/nmeth.4285>
- Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30, 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kinjo, Y., Lo, N., Martín, P.V., Tokuda, G., Pigolotti, S., Bourguignon, T., 2021. Enhanced Mutation Rate, Relaxed Selection, and the “Domino Effect” are associated with Gene Loss in *Blattabacterium*, A Cockroach Endosymbiont. *Mol. Biol. Evol.* 38, 3820–3831. <https://doi.org/10.1093/molbev/msab159>
- Klasson, L., Andersson, S.G.E., 2006. Strong Asymmetric Mutation Bias in Endosymbiont Genomes Coincide with Loss of Genes for Replication Restart Pathways. *Mol. Biol. Evol.* 23, 1031–1039. <https://doi.org/10.1093/molbev/msj107>
- Koga, R., Bennett, G.M., Cryan, J.R., Moran, N.A., 2013. Evolutionary replacement of obligate symbionts in an ancient and diverse insect lineage: Replacement of symbionts in sap-feeding insects. *Environ. Microbiol.* 15, 2073–2081. <https://doi.org/10.1111/1462-2920.12121>
- Koga, R., Meng, X.-Y., Tsuchida, T., Fukatsu, T., 2012. Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. *Proc. Natl. Acad. Sci.* 109, E1230–E1237. <https://doi.org/10.1073/pnas.1119212109>

- Koga, R., Moran, N.A., 2014. Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. *ISME J.* 8, 1237–1246. <https://doi.org/10.1038/ismej.2013.235>
- Koga, R., Tsuchida, T., Fukatsu, T., 2009. Quenching autofluorescence of insect tissues for in situ detection of endosymbionts. *Appl. Entomol. Zool.* 44, 281–291. <https://doi.org/10.1303/aez.2009.281>
- Komaki, K., Ishikawa, H., 2000. Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. *Insect Biochem. Mol. Biol.* 30, 253–258. [https://doi.org/10.1016/S0965-1748\(99\)00125-3](https://doi.org/10.1016/S0965-1748(99)00125-3)
- Komaki, K., Ishikawa, H., 1999. Intracellular Bacterial Symbionts of Aphids Possess Many Genomic Copies per Bacterium. *J. Mol. Evol.* 48, 717–722. <https://doi.org/10.1007/PL00006516>
- Kremer, N., Charif, D., Henri, H., Bataille, M., Prévost, G., Kraaijeveld, K., Vavre, F., 2009. A new case of *Wolbachia* dependence in the genus *Asobara*: evidence for parthenogenesis induction in *Asobara japonica*. *Heredity* 103, 248–256. <https://doi.org/10.1038/hdy.2009.63>
- Kuo, C.-H., Ochman, H., 2009. Deletional Bias across the Three Domains of Life. *Genome Biol. Evol.* 1, 145–152. <https://doi.org/10.1093/gbe/evp016>
- Lamelas, A., Gosalbes, M.J., Manzano-Marín, A., Peretó, J., Moya, A., Latorre, A., 2011a. *Serratia symbiotica* from the Aphid *Cinara cedri*: A Missing Link from Facultative to Obligate Insect Endosymbiont. *PLoS Genet.* 7, e1002357. <https://doi.org/10.1371/journal.pgen.1002357>
- Lamelas, A., Gosalbes, M.J., Moya, A., Latorre, A., 2011b. New Clues about the Evolutionary History of Metabolic Losses in Bacterial Endosymbionts, Provided by the Genome of *Buchnera aphidicola* from the Aphid *Cinara tujafilina*. *Appl. Environ. Microbiol.* 77, 4446–4454. <https://doi.org/10.1128/AEM.00141-11>
- Lamelas, A., Pérez-Brocail, V., Gómez-Valero, L., Gosalbes, M.J., Moya, A., Latorre, A., 2008. Evolution of the Secondary Symbiont “*Candidatus Serratia symbiotica*” in Aphid Species of the Subfamily Lachninae. *Appl. Environ. Microbiol.* 74, 4236–4240. <https://doi.org/10.1128/AEM.00022-08>
- Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2016. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Mol. Biol. Evol.* msw260. <https://doi.org/10.1093/molbev/msw260>
- Lanning, D., Rhee, K., Knight, K., 2005. Intestinal bacteria and development of the B-lymphocyte repertoire. *Trends Immunol.* 26, 419–425. <https://doi.org/10.1016/j.it.2005.06.001>
- Lartillot, N., Lepage, T., Blanquart, S., 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25, 2286–2288. <https://doi.org/10.1093/bioinformatics/btp368>
- Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.-M., Kennedy, S., Leonard, P., Li, J., Burgdorf, K., Garup, N., Jørgensen, T., Brandslund, I., Nielsen, H.B., Juncker, A.S., Bertalan, M., Levenez, F., Pons, N., Rasmussen, S., Sunagawa, S., Tap, J., Tims, S., Zoetendal, E.G., Brunak, S., Clément, K., Doré, J., Kleerebezem, M., Kristiansen, K., Renault, P., Sicheritz-Ponten, T., de Vos, W.M., Zucker, J.-D., Raes, J., Hansen, T., MetaHIT consortium, Guedon, E., Delorme, C., Layec, S., Khaci, G., van de Guchte, M., Vandemeulebrouck, G., Jamet, A., Dervyn, R., Sanchez, N., Maguin, E., Haimet, F., Winogradski, Y., Cultrone, A., Leclerc, M., Juste, C., Blottière, H., Pelletier, E., LePaslier, D., Artiguenave, F., Bruls, T., Weissenbach, J., Turner, K., Parkhill, J., Antolin, M.,

- Manichanh, C., Casellas, F., Boruel, N., Varela, E., Torrejon, A., Guarner, F., Denariáz, G., Derrien, M., van Hylckama Vlieg, J.E.T., Veiga, P., Oozeer, R., Knol, J., Rescigno, M., Brechot, C., M'Rini, C., Mérieux, A., Yamada, T., Bork, P., Wang, J., Ehrlich, S.D., Pedersen, O., 2013. Richness of human gut microbiome correlates with metabolic markers. *Nature* 500, 541–546. <https://doi.org/10.1038/nature12506>
- Leigh, E.G., 2010. The evolution of mutualism: The evolution of mutualism. *J. Evol. Biol.* 23, 2507–2528. <https://doi.org/10.1111/j.1420-9101.2010.02114.x>
- Lenhart, P.A., White, J.A., 2017. A defensive endosymbiont fails to protect aphids against the parasitoid community present in the field: *Hamiltonella* provides no protection in the field. *Ecol. Entomol.* 42, 680–684. <https://doi.org/10.1111/een.12419>
- Leybourne, D.J., Bos, J.I.B., Valentine, T.A., Karley, A.J., 2020. The price of protection: a defensive endosymbiont impairs nymph growth in the bird cherry-oat aphid, *Rhopalosiphum padi*. *Insect Sci.* 27, 69–85. <https://doi.org/10.1111/1744-7917.12606>
- Lind, P.A., Andersson, D.I., 2008. Whole-genome mutational biases in bacteria. *Proc. Natl. Acad. Sci.* 105, 17878–17883. <https://doi.org/10.1073/pnas.0804445105>
- Liu, L., Huang, X., Zhang, R., Jiang, L., Qiao, G., 2013. Phylogenetic congruence between *Mollitrichosiphum* (Aphididae: Greenideinae) and *Buchnera* indicates insect-bacteria parallel evolution. *Syst. Entomol.* 38, 81–92. <https://doi.org/10.1111/j.1365-3113.2012.00647.x>
- Liu, L., Li, L., Liu, Jiannan, Yu, Z., Yang, X., Liu, Jingze, 2016. Population dynamics of multiple symbionts in the hard tick, *Dermacentor silvarum* Olenov (Acari: Ixodidae). *Ticks Tick-Borne Dis.* 7, 188–192. <https://doi.org/10.1016/j.ttbdis.2015.10.002>
- Login, F.H., Balmand, S., Vallier, A., Vincent-Monégat, C., Vigneron, A., Weiss-Gayet, M., Rochat, D., Heddi, A., 2011a. Antimicrobial peptides keep insect endosymbionts under control. *Science* 334, 362–365. <https://doi.org/10.1126/science.1209728>
- Login, F.H., Balmand, S., Vallier, A., Vincent-Monégat, C., Vigneron, A., Weiss-Gayet, M., Rochat, D., Heddi, A., 2011b. Antimicrobial Peptides Keep Insect Endosymbionts Under Control. *Science* 334, 362–365. <https://doi.org/10.1126/science.1209728>
- Login, F.H., Heddi, A., 2013. Insect immune system maintains long-term resident bacteria through a local response. *J. Insect Physiol.* 59, 232–239. <https://doi.org/10.1016/j.jinsphys.2012.06.015>
- López-Madrugal, S., Duarte, E.H., 2019. Titer regulation in arthropod-Wolbachia symbioses. *FEMS Microbiol. Lett.* 366, fnz232. <https://doi.org/10.1093/femsle/fnz232>
- López-Sánchez, M.J., Neef, A., Peretó, J., Patiño-Navarrete, R., Pignatelli, M., Latorre, A., Moya, A., 2009. Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. *PLoS Genet.* 5. <https://doi.org/10.1371/journal.pgen.1000721>
- Lu, P., Bian, G., Pan, X., Xi, Z., 2012. Wolbachia Induces Density-Dependent Inhibition to Dengue Virus in Mosquito Cells. *PLoS Negl. Trop. Dis.* 6, e1754. <https://doi.org/10.1371/journal.pntd.0001754>
- Lu, W.N., Chiu, M.C., Kuo, M.H., 2014. Host life stage- and temperature-dependent density of the symbiont *Buchnera aphidicola* in a subtropical pea aphid (*Acyrtosiphon pisum*) population. *J. Asia-Pac. Entomol.* 17, 537–541. <https://doi.org/10.1016/j.aspen.2014.03.012>

- Luan, J.-B., Chen, W., Hasegawa, D.K., Simmons, A.M., Wintermantel, W.M., Ling, K.-S., Fei, Z., Liu, S.-S., Douglas, A.E., 2015. Metabolic Coevolution in the Bacterial Symbiosis of Whiteflies and Related Plant Sap-Feeding Insects. *Genome Biol. Evol.* 7, 2635–2647. <https://doi.org/10.1093/gbe/evv170>
- Luan, J.-B., Shan, H.-W., Isermann, P., Huang, J.-H., Lammerding, J., Liu, S.-S., Douglas, A.E., 2016. Cellular and molecular remodelling of a host cell for vertical transmission of bacterial symbionts. *Proc. R. Soc. B Biol. Sci.* 283, 20160580. <https://doi.org/10.1098/rspb.2016.0580>
- Lynch, M., 2006. Streamlining and Simplification of Microbial Genome Architecture. *Annu. Rev. Microbiol.* 60, 327–349. <https://doi.org/10.1146/annurev.micro.60.080805.142300>
- Lynn-Bell, N.L., Strand, M.R., Oliver, K.M., 2019. Bacteriophage acquisition restores protective mutualism. *Microbiology* 165, 985–989. <https://doi.org/10.1099/mic.0.000816>
- Macdonald, S.J., Thomas, G.H., Douglas, A.E., 2011. Genetic and metabolic determinants of nutritional phenotype in an insect-bacterial symbiosis: phenotype of insect-bacterial symbiosis. *Mol. Ecol.* 20, 2073–2084. <https://doi.org/10.1111/j.1365-294X.2011.05031.x>
- Manzano-marín, 2022. Co-obligate symbioses have repeatedly evolved across aphids, but partner identity and nutritional contributions vary across lineages.
- Manzano-Marín, A., Coeur d’acier, A., Clamens, A.-L., Orvain, C., Cruaud, C., Barbe, V., Jousset, E., 2018. A Freeloader? The Highly Eroded Yet Large Genome of the *Serratia symbiotica* Symbiont of *Cinara strobilifera*. *Genome Biol. Evol.* 10, 2178–2189. <https://doi.org/10.1093/gbe/evy173>
- Manzano-Marín, A., Latorre, A., 2016. Snapshots of a shrinking partner: Genome reduction in *Serratia symbiotica*. *Sci. Rep.* 6, 32590. <https://doi.org/10.1038/srep32590>
- Manzano-Marín, A., 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 Biol. Evol.* 6, 1683–1698. <https://doi.org/10.1093/gbe/evu133>
- Manzano-Marín, A., Ocegüera-Figueroa, A., Latorre, A., Jiménez-García, L.F., Moya, A., 2015. Solving a Bloody Mess: B-Vitamin Independent Metabolic Convergence among Gammaproteobacterial Obligate Endosymbionts from Blood-Feeding Arthropods and the Leech *Haementeria officinalis*. *Genome Biol. Evol.* 7, 2871–2884. <https://doi.org/10.1093/gbe/evv188>
- Manzano-Marín, A., Simon, J.-C., Latorre, A., 2016. Reinventing the Wheel and Making It Round Again: Evolutionary Convergence in *Buchnera* – *Serratia* Symbiotic Consortia between the Distantly Related Lachninae Aphids *Tuberochnebrionus salignus* and *Cinara cedri*. *Genome Biol. Evol.* 8, 1440–1458. <https://doi.org/10.1093/gbe/evw085>
- Manzano-Marín, A., Szabó, G., Simon, J., Horn, M., Latorre, A., 2017. Happens in the best of subfamilies: establishment and repeated replacements of co-obligate secondary endosymbionts within Lachninae aphids. *Environ. Microbiol.* 19, 393–408. <https://doi.org/10.1111/1462-2920.13633>
- Manzano-Marín, A., Szabó, G., Simon, J.C., Horn, M., Latorre, A., 2017. Happens in the best of subfamilies: establishment and repeated replacements of co-obligate secondary endosymbionts within Lachninae aphids. *Environ. Microbiol.* 19, 393–408. <https://doi.org/10.1111/1462-2920.13633>
- Manzano-Marín, A., Coeur d’acier, A., Clamens, A.-L., Orvain, C., Cruaud, C., Barbe, V., Jousset, E., 2020a. Serial horizontal transfer of vitamin-biosynthetic genes enables the establishment of

- new nutritional symbionts in aphids' di-symbiotic systems. *ISME J.* 14, 259–273. <https://doi.org/10.1038/s41396-019-0533-6>
- Manzano-Marín, A., Coeur d'acier, A., Clamens, A.L., Orvain, C., Cruaud, C., Barbe, V., Jousset, E., 2020b. Serial horizontal transfer of vitamin-biosynthetic genes enables the establishment of new nutritional symbionts in aphids' di-symbiotic systems. *ISME J.* 14, 259–273. <https://doi.org/10.1038/s41396-019-0533-6>
- Mao, M., Yang, X., Bennett, G.M., 2018. Evolution of host support for two ancient bacterial symbionts with differentially degraded genomes in a leafhopper host. *Proc. Natl. Acad. Sci.* 115, E11691–E11700. <https://doi.org/10.1073/pnas.1811932115>
- Margulis, L., 1981. *Symbiosis in cell evolution: Life and its environment on the early earth.*
- Margulis L., T., Fester R., 1991. Symbiosis as a source of evolutionary innovation: Speciation and morphogenesis. *Camb. Mass. MIT Press* 7, 422–423. [https://doi.org/10.1016/0169-5347\(92\)90028-A](https://doi.org/10.1016/0169-5347(92)90028-A)
- Martin, B.D., Schwab, E., 2012. Current Usage of Symbiosis and Associated Terminology. *Int. J. Biol.* 5, p32. <https://doi.org/10.5539/ijb.v5n1p32>
- Mazmanian, S.K., Liu, C.H., Tzianabos, A.O., Kasper, D.L., 2005. An Immunomodulatory Molecule of Symbiotic Bacteria Directs Maturation of the Host Immune System 12.
- McCutcheon, J.P., Boyd, B.M., Dale, C., 2019. The Life of an Insect Endosymbiont from the Cradle to the Grave. *Curr. Biol.* 29, R485–R495. <https://doi.org/10.1016/j.cub.2019.03.032>
- McCutcheon, J.P., McDonald, B.R., Moran, N.A., 2009. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc. Natl. Acad. Sci.* 106, 15394–15399. <https://doi.org/10.1073/pnas.0906424106>
- McCutcheon, J.P., Moran, N.A., 2012. Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10, 13–26. <https://doi.org/10.1038/nrmicro2670>
- McCutcheon, J.P., Moran, N.A., 2010. Functional Convergence in Reduced Genomes of Bacterial Symbionts Spanning 200 My of Evolution. *Genome Biol. Evol.* 2, 708–718. <https://doi.org/10.1093/gbe/evq055>
- McCutcheon, J.P., Moran, N.A., 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc. Natl. Acad. Sci.* 104, 19392–19397. <https://doi.org/10.1073/pnas.0708855104>
- McCutcheon, J.P., von Dohlen, C.D., 2011. An Interdependent Metabolic Patchwork in the Nested Symbiosis of Mealybugs. *Curr. Biol.* 21, 1366–1372. <https://doi.org/10.1016/j.cub.2011.06.051>
- McFall-Ngai, M., Hadfield, M.G., Bosch, T.C.G., Carey, H.V., Domazet-Lošo, T., Douglas, A.E., Dübilier, N., Eberl, G., Fukami, T., Gilbert, S.F., Hentschel, U., King, N., Kjelleberg, S., Knoll, A.H., Kremer, N., Mazmanian, S.K., Metcalf, J.L., Neelson, K., Pierce, N.E., Rawls, J.F., Reid, A., Ruby, E.G., Rumpho, M., Sanders, J.G., Tautz, D., Wernegreen, J.J., 2013. Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad. Sci.* 110, 3229–3236. <https://doi.org/10.1073/pnas.1218525110>
- McFall-Ngai, M., Heath-Heckman, E.A.C., Gillette, A.A., Peyer, S.M., Harvie, E.A., 2012. The secret languages of coevolved symbioses: Insights from the *Euprymna scolopes*–*Vibrio fischeri* symbiosis. *Semin. Immunol.* 24, 3–8. <https://doi.org/10.1016/j.smim.2011.11.006>

- Mendiola, S.Y., Civitello, D.J., Gerardo, N.M., 2020. An integrative approach to symbiont-mediated vector control for agricultural pathogens. *Curr. Opin. Insect Sci.* 39, 57–62. <https://doi.org/10.1016/j.cois.2020.02.007>
- Meseguer, A.S., Coeur d’acier, A., Genson, G., Joussetin, E., 2015. Unravelling the historical biogeography and diversification dynamics of a highly diverse conifer-feeding aphid genus. *J. Biogeogr.* 42, 1482–1492. <https://doi.org/10.1111/jbi.12531>
- Meseguer, A.S., Manzano-Marín, A., Coeur d’Acier, A., Clamens, A.-L., Godefroid, M., Joussetin, E., 2017. Buchnera has changed flatmate but the repeated replacement of co-obligate symbionts is not associated with the ecological expansions of their aphid hosts. *Mol. Ecol.* 26, 2363–2378. <https://doi.org/10.1111/mec.13910>
- Meseguer, Manzano-Marín, Coeur d’Acier, Clamens, Godefroid, Joussetin, 2017. Buchnera has changed flatmate but the repeated replacement of co-obligate symbionts is not associated with the ecological expansions of their aphid hosts. *Mol. Ecol.* 26, 2363–2378. <https://doi.org/10.1111/mec.13910>
- Michalik, A., 2010. Ovary structure and transovarial transmission of endosymbiotic microorganisms in *Clethrobium comes*, *Myzocallis walshii* and *Sipha maydis* /Hemiptera, Aphididae: Drepanosiphinae/ 8.
- Michalik, A., Franco, D.C., Kobińska, M., Szklarzewicz, T., Stroiński, A., Łukasik, P., 2021. Alternative transmission patterns in independently acquired nutritional co-symbionts of Dictyopharidae planthoppers 19.
- Michalik, A., Szklarzewicz, T., Jankowska, W., Wieczorek, K., 2014. Endosymbiotic microorganisms of aphids (Hemiptera: Sternorrhyncha: Aphidoidea): Ultrastructure, distribution and transovarial transmission. *Eur. J. Entomol.* 111, 91–104. <https://doi.org/10.14411/eje.2014.011>
- Mira, A., Ochman, H., Moran, N.A., 2001. Deletional bias and the evolution of bacterial genomes. *Trends Genet.* 17, 589–596. [https://doi.org/10.1016/S0168-9525\(01\)02447-7](https://doi.org/10.1016/S0168-9525(01)02447-7)
- Mittler, T.E., 1971. Dietary Amino Acid Requirements of the Aphid *Myzus persicae* Affected by Antibiotic Uptake. *J. Nutr.* 101, 1023–1028. <https://doi.org/10.1093/jn/101.8.1023>
- Monnin, D., Jackson, R., Kiers, E.T., Bunker, M., Ellers, J., Henry, L.M., 2020. Parallel Evolution in the Integration of a Co-obligate Aphid Symbiosis. *Curr. Biol.* 30, 1949-1957.e6. <https://doi.org/10.1016/j.cub.2020.03.011>
- Montllor, C.B., Maxmen, A., Purcell, A.H., 2002. Facultative bacterial endosymbionts benefit pea aphids *Acyrtosiphon pisum* under heat stress: Aphid symbionts and heat stress. *Ecol. Entomol.* 27, 189–195. <https://doi.org/10.1046/j.1365-2311.2002.00393.x>
- Moran, 1993. A Molecular Clock in Endosymbiotic Bacteria Is Calibrated Using the Insect.
- Moran, N.A., 2007. Symbiosis as an adaptive process and source of phenotypic complexity. *Proc. Natl. Acad. Sci.* 104, 8627–8633. <https://doi.org/10.1073/pnas.0611659104>
- Moran, N.A., 2002. Microbial Minimalism. *Cell* 108, 583–586. [https://doi.org/10.1016/S0092-8674\(02\)00665-7](https://doi.org/10.1016/S0092-8674(02)00665-7)
- Moran, N.A., 1996. Accelerated evolution and Muller’s ratchet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci.* 93, 2873–2878. <https://doi.org/10.1073/pnas.93.7.2873>

- Moran, N.A., Degnan, P.H., Santos, S.R., Dunbar, H.E., Ochman, H., 2005. The players in a mutualistic symbiosis: Insects, bacteria, viruses, and virulence genes. *Proc. Natl. Acad. Sci.* 102, 16919–16926. <https://doi.org/10.1073/pnas.0507029102>
- Moran, N.A., McCutcheon, J.P., Nakabachi, A., 2008a. Genomics and Evolution of Heritable Bacterial Symbionts. *Annu. Rev. Genet.* 42, 165–190. <https://doi.org/10.1146/annurev.genet.41.110306.130119>
- Moran, N.A., McCutcheon, J.P., Nakabachi, A., 2008b. Genomics and Evolution of Heritable Bacterial Symbionts. *Annu. Rev. Genet.* 42, 165–190. <https://doi.org/10.1146/annurev.genet.41.110306.130119>
- Moran, N.A., McLaughlin, H.J., Sorek, R., 2009. The Dynamics and Time Scale of Ongoing Genomic Erosion in Symbiotic Bacteria. *Science* 323, 379–382. <https://doi.org/10.1126/science.1167140>
- Moran, N.A., Mira, A., 2001. The process of genome shrinkage in the obligate symbiont 12.
- Moran, N.A., Sloan, D.B., 2015. The Hologenome Concept: Helpful or Hollow? *PLOS Biol.* 13, e1002311. <https://doi.org/10.1371/journal.pbio.1002311>
- Moran, N.A., Wernegreen, J.J., 2000. Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends Ecol. Evol.* 15, 321–326. [https://doi.org/10.1016/S0169-5347\(00\)01902-9](https://doi.org/10.1016/S0169-5347(00)01902-9)
- Moreira, D., Lopez-Garcia, P., 1998. Symbiosis Between Methanogenic Archaea and α -Proteobacteria as the Origin of Eukaryotes: The Syntrophic Hypothesis 47, 517–530.
- Morris, J.J., Lenski, R.E., Zinser, E.R., 2012. The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. *mBio* 3, e00036-12. <https://doi.org/10.1128/mBio.00036-12>
- 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. <https://doi.org/10.1038/nrg2319>
- Muñoz-Gómez, S.A., Susko, E., Williamson, K., Eme, L., Slamovits, C.H., Moreira, D., López-García, P., Roger, A.J., 2022. Site-and-branch-heterogeneous analyses of an expanded dataset favour mitochondria as sister to known Alphaproteobacteria. *Nat. Ecol. Evol.* 6, 253–262. <https://doi.org/10.1038/s41559-021-01638-2>
- Muñoz-Gómez, S.A., Wideman, J.G., Roger, A.J., Slamovits, C.H., 2017. The origin of mitochondrial cristae from alphaproteobacteria. *Mol. Biol. Evol.* msw298. <https://doi.org/10.1093/molbev/msw298>
- Munson, M.A., Baumann, P., Clark, M.A., Baumann, L., Moran, N.A., Voegtlin, D.J., Campbell, B.C., 1991. Evidence for the Establishment of Aphid-Eubacterium Endosymbiosis in an Ancestor of Four Aphid Families. *J BACTERIOL.* 4.
- Murray, G.G.R., Weinert, L.A., Rhule, E.L., Welch, J.J., 2016. The Phylogeny of *Rickettsia* Using Different Evolutionary Signatures: How Tree-Like is Bacterial Evolution? *Syst. Biol.* 65, 265–279. <https://doi.org/10.1093/sysbio/syv084>
- Murrell, B., Weaver, S., Smith, M.D., Wertheim, J.O., Murrell, S., Aylward, A., Eren, K., Pollner, T., Martin, D.P., Smith, D.M., Scheffler, K., Kosakovsky Pond, S.L., 2015. Gene-Wide Identification of Episodic Selection. *Mol. Biol. Evol.* 32, 1365–1371. <https://doi.org/10.1093/molbev/msv035>

- Nakabachi, A., Ishikawa, H., 1999. Provision of riboflavin to the host aphid, *Acyrtosiphon pisum*, by endosymbiotic bacteria, *Buchnera*. *J. Insect Physiol.* 45, 1–6. [https://doi.org/10.1016/S0022-1910\(98\)00104-8](https://doi.org/10.1016/S0022-1910(98)00104-8)
- Nakabachi, A., Yamashita, A., Toh, H., Ishikawa, H., Dunbar, H.E., Moran, N.A., Hattori, M., 2006. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* 314, 267. <https://doi.org/10.1126/science.1134196>
- Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* 32, 268–274. <https://doi.org/10.1093/molbev/msu300>
- Nishikori, K., Morioka, K., Kubo, T., Morioka, M., 2009. Age- and morph-dependent activation of the lysosomal system and *Buchnera* degradation in aphid endosymbiosis. *J. Insect Physiol.* 55, 351–357. <https://doi.org/10.1016/j.jinsphys.2009.01.001>
- Nováková, E., Hypša, V., 2007. A new *Sodalis* lineage from bloodsucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*. *FEMS Microbiol. Lett.* 269, 131–135. <https://doi.org/10.1111/j.1574-6968.2006.00620.x>
- Nováková, E., Hypša, V., Moran, N.A., 2009. Arsenophonus, an emerging clade of intracellular symbionts with a broad host distribution. *BMC Microbiol.* 9, 143. <https://doi.org/10.1186/1471-2180-9-143>
- Ochman, H., Elwyn, S., Moran, N.A., 1999. Calibrating bacterial evolution 6.
- 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. *Annu. Rev. Entomol.* 55, 247–266. <https://doi.org/10.1146/annurev-ento-112408-085305>
- Oliver, K.M., Degnan, P.H., Hunter, M.S., Moran, N.A., 2009. Bacteriophages Encode Factors Required for Protection in a Symbiotic Mutualism. *Science* 325, 992–994. <https://doi.org/10.1126/science.1174463>
- Oliver, K.M., Higashi, C.H., 2019. Variations on a protective theme: *Hamiltonella defensa* infections in aphids variably impact parasitoid success. *Curr. Opin. Insect Sci.* 32, 1–7. <https://doi.org/10.1016/j.cois.2018.08.009>
- Oliver, K.M., Moran, N.A., Hunter, M.S., 2006. Costs and benefits of a superinfection of facultative symbionts in aphids. *Proc. R. Soc. B Biol. Sci.* 273, 1273–1280. <https://doi.org/10.1098/rspb.2005.3436>
- Oliver, K.M., Russell, J.A., Moran, N.A., Hunter, M.S., 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proc. Natl. Acad. Sci.* 100, 1803–1807. <https://doi.org/10.1073/pnas.0335320100>
- Olivier, H.M., Moon, B.R., 2010. The effects of atrazine on spotted salamander embryos and their symbiotic alga. *Ecotoxicology* 19, 654–661. <https://doi.org/10.1007/s10646-009-0437-8>
- Osborne, S.E., Iturbe-Ormaetxe, I., Brownlie, J.C., O'Neill, S.L., Johnson, K.N., 2012. Antiviral Protection and the Importance of *Wolbachia* Density and Tissue Tropism in *Drosophila simulans*. *Appl. Environ. Microbiol.* 78, 6922–6929. <https://doi.org/10.1128/AEM.01727-12>

- Pannebakker, B.A., Loppin, B., Elemans, C.P.H., Humblot, L., Vavre, F., 2007. Parasitic inhibition of cell death facilitates symbiosis. *Proc. Natl. Acad. Sci.* 104, 213–215. <https://doi.org/10.1073/pnas.0607845104>
- Parker, B.J., 2021. Mechanisms and Evolution of Heritable Microbial Density in Insect Hosts. *mSystems* 6. <https://doi.org/10.1128/mSystems.00728-21>
- Patiño-Navarrete, R., Moya, A., Latorre, A., Peretó, J., 2013. Comparative Genomics of *Blattabacterium cuenoti*: The Frozen Legacy of an Ancient Endosymbiont Genome. *Genome Biol. Evol.* 5, 351–361. <https://doi.org/10.1093/gbe/evt011>
- Peccoud, J., Simon, J.-C., von Dohlen, C., Coeur d’acier, A., Plantegenest, M., Vanlerberghe-Masutti, F., Jousset, E., 2010. Evolutionary history of aphid-plant associations and their role in aphid diversification. *C. R. Biol.* 333, 474–487. <https://doi.org/10.1016/j.crv.2010.03.004>
- Pérez-Brocal, V., Gil, R., Moya, A., Latorre, A., 2011. New Insights on the Evolutionary History of Aphids and Their Primary Endosymbiont *Buchnera aphidicola*. *Int. J. Evol. Biol.* 2011, 1–9. <https://doi.org/10.4061/2011/250154>
- Pérez-Brocal, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J.M., Silva, F.J., Moya, A., Latorre, A., 2006. A Small Microbial Genome: The End of a Long Symbiotic Relationship? *Science* 314, 312–313. <https://doi.org/10.1126/science.1130441>
- Perotti, M.A., Clarke, H.K., Turner, B.D., Braig, H.R., Perotti, M.A., Clarke, H.K., Turner, B.D., Braig, H.R., 2006. Rickettsia as obligate and mycetozoa. *FASEB J.* 20, 2372–2374. <https://doi.org/10.1096/fj.06-5870fje>
- Perreau, J., Moran, N.A., 2021. Genetic innovations in animal–microbe symbioses. *Nat. Rev. Genet.* <https://doi.org/10.1038/s41576-021-00395-z>
- Perreau, J., Zhang, B., Maeda, G.P., Kirkpatrick, M., Moran, N.A., 2021. Strong within-host selection in a maternally inherited obligate symbiont: *Buchnera* and aphids. *Proc. Natl. Acad. Sci.* 118, e2102467118. <https://doi.org/10.1073/pnas.2102467118>
- Pers, D., Hansen, A.K., 2021. The boom and bust of the aphid’s essential amino acid metabolism across nymphal development. *G3: GenesGenomesGenetics* 11, jkab115. <https://doi.org/10.1093/g3journal/jkab115>
- Philippe, H., Roure, B., 2011. Difficult phylogenetic questions: more data, maybe; better methods, certainly. *BMC Biol.* 9, 91. <https://doi.org/10.1186/1741-7007-9-91>
- Philippot, L., Raaijmakers, J.M., Lemanceau, P., van der Putten, W.H., 2013. Going back to the roots: the microbial ecology of the rhizosphere. *Nat. Rev. Microbiol.* 11, 789–799. <https://doi.org/10.1038/nrmicro3109>
- Plague, G.R., Dunbar, H.E., Tran, P.L., Moran, N.A., 2008. Extensive Proliferation of Transposable Elements in Heritable Bacterial Symbionts. *J. Bacteriol.* 190, 777–779. <https://doi.org/10.1128/JB.01082-07>
- Poliakov, A., Russell, C.W., Ponnala, L., Hoops, H.J., Sun, Q., Douglas, A.E., van Wijk, K.J., 2011. Large-Scale Label-Free Quantitative Proteomics of the Pea aphid-Buchnera Symbiosis. *Mol. Cell. Proteomics* 10, M110.007039. <https://doi.org/10.1074/mcp.M110.007039>
- Polin, S., Le Gallic, J.-F., Simon, J.-C., Tsuchida, T., Outreman, Y., 2015. Conditional Reduction of Predation Risk Associated with a Facultative Symbiont in an Insect. *PLOS ONE* 10, e0143728. <https://doi.org/10.1371/journal.pone.0143728>

- Polin, S., Simon, J., Outreman, Y., 2014. An ecological cost associated with protective symbionts of aphids. *Ecol. Evol.* 4, 836–840. <https://doi.org/10.1002/ece3.991>
- Pond, S.L.K., Frost, S.D.W., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679. <https://doi.org/10.1093/bioinformatics/bti079>
- Pons, I., Renoz, F., Hance, T., 2019. Fitness costs of the cultivable symbiont *Serratia symbiotica* and its phenotypic consequences to aphids in presence of environmental stressors. *Evol. Ecol.* 33, 825–838. <https://doi.org/10.1007/s10682-019-10012-5>
- Pons, I., Scieur, N., Dhondt, L., Renard, M.-E., Renoz, F., Hance, T., 2021. Ubiquity of the Symbiont *Serratia symbiotica* in the Aphid Natural Environment: Distribution, Diversity and Evolution at a Multitrophic Level (preprint). *Evolutionary Biology*. <https://doi.org/10.1101/2021.04.18.440331>
- Prosser, W.A., Douglas, A.E., 1991. The aposymbiotic aphid: An analysis of chlortetracycline-treated pea aphid, *Acyrtosiphon pisum*. *J. Insect Physiol.* 37, 713–719. [https://doi.org/10.1016/0022-1910\(91\)90104-8](https://doi.org/10.1016/0022-1910(91)90104-8)
- Pyka-Fosciak, G., Szklarzewicz, T., 2008. Germ cell cluster formation and ovariole structure in viviparous and oviparous generations of the aphid *Stomaphis quercus*. *Int. J. Dev. Biol.* 522-3 259-265. <https://doi.org/10.1387/ijdb.072338gp>
- R Core Team, 2022. R: A language and environment for statistica computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. 12.
- Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* 67, 901–904. <https://doi.org/10.1093/sysbio/syy032>
- Ranwez, V., Douzery, E.J.P., Cambon, C., Chantret, N., Delsuc, F., 2018. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Mol. Biol. Evol.* 35, 2582–2584. <https://doi.org/10.1093/molbev/msy159>
- Ratzka, C., Gross, R., Feldhaar, H., 2012. Endosymbiont Tolerance and Control within Insect Hosts. *Insects* 3, 553–572. <https://doi.org/10.3390/insects3020553>
- Renoz, F., Lopes, M.R., Gaget, K., Duport, G., Eloy, M.-C., Geelhand de Merxem, B., Hance, T., Calevro, F., 2022. Compartmentalized into Bacteriocytes but Highly Invasive: the Puzzling Case of the Co-Obligate Symbiont *Serratia symbiotica* in the Aphid *Periphyllus lyropictus*. *Microbiol. Spectr.* e00457-22. <https://doi.org/10.1128/spectrum.00457-22>
- Rio, R.V.M., Wu, Y., Filardo, G., Aksoy, S., 2006. Dynamics of multiple symbiont density regulation during host development: tsetse fly and its microbial flora. *Proc. R. Soc. B Biol. Sci.* 273, 805–814. <https://doi.org/10.1098/rspb.2005.3399>
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst. Biol.* 61, 539–542. <https://doi.org/10.1093/sysbio/sys029>
- Rosenberg, E., Koren, O., Reshef, L., Efrony, R., Zilber-Rosenberg, I., 2007. The role of microorganisms in coral health, disease and evolution. *Nat. Rev. Microbiol.* 5, 355–362. <https://doi.org/10.1038/nrmicro1635>

- Rosenblueth, M., Sayavedra, L., Sámano-Sánchez, H., Roth, A., Martínez-Romero, E., 2012. Evolutionary relationships of flavobacterial and enterobacterial endosymbionts with their scale insect hosts (Hemiptera: Coccoidea). *J. Evol. Biol.* 25, 2357–2368. <https://doi.org/10.1111/j.1420-9101.2012.02611.x>
- Russell, C.W., Poliakov, A., Haribal, M., Jander, G., van Wijk, K.J., Douglas, A.E., 2014. Matching the supply of bacterial nutrients to the nutritional demand of the animal host. *Proc. R. Soc. B Biol. Sci.* 281, 20141163. <https://doi.org/10.1098/rspb.2014.1163>
- Russell, J.A., Moran, N.A., 2006. Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. *Proc. R. Soc. B Biol. Sci.* 273, 603–610. <https://doi.org/10.1098/rspb.2005.3348>
- Sakurai, M., Koga, R., Tsuchida, T., Meng, X.-Y., Fukatsu, T., 2005. *Rickettsia* Symbiont in the Pea Aphid *Acyrtosiphon pisum*: Novel Cellular Tropism, Effect on Host Fitness, and Interaction with the Essential Symbiont *Buchnera*. *Appl. Environ. Microbiol.* 71, 4069–4075. <https://doi.org/10.1128/AEM.71.7.4069-4075.2005>
- Sandström, J., Moran, N., 1999. How nutritionally imbalanced is phloem sap for aphids? *Entomol. Exp. Appl.* 91, 203–210. <https://doi.org/10.1046/j.1570-7458.1999.00485.x>
- Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. <https://doi.org/10.1093/bioinformatics/btr026>
- Serbus, L.R., Casper-Lindley, C., Landmann, F., Sullivan, W., 2008. The Genetics and Cell Biology of *Wolbachia* -Host Interactions. *Annu. Rev. Genet.* 42, 683–707. <https://doi.org/10.1146/annurev.genet.41.110306.130354>
- Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., Ishikawa, H., 2000a. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. *APS. Nature* 407, 81–86. <https://doi.org/10.1038/35024074>
- Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., Ishikawa, H., 2000b. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. *APS* 407, 6.
- Shigenobu, S., Wilson, A.C.C., 2011. Genomic revelations of a mutualism: the pea aphid and its obligate bacterial symbiont. *Cell. Mol. Life Sci.* 68, 1297–1309. <https://doi.org/10.1007/s00018-011-0645-2>
- Sicard, M., Bouchon, D., Ceyrac, L., Raimond, R., Thierry, M., Le Clec'h, W., Marcadé, I., Caubet, Y., Grève, P., 2014a. Bidirectional cytoplasmic incompatibility caused by *Wolbachia* in the terrestrial isopod *Porcellio dilatatus*. *J. Invertebr. Pathol.* 121, 28–36. <https://doi.org/10.1016/j.jip.2014.06.007>
- Sicard, M., Dittmer, J., Grève, P., Bouchon, D., Braquart-Varnier, C., 2014b. A host as an ecosystem: *Wolbachia* coping with environmental constraints: The *Wolbachia* in their 'host ecosystem.' *Environ. Microbiol.* 16, 3583–3607. <https://doi.org/10.1111/1462-2920.12573>
- Silva, F.J., Santos-Garcia, D., 2015. Slow and Fast Evolving Endosymbiont Lineages: Positive Correlation between the Rates of Synonymous and Non-Synonymous Substitution. *Front. Microbiol.* 6. <https://doi.org/10.3389/fmicb.2015.01279>
- Simonet, P., Duport, G., Gaget, K., Weiss-Gayet, M., Colella, S., Febvay, G., Charles, H., Viñuelas, J., Heddi, A., Calevro, F., 2016a. Direct flow cytometry measurements reveal a fine-tuning of symbiotic cell dynamics according to the host developmental needs in aphid symbiosis. *Sci. Rep.* 6, 1–13. <https://doi.org/10.1038/srep19967>

- Simonet, P., Duport, G., Gaget, K., Weiss-Gayet, M., Colella, S., Febvay, G., Charles, H., Viñuelas, J., Heddi, A., Calevro, F., 2016b. Direct flow cytometry measurements reveal a fine-tuning of symbiotic cell dynamics according to the host developmental needs in aphid symbiosis. *Sci. Rep.* 6, 19967. <https://doi.org/10.1038/srep19967>
- Simonet, P., Gaget, K., Balmand, S., Lopes, M.R., Parisot, N., Buhler, K., Duport, G., Vulsteke, V., Febvay, G., Heddi, A., Charles, H., Callaerts, P., Calevro, F., 2018a. Bacteriocyte cell death in the pea aphid/*Buchnera* symbiotic system. *Proc. Natl. Acad. Sci. U. S. A.* 115, E1819–E1827. <https://doi.org/10.1073/pnas.1720237115>
- Simonet, P., Gaget, K., Balmand, S., Ribeiro Lopes, M., Parisot, N., Buhler, K., Duport, G., Vulsteke, V., Febvay, G., Heddi, A., Charles, H., Callaerts, P., Calevro, F., 2018b. Bacteriocyte cell death in the pea aphid/ *Buchnera* symbiotic system. *Proc. Natl. Acad. Sci.* 115, E1819–E1828. <https://doi.org/10.1073/pnas.1720237115>
- Six, D. L., & Klepzig, K. D. (2004). Dendroctonus bark beetles as model systems for studies on symbiosis. *Symbiosis*.
- Sloan, D.B., Moran, N.A., 2012. Genome Reduction and Co-evolution between the Primary and Secondary Bacterial Symbionts of Psyllids. *Mol. Biol. Evol.* 29, 3781–3792. <https://doi.org/10.1093/molbev/mss180>
- Smith, T.E., Li, Y., Perreau, J., Moran, N.A., 2022. Elucidation of host and symbiont contributions to peptidoglycan metabolism based on comparative genomics of eight aphid subfamilies and their *Buchnera*. *PLOS Genet.* 18, e1010195. <https://doi.org/10.1371/journal.pgen.1010195>
- Snyder, A.K., Deberry, J.W., Runyen-Janecky, L., Rio, R.V.M., 2010. Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. *Proc. R. Soc. B Biol. Sci.* 277, 2389–2397. <https://doi.org/10.1098/rspb.2010.0364>
- Snyder, A.K., McLain, C., Rio, R.V.M., 2012. The Tsetse Fly Obligate Mutualist *Wigglesworthia morsitans* Alters Gene Expression and Population Density via Exogenous Nutrient Provisioning. *Appl. Environ. Microbiol.* 78, 7792–7797. <https://doi.org/10.1128/AEM.02052-12>
- Sochard, C., Dupont, C., Simon, J.-C., Outreman, Y., 2021. Secondary Symbionts Affect Foraging Capacities of Plant-Specialized Genotypes of the Pea Aphid. *Microb. Ecol.* 82, 1009–1019. <https://doi.org/10.1007/s00248-021-01726-7>
- Stoll, S., Feldhaar, H., Fraunholz, M.J., Gross, R., 2010. Bacteriocyte dynamics during development of a holometabolous insect, the carpenter ant *Camponotus floridanus*. *BMC Microbiol.* 10. <https://doi.org/10.1186/1471-2180-10-308>
- Sudakaran, S., Kost, C., Kaltenpoth, M., 2017. Symbiont Acquisition and Replacement as a Source of Ecological Innovation. *Trends Microbiol.* 25, 375–390. <https://doi.org/10.1016/j.tim.2017.02.014>
- Sudakaran, S., Retz, F., Kikuchi, Y., Kost, C., Kaltenpoth, M., 2015. Evolutionary transition in symbiotic syndromes enabled diversification of phytophagous insects on an imbalanced diet. *ISME J.* 9, 2587–2604. <https://doi.org/10.1038/ismej.2015.75>
- Szabó, G., Schulz, F., Manzano-Marín, A., Toenshoff, E.R., Horn, M., 2020. Evolutionary recent dual obligatory symbiosis among adelgids indicates a transition between fungus and insect associated lifestyles (preprint). *Microbiology*. <https://doi.org/10.1101/2020.10.16.342642>

- Szabó, G., Schulz, F., Toenshoff, E.R., Volland, J.-M., Finkel, O.M., Belkin, S., Horn, M., 2017. Convergent patterns in the evolution of mealybug symbioses involving different intrabacterial symbionts. *ISME J.* 11, 715–726. <https://doi.org/10.1038/ismej.2016.148>
- Takiya, D.M., Tran, P.L., Dietrich, C.H., Moran, N.A., 2006. Co-cladogenesis spanning three phyla: leafhoppers (Insecta: Hemiptera: Cicadellidae) and their dual bacterial symbionts: codiversification of sharpshooter endosymbionts. *Mol. Ecol.* 15, 4175–4191. <https://doi.org/10.1111/j.1365-294X.2006.03071.x>
- Talavera, G., Castresana, J., 2007. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Syst. Biol.* 56, 564–577. <https://doi.org/10.1080/10635150701472164>
- Tamas, I., Klasson, L., Canbäck, B., Näslund, A.K., Eriksson, A.-S., Wernegreen, J.J., Sandström, J.P., Moran, N.A., Andersson, S.G.E., 2002. 50 Million Years of Genomic Stasis in Endosymbiotic Bacteria. *Science* 296, 2376–2379. <https://doi.org/10.1126/science.1071278>
- Tamas, I., Klasson, L.M., Sandström, J.P., Andersson, S.G.E., 2001. Mutualists and parasites: how to paint yourself into a (metabolic) corner. *FEBS Lett.* 498, 135–139. [https://doi.org/10.1016/S0014-5793\(01\)02459-0](https://doi.org/10.1016/S0014-5793(01)02459-0)
- Tamas, I., Wernegreen, J.J., Nystedt, B., Kauppinen, S.N., Darby, A.C., Gomez-Valero, L., Lundin, D., Poole, A.M., Andersson, S.G.E., 2008. Endosymbiont gene functions impaired and rescued by polymerase infidelity at poly(A) tracts. *Proc. Natl. Acad. Sci.* 105, 14934–14939. <https://doi.org/10.1073/pnas.0806554105>
- Taylor, G.P., Coghlin, P.C., Floate, K.D., Perlman, S.J., 2011. The host range of the male-killing symbiont *Arsenophonus nasoniae* in filth fly parasitoids. *J. Invertebr. Pathol.* 106, 371–379. <https://doi.org/10.1016/j.jip.2010.12.004>
- Toenshoff, Elena R., Gruber, D., Horn, M., 2012. Co-evolution and symbiont replacement shaped the symbiosis between adelgids (Hemiptera: Adelgidae) and their bacterial symbionts: Bacteriocyte-associated symbionts of adelgids. *Environ. Microbiol.* 14, 1284–1295. <https://doi.org/10.1111/j.1462-2920.2012.02712.x>
- Toenshoff, Elena R., Penz, T., Narzt, T., Collingro, A., Schmitz-Esser, S., Pfeiffer, S., Klepal, W., Wagner, M., Weinmaier, T., Rattei, T., Horn, M., 2012. Bacteriocyte-associated gammaproteobacterial symbionts of the *Adelges nordmannianae/piceae* complex (Hemiptera: Adelgidae). *ISME J.* 6, 384–396. <https://doi.org/10.1038/ismej.2011.102>
- Toft, C., Andersson, S.G.E., 2010. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat. Rev. Genet.* 11, 465–475. <https://doi.org/10.1038/nrg2798>
- Toh, H., Weiss, B.L., Perkin, S.A.H., Yamashita, A., Oshima, K., Hattori, M., Aksoy, S., 2006. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Res.* 16, 149–156. <https://doi.org/10.1101/gr.4106106>
- Urban, J.M., Cryan, J.R., 2012. Two ancient bacterial endosymbionts have coevolved with the planthoppers (Insecta: Hemiptera: Fulgoroidea). *BMC Evol. Biol.* 12, 87. <https://doi.org/10.1186/1471-2148-12-87>
- Vallenet, D., Calteau, A., Dubois, M., Amours, P., Bazin, A., Beuvin, M., Burlot, L., Bussell, X., Fouteau, S., Gautreau, G., Lajus, A., Langlois, J., Planel, R., Roche, D., Rollin, J., Rouy, Z., Sabatet, V., Médigue, C., 2019. MicroScope: an integrated platform for the annotation and

- exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Res.* gkz926. <https://doi.org/10.1093/nar/gkz926>
- van der Wilk, F., Dullemans, A.M., Verbeek, M., van den Heuvel, J.F.J.M., 1999. Isolation and Characterization of APSE-1, a Bacteriophage Infecting the Secondary Endosymbiont of *Acyrtosiphon pisum*. *Virology* 262, 104–113. <https://doi.org/10.1006/viro.1999.9902>
- van Ham, R.C.H.J., Kamerbeek, J., Palacios, C., Rausell, C., Abascal, F., Bastolla, U., Fernández, J.M., Jiménez, L., Postigo, M., Silva, F.J., Tamames, J., Viguera, E., Latorre, A., Valencia, A., Morán, F., Moya, A., 2003. Reductive genome evolution in *Buchnera aphidicola*. *Proc. Natl. Acad. Sci.* 100, 581–586. <https://doi.org/10.1073/pnas.0235981100>
- Van rhijn, P., Vanderleyden, J., 1995. *The Rhizobium-Plant Symbiosis*.
- Vasquez, Y.M., Bennett, G.M., 2022. A complex interplay of evolutionary forces continues to shape ancient co-occurring symbiont genomes. *iScience* 104786. <https://doi.org/10.1016/j.isci.2022.104786>
- Vigneron, A., Masson, F., Vallier, A., Balmand, S., Rey, M., Vincent-Monégat, C., Aksoy, E., Aubailly-Giraud, E., Zaidman-Rémy, A., Heddi, A., 2014. Insects Recycle Endosymbionts when the Benefit Is Over. *Curr. Biol.* 24, 2267–2273. <https://doi.org/10.1016/j.cub.2014.07.065>
- Viñuelas, J., Febvay, G., Duport, G., Colella, S., Fayard, J.-M., Charles, H., Rahbé, Y., Calevro, F., 2011. Multimodal dynamic response of the *Buchnera aphidicola* pLeu plasmid to variations in leucine demand of its host, the pea aphid *Acyrtosiphon pisum*: pLeu plasmid regulation in aphid/*Buchnera* symbiosis. *Mol. Microbiol.* 81, 1271–1285. <https://doi.org/10.1111/j.1365-2958.2011.07760.x>
- Von Dohlen, C.D., Moran, N.A., 2000. Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation. *Biol. J. Linn. Soc.* 71, 689–717. <https://doi.org/10.1006/bjil.2000.0470>
- von Dohlen, C.D., Spaulding, U., Patch, K.B., Weglarz, K.M., Footitt, R.G., Havill, N.P., Burke, G.R., 2017. Dynamic Acquisition and Loss of Dual-Obligate Symbionts in the Plant-Sap-Feeding Adelgidae (Hemiptera: Sternorrhyncha: Aphidoidea). *Front. Microbiol.* 8, 1037. <https://doi.org/10.3389/fmicb.2017.01037>
- Wang, Y., Huang, X.-L., Qiao, G.-X., 2013. Comparative Analysis of Mitochondrial Genomes of Five Aphid Species (Hemiptera: Aphididae) and Phylogenetic Implications. *PLoS ONE* 8, e77511. <https://doi.org/10.1371/journal.pone.0077511>
- Wang, Y.-B., Ren, F.-R., Yao, Y.-L., Sun, X., Walling, L.L., Li, N.-N., Bai, B., Bao, X.-Y., Xu, X.-R., Luan, J.-B., 2020. Intracellular symbionts drive sex ratio in the whitefly by facilitating fertilization and provisioning of B vitamins. *ISME J.* 14, 2923–2935. <https://doi.org/10.1038/s41396-020-0717-0>
- Weber, M.G., Agrawal, A.A., 2014. Defense mutualisms enhance plant diversification. *Proc. Natl. Acad. Sci.* 111, 16442–16447. <https://doi.org/10.1073/pnas.1413253111>
- Weglarz, K.M., Havill, N.P., Burke, G.R., von Dohlen, C.D., 2018. Partnering With a Pest: Genomes of Hemlock Woolly Adelgid Symbionts Reveal Atypical Nutritional Provisioning Patterns in Dual-Obligate Bacteria. *Genome Biol. Evol.* 10, 1607–1621. <https://doi.org/10.1093/gbe/evy114>
- Weiss, B.L., Maltz, M., Aksoy, S., 2012. Obligate Symbionts Activate Immune System Development in the Tsetse Fly. *J. Immunol.* 188, 3395–3403. <https://doi.org/10.4049/jimmunol.1103691>

- Wernegreen, J.J., 2015. Endosymbiont evolution: predictions from theory and surprises from genomes: Endosymbiont genome evolution. *Ann. N. Y. Acad. Sci.* 1360, 16–35. <https://doi.org/10.1111/nyas.12740>
- Wernegreen, J.J., 2012. Endosymbiosis. *Curr. Biol.* 22, R555–R561. <https://doi.org/10.1016/j.cub.2012.06.010>
- Wernegreen, J.J., 2002. Genome evolution in bacterial endosymbionts of insects. *Nat. Rev. Genet.* 3, 850–861. <https://doi.org/10.1038/nrg931>
- Werren, J.H., 1997. BIOLOGY OF *WOLBACHIA*. *Annu. Rev. Entomol.* 42, 587–609. <https://doi.org/10.1146/annurev.ento.42.1.587>
- Werren, J.H., Baldo, L., Clark, M.E., 2008. Wolbachia: master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* 6, 741–751. <https://doi.org/10.1038/nrmicro1969>
- Werren, J.H., O'Neill, S.L., 1997. The evolution of heritable symbionts 11.
- Wertheim, J.O., Murrell, B., Smith, M.D., Kosakovsky Pond, S.L., Scheffler, K., 2015. RELAX: Detecting Relaxed Selection in a Phylogenetic Framework. *Mol. Biol. Evol.* 32, 820–832. <https://doi.org/10.1093/molbev/msu400>
- Whittle, M., Barreaux, A.M.G., Bonsall, M.B., Ponton, F., 2021. Insect-host control of obligate, intracellular symbiont density 21.
- Wilkinson, T.L., Douglas, A.E., 1998. Host cell allometry and regulation of the symbiosis between pea aphids, *Acyrtosiphon pisum*, and bacteria, *Buchnera*. *J. Insect Physiol.* 44, 629–635. [https://doi.org/10.1016/S0022-1910\(98\)00030-4](https://doi.org/10.1016/S0022-1910(98)00030-4)
- Wilkinson, T.L., Fukatsu, T., Ishikawa, H., 2003. Transmission of symbiotic bacteria *Buchnera* to parthenogenetic embryos in the aphid *Acyrtosiphon pisum* (Hemiptera: Aphidoidea). *Arthropod Struct. Dev.* 32, 241–245. [https://doi.org/10.1016/S1467-8039\(03\)00036-7](https://doi.org/10.1016/S1467-8039(03)00036-7)
- Wilkinson, T.L., Ishikawa, H., 2000. Injection of essential amino acids substitutes for bacterial supply in aposymbiotic pea aphids (*Acyrtosiphon pisum*). *Entomol. Exp. Appl.* 94, 85–91. <https://doi.org/10.1046/j.1570-7458.2000.00607.x>
- Wilkinson, T.L., Koga, R., Fukatsu, T., 2007. Role of Host Nutrition in Symbiont Regulation: Impact of Dietary Nitrogen on Proliferation of Obligate and Facultative Bacterial Endosymbionts of the Pea Aphid *Acyrtosiphon pisum*. *Appl. Environ. Microbiol.* 73, 1362–1366. <https://doi.org/10.1128/AEM.01211-06>
- Wolschin, F., Hölldobler, B., Gross, R., Zientz, E., 2004. Replication of the Endosymbiotic Bacterium *Blochmannia floridanus* Is Correlated with the Developmental and Reproductive Stages of Its Ant Host. *Appl. Environ. Microbiol.* 70, 4096–4102. <https://doi.org/10.1128/AEM.70.7.4096-4102.2004>
- Woolfit, M., 2003. Increased Rates of Sequence Evolution in Endosymbiotic Bacteria and Fungi with Small Effective Population Sizes. *Mol. Biol. Evol.* 20, 1545–1555. <https://doi.org/10.1093/molbev/msg167>
- Wu, D., Daugherty, S.C., Van Aken, S.E., Pai, G.H., Watkins, K.L., Khouri, H., Tallon, L.J., Zaborsky, J.M., Dunbar, H.E., Tran, P.L., Moran, N.A., Eisen, J.A., 2006. Metabolic Complementarity and Genomics of the Dual Bacterial Symbiosis of Sharpshooters. *PLoS Biol.* 4, e188. <https://doi.org/10.1371/journal.pbio.0040188>

- Yamamura, 1993. Vertical transmission and evolution of mutualism from parasitism. *Theoretical Population Biology*, 44(1), 95-109.
- Yang, Z., 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* 24, 1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yang, Z., Bielawski, J.P., 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* 15, 496–503. [https://doi.org/10.1016/S0169-5347\(00\)01994-7](https://doi.org/10.1016/S0169-5347(00)01994-7)
- Yorimoto, S., Hattori, M., Kondo, M., Shigenobu, S., 2022. Complex host/symbiont integration of a multi-partner symbiotic system in the eusocial aphid *Ceratovacuna japonica*. *iScience* 25, 105478. <https://doi.org/10.1016/j.isci.2022.105478>
- Zchori-Fein, E., Perlman, S.J., 2004. Distribution of the bacterial symbiont *Cardinium* in arthropods: DISTRIBUTION OF CARDINIUM IN ARTHROPODS. *Mol. Ecol.* 13, 2009–2016. <https://doi.org/10.1111/j.1365-294X.2004.02203.x>
- Zientz, E., Dandekar, T., Gross, R., 2004. Metabolic Interdependence of Obligate Intracellular Bacteria and Their Insect Hosts. *Microbiol. Mol. Biol. Rev.* 68, 745–770. <https://doi.org/10.1128/MMBR.68.4.745-770.2004>
- Zindel, R., Gottlieb, Y., Aebi, A., 2011. Arthropod symbioses: a neglected parameter in pest- and disease-control programmes: Endosymbionts and biological control. *J. Appl. Ecol.* 48, 864–872. <https://doi.org/10.1111/j.1365-2664.2011.01984.x>
- Zug, R., Hammerstein, P., 2015. Bad guys turned nice? A critical assessment of *Wolbachia* mutualisms in arthropod hosts. *Biol. Rev.* 90, 89–111. <https://doi.org/10.1111/brv.12098>
- Zug, R., Koehncke, A., Hammerstein, P., 2012. Epidemiology in evolutionary time: the case of *Wolbachia* horizontal transmission between arthropod host species. *J. Evol. Biol.* 25, 2149–2160. <https://doi.org/10.1111/j.1420-9101.2012.02601.x>
- Zytynska, S.E., Weisser, W.W., 2016. The natural occurrence of secondary bacterial symbionts in aphids: Aphid secondary symbionts. *Ecol. Entomol.* 41, 13–26. <https://doi.org/10.1111/een.12281>

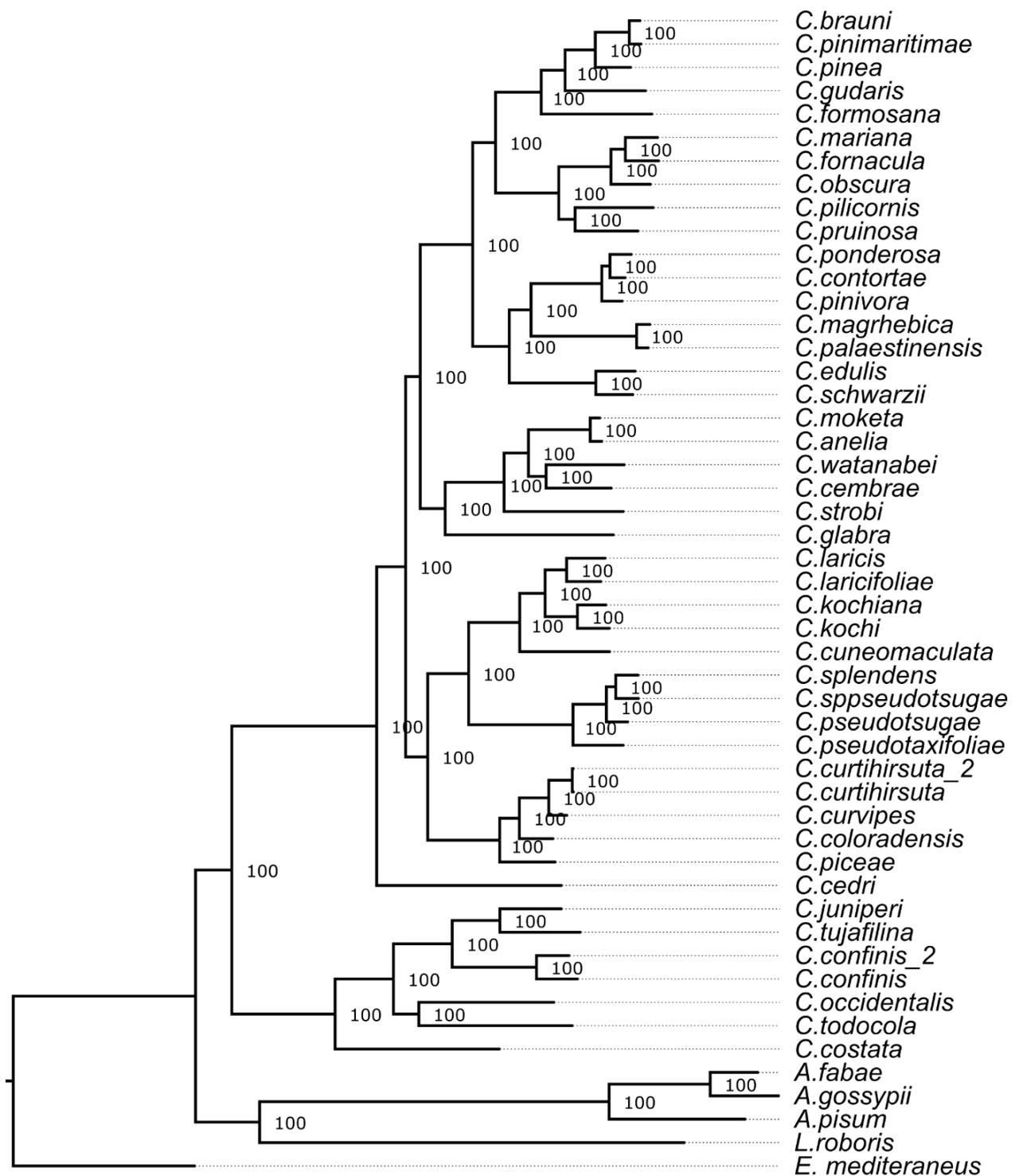
Annexes 1 : Matériel supplémentaire du chapitre 1

Table S1: Résumé des données génomiques utilisées. Mb= Mégabases.

Classification des pucerons hôtes			Buchnera associée			Serratia associée		
Sous-famille	Tribu	Nom d'espèce	Taille du génome (Mb)	GC% génomique	n° GenBank	Taille du génome (Mb)	GC% génomique	n° GenBank
Lachninae	Eulachnini	<i>Eulachnus mediterraneus</i>	0.420251	21.7				
Yersiniaceae		<i>Serratia marcescens</i> souche Db11*				5.177725	59.5	NZ_HG326223.1
Aphidinae	Aphidini	<i>Aphis gossypii</i> **	0.628098	25.4		3.191354	51.9	
Aphidinae	Aphidini	<i>Aphis fabae</i> **	0.634931	24.2		3.581747	52.1	CCES00000000.1
Aphidinae	Macrosiphini	<i>Acyrtosiphon pisum</i> **	0.655725	26.4	BA000003.2	2.78922	52	AENX00000000.1
Lachninae	Lachnini	<i>Lachnus roboris</i>	0.42432	20.7		0.712306	22.5	
Lachninae	Eulachnini	<i>Tuberolachnus salignus</i>	0.421426	21.6	LN890285.1	0.650317	20.9	LN890288.1
Lachninae	Eulachnini	<i>Cinara anelia</i>	0.447517	24.2		1.548732	42.2	
Lachninae	Eulachnini	<i>Cinara brauni</i>	0.449249	25		1.129658	30.9	
Lachninae	Eulachnini	<i>Cinara cedri</i>	0.426561	20.4	CP000263.1	1.785389	29.2	CP002295.1
Lachninae	Eulachnini	<i>Cinara cembrae mongolica</i> **	0.458493	25.8		2.907517	51.4	
Lachninae	Eulachnini	<i>Cinara contortae</i>	0.445934	25.9		1.130103	35	
Lachninae	Eulachnini	<i>Cinara curtihirsuta</i> **	0.433729	21.2		2.801016	51.7	
Lachninae	Eulachnini	<i>Cinara edulis</i>	0.441743	25.2		1.200872	34.6	
Lachninae	Eulachnini	<i>Cinara formosana</i>	0.445712	24.3		1.126283	28.4	
Lachninae	Eulachnini	<i>Cinara fornacula</i>	0.448244	28.1	LT667500.1	1.147624	32.2	
Lachninae	Eulachnini	<i>Cinara gudarisi</i>	0.446616	24.7		1.101076	30.7	
Lachninae	Eulachnini	<i>Cinara maghrebica</i>	0.445319	26.5		1.109775	35.8	
Lachninae	Eulachnini	<i>Cinara mariana</i>	0.450741	27.7		1.183691	32.5	
Lachninae	Eulachnini	<i>Cinara moketa</i>	0.459159	24.2		1.551839	42.5	
Lachninae	Eulachnini	<i>Cinara occidentalis</i>	0.440589	23.2		1.079026	34.4	
Lachninae	Eulachnini	<i>Cinara palaestinensis</i>	0.445286	26		1.133647	35.5	
Lachninae	Eulachnini	<i>Cinara pinea</i>	0.448598	25.4		1.139137	30.2	
Lachninae	Eulachnini	<i>Cinara pinimaritimae</i>	0.448675	24.9		1.127874	31	
Lachninae	Eulachnini	<i>Cinara pruinosa</i>	0.455279	25.6		1.275538	33.5	
Lachninae	Eulachnini	<i>Cinara schwarzi</i>	0.454085	25.7		1.197744	35.4	
Lachninae	Eulachnini	<i>Cinara strobili</i>	0.440065	23.9		2.436813	40.3	
Lachninae	Eulachnini	<i>Cinara tujafilina</i>	0.444925	23	CP001817.1	2.500326	52.1	GCA_900002265.1
Lachninae	Eulachnini	<i>Cinara watanabei</i>	0.444552	24.6		2.004754	41.9	
Lachninae	Eulachnini	<i>Cinara confinis</i>	0.443747	24		1.251054	26.2	
Lachninae	Eulachnini	<i>Cinara obscura</i>	0.448034	29.5		1.227235	31.4	
Lachninae	Eulachnini	<i>Cinara pilicornis</i>	0.452523	25.9		1.272682	32.8	
Lachninae	Eulachnini	<i>Cinara ponderosa</i>	0.443838	25.6		1.109917	34.9	
Lachninae	Eulachnini	<i>Cinara todocola</i>	0.429364	21.9		0.970484	32.5	
Lachninae	Eulachnini	<i>Cinara cf costata</i>	0.434665	21.2				
Lachninae	Eulachnini	<i>Cinara coloradensis</i>	0.435339	20.9				
Lachninae	Eulachnini	<i>Cinara cuneomaculata</i>	0.431261	23.6				
Lachninae	Eulachnini	<i>Cinara curtihirsuta</i>	0.433448	21.2				
Lachninae	Eulachnini	<i>Cinara curvipes</i>	0.434179	21.2				
Lachninae	Eulachnini	<i>Cinara glabra</i>	0.445575	25.3				
Lachninae	Eulachnini	<i>Cinara juniperi</i>	0.446614	23.5				
Lachninae	Eulachnini	<i>Cinara kochi</i>	0.434877	23.2				
Lachninae	Eulachnini	<i>Cinara kochiana</i>	0.433965	23.3				
Lachninae	Eulachnini	<i>Cinara laricifoliae</i>	0.437074	22.3				
Lachninae	Eulachnini	<i>Cinara laricis</i>	0.435435	22.8				
Lachninae	Eulachnini	<i>Cinara piceae</i>	0.435435	21.9				
Lachninae	Eulachnini	<i>Cinara pseudotaxifoliae</i>	0.447111	24.1	LT635893.1			
Lachninae	Eulachnini	<i>Cinara pseudotsugae sensu Palmer</i>	0.446142	23.8				
Lachninae	Eulachnini	<i>Cinara splendens pseudotsugae</i>	0.445035	23.9				
Lachninae	Eulachnini	<i>Cinara splendens sensu Palmer</i>	0.445737	23.6				

*Nom de la bactérie libre, pas d'hôte

** Serratia facultative



0.2

Figure S1 : Phylogénie de *Buchnera aphidicola* reconstruite en Maximum de Vraisemblance (IQtree1.6.2). Les noms des espèces-hôtes (pucerons) sont indiqués aux feuilles. Les valeurs de soutien aux nœuds sont indiquées en pourcentages de bootstraps non-paramétriques.

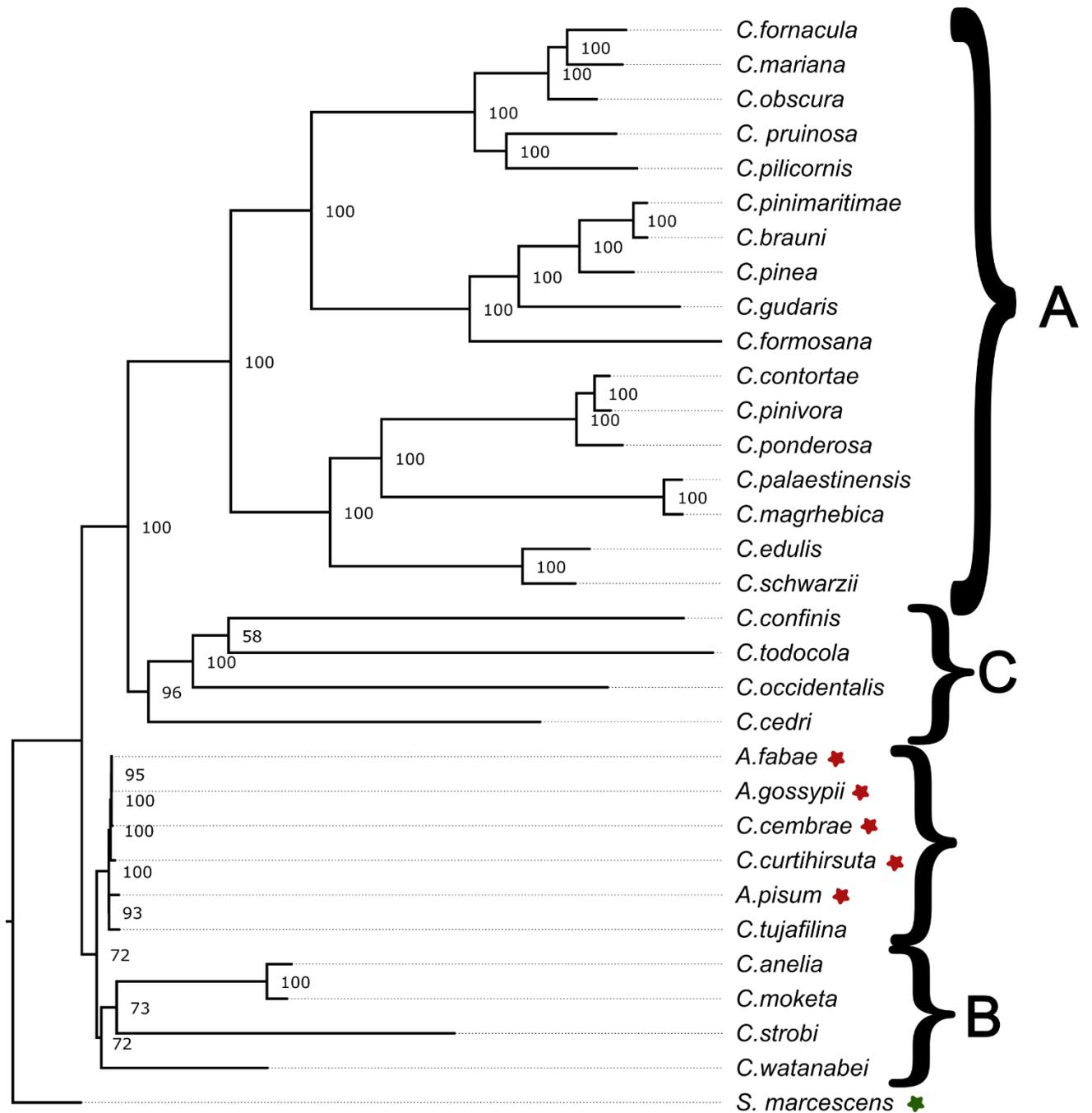


Figure S2 : Phylogénie des *Serratia symbiotica* reconstruite en Maximum de Vraisemblance (IQtree1.6.2). Les noms des espèces-hôtes (pucerons) sont indiqués aux feuilles. Les valeurs de soutien aux nœuds sont indiquées en pourcentages de bootstraps non-paramétriques. Les *Serratia* facultatives sont représentées avec une étoile rouge et la libre avec une verte.

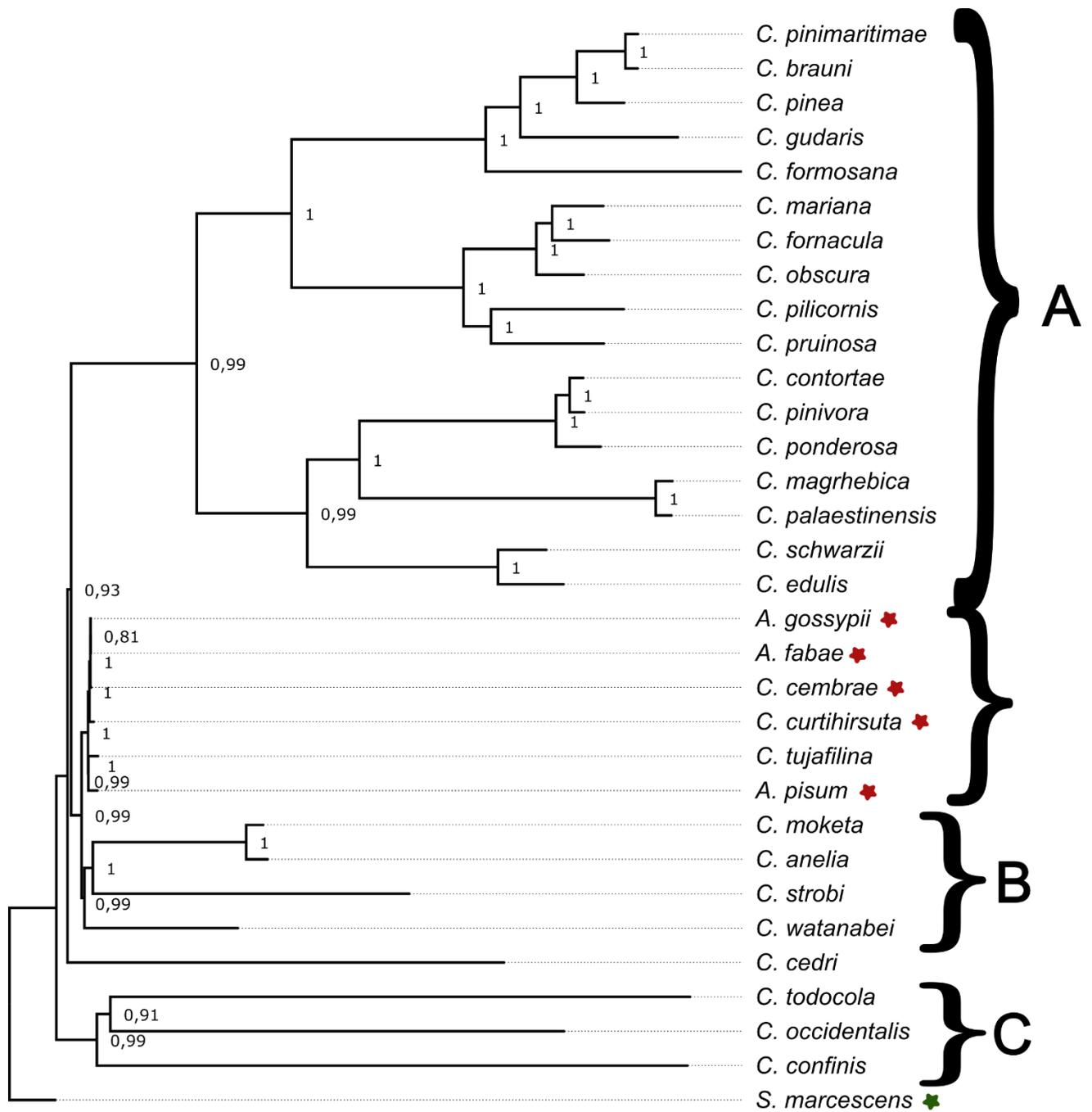


Figure S4 : Phylogénie de *Serratia symbiotica* reconstruite par inférence Bayésienne (Phylobayes). Les noms des espèces-hôtes (pucerons) sont indiqués aux feuilles. Les probabilités postérieures en soutien aux nœuds sont indiquées. Les *Serratia* facultatives sont représentées avec une étoile rouge et la libre avec une verte.

Annexes 2 : Matériel supplémentaire du chapitre 2

Table S1 : Collection data and accession numbers for *Cinara* species analysed in this study

Voucher ID	Species	Latitude	Longitude	Collection date	Host plant
ACDA2806	<i>Cinara pruinosa</i>	43.25833333	76.95638889	19_May_2010	<i>Picea sp.</i>
ACDA2827	<i>Cinara brauni</i>	43.23416667	76.95000000	26_May_2010	<i>Pinus nigra</i>
ACDA2836	<i>Cinara pinimaritimae</i>	43.75619888	-1.32882226	08_Jun_2010	<i>Pinus pinaster</i>
ACDA2891	<i>Cinara edulis</i>	39.66666667	-106.83194444	06_Jul_2010	<i>Pinus edulis</i>
ACDA2912	<i>Cinara fornacula</i>	37.51583333	-106.76916667	11_Jul_2010	<i>Picea sp.</i>
ACDA2925	<i>Cinara gudarisi</i>	43.77916667	3.54416667	13_Aug_2010	<i>Pinus nigra var. salzmannii</i>
ACDA2971	<i>Cinara schwarzii</i>	34.19337800	-116.71914900	31_May_2011	<i>Pinus ponderosa</i>
ACDA3032	<i>Cinara pinea</i>	48.08069800	-123.10309600	17_Jun_2011	<i>Pinus contorta</i>
ACDA3046	<i>Cinara mariana</i>	47.73658100	-122.64060600	17_Jun_2011	<i>Picea sp.</i>
ACDA3310	<i>Cinara formosana</i>	42.55998600	140.74579700	22_Aug_2012	<i>Pinus sp</i>
ACDA3375	<i>Cinara contortae</i>	46.438109	-115.098486	16_Aug_2013	<i>Pinus contorta</i>
ACDA3460	<i>Cinara palaestinensis</i>	34.890530	32.866229	16_Apr_2015	<i>Pinus brutia</i>
ACDA3544	<i>Cinara magrhebica</i>	35.257500	6.351100	08_Jun_2013	<i>Pinus halepensis</i>

Table S2. Genomic features of sequenced *Buchnera* strains. Min.cov = minimum coverage observed in one contigs, G+C (%) = G+C content, NCDS = number of CDS, Ncontigs = number of contigs, condng P =condng proportion. G+C content and coding density was calculated from the chromosome.

<i>Buchnera aphidicola</i>														
host	<i>C. edulis</i>	<i>C. schwarzii</i>	<i>C. contortae</i>	<i>C. palaestinensis</i>	<i>C. magrhebica</i>	<i>C. brauni</i>	<i>C. pinimarinimae</i>	<i>C. pinea</i>	<i>C. gudarisi</i>	<i>C. formosana</i>	<i>C. pruinosa</i>	<i>C. fornacula</i>	<i>C. mariana</i>	mean
contigs number	2	9	2	4	2	7	2	4	7	3	40	3	1	6.6
genome size (Kbp)	451.0	464.5	455.2	454.3	454.6	457.9	456.8	454.1	453.9	455.5	456.0	457.3	460.0	456.2
G+C content (%)	25.2	25.7	25.9	26	26.5	25	24.9	25.3	24.7	24.3	25.6	28.1	27.7	25.8
CDS (chr +pLeu +pTrp)	369 +5 +2	370 +5 +4	371 +5 +2	369 +5 +2	369 +5 +2	372 +5 +2	372 +5 +2	368 +5 +2	371 +5 +2	370 +5 +2	366 +5 +2	366 +5 +2	366 +5 +4	369 +5 +2
rRNA (23S,16S,5S)	1,1,1	1,1,1	1,1,1	1,1,1	1,1,1	1,1,1	1,1,1	1,1,1	1,1,1	1,1,1	1,1,1	1,1,1	1,1,1	1,1,1
tRNA	31	31	31	31	31	31	31	31	31	31	31	31	31	31
coding density (%)	85.5	84.2	85.3	85.0	85.0	84.7	85.0	84.7	85.3	84.9	84.2	84.0	84.1	84.8
min. coverage (chr:pLeu:pTrp)	53 :15 :80	62 :48 :744	136 :29 :201	62 :12 :74	31 :12 :150	56 :15 :87	106 :30 :300	47 :9 :382	63 :20 :530	125 :26 :346	100 :23 :268	139 :33 :476	NA :NA :NA	81 :22 :303

Table S3. Genomic features of sequenced *Serratia* strains. Min.cov = minimum coverage observed in one contigs, G+C (%) = G+C content, NCDS = number of CDS, Ncontigs = number of contigs, condng P = coding proportion. G+C content and coding density was calculated from the chromosome.

<i>Serratia symbiotica</i>														
host	<i>C. edulis</i>	<i>C. schwarzi</i>	<i>C. contortae</i>	<i>C. palaestinensis</i>	<i>C. magrhebica</i>	<i>C. brauni</i>	<i>C. pinimariitiae</i>	<i>C. pinea</i>	<i>C. gudarisi</i>	<i>C. formosana</i>	<i>C. pruinosa</i>	<i>C. fornacula</i>	<i>C. mariana</i>	mean
contigs number	7	5	4	5	3	2	2	2	2	2	7	2	NA	3.6
genome size (Kbp)	1200.9	1197.7	1130.1	1133.6	1109.8	1129.7	1127.9	1139.1	1101.1	1126.3	1275.5	1147.6	1183.7	1154.1
G+C content (%)	34.5	35.3	34.8	35.4	35.8	30.8	30.9	30.1	30.6	28.3	33.4	32.2	32.5	32.7
CDS (chr)	674	681	663	661	643	653	654	656	654	640	671	652	631	656.4
tRNA	34	34	34	35	34	34	34	34	34	34	35	35	35	34.3
coding density (%)	56.3	56.9	57.9	57.6	57.7	58.0	58.2	57.7	58.8	57.4	52.3	55.7	53.7	56.8
Min. coverage (chr)	136	77	26	60	38	27	24	50	34	65	95	363	NA	139.4

Table S4: Summary of genomic data from NCBI used as outgroup. Mb= Megabases. Nb= Number. The numbers of CDS correspond to the outputs of Prodigals. The numbers of contigs correspond to the results of the assemblies. ** facultative *Serratia*

sub-family	Trib	species	<i>Buchnera</i>					<i>Serratia</i>				
			genome size (Mb)	% GC	CDS Nb	contig Nb	accession Nb	genome size (Mb)	% GC	CDS Nb	contig Nb	accession Nb
Aphidinae	Macrosiphini	<i>Acyrtosiphon pisum</i> **	0.655725	26.4	594	1	BA000003.2	2.78922	52	5742	600	AENX00000000.1
Lachninae	Eulachnini	<i>Cinara cedri</i>	0.426561	20.4	380	8	CP000263.1	1.785389	29.3	757	4	CP002295.1
Lachninae	Eulachnini	<i>Cinara tujafilina</i>	0.444925	23	396	1	CP001817.1	2.500326	52.1	2924	22	GCA_900002265.1

Table S5 : Number and proportion of genes lost at least once during *Serratia*'s diversification in each Clusters of Orthologous Groups (COGs) functional categories (Tatusov et al., 2000).

function	COG	NB genes lost at least once	NB genes total	% genes lost at least once
RNA processing and modification	X	2	2	100.0
Chromatin structure and dynamics	Z	2	2	100.0
Energy production and conversion	B	1	1	100.0
Cell cycle control, cell division, chromosome partitioning	R	8	17	47.1
Amino acid transport and metabolism	S	5	12	41.7
Nucleotide transport and metabolism	E	13	35	37.1
Carbohydrate transport and metabolism	T	7	19	36.8
Coenzyme transport and metabolism	C	16	45	35.6
Lipid transport and metabolism	K	11	33	33.3
Translation, ribosomal structure and biogenesis	P	9	27	33.3
Transcription	D	8	27	29.6
Replication, recombination and repair	N	1	4	25.0
Cell wall/membrane/envelope biogenesis	L	11	46	23.9
Cell motility	F	8	35	22.9
Post-translational modification, protein turnover, chaperones	O	10	47	21.3
Inorganic ion transport and metabolism	M	14	72	19.4
Secondary metabolites biosynthesis, transport, catabolism	H	13	70	18.6
General function prediction only	G	4	27	14.8
Function unknown	V	1	7	14.3
Signal transduction mechanisms	U	1	14	7.1
Intracellular trafficking, secretion, and vesicular transport	I	2	33	6.1
Defense mechanisms	J	8	150	5.3
Extracellular structures	W	0	2	0.0
Mobilome: prophages, transposons	A	0	1	0.0
Cytoskeleton	Q	0	1	0.0

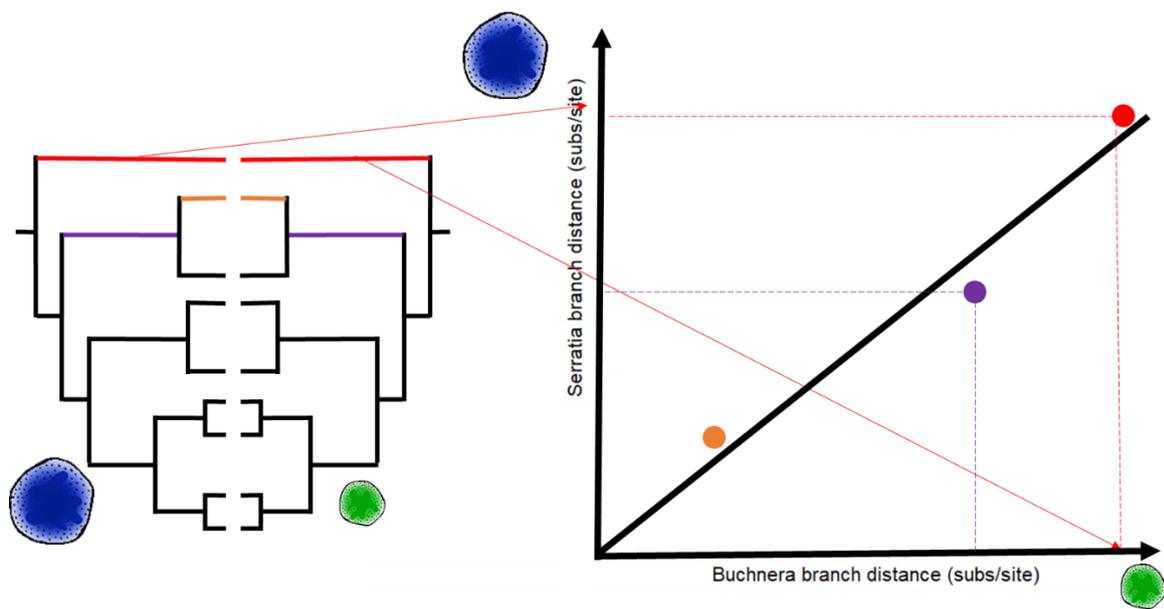


Figure S2 : Schematic representation of the branch correlation analysis. On the right part *Serratia* and *Buchnera* co-phylogeny. On the left part the correlation plot.

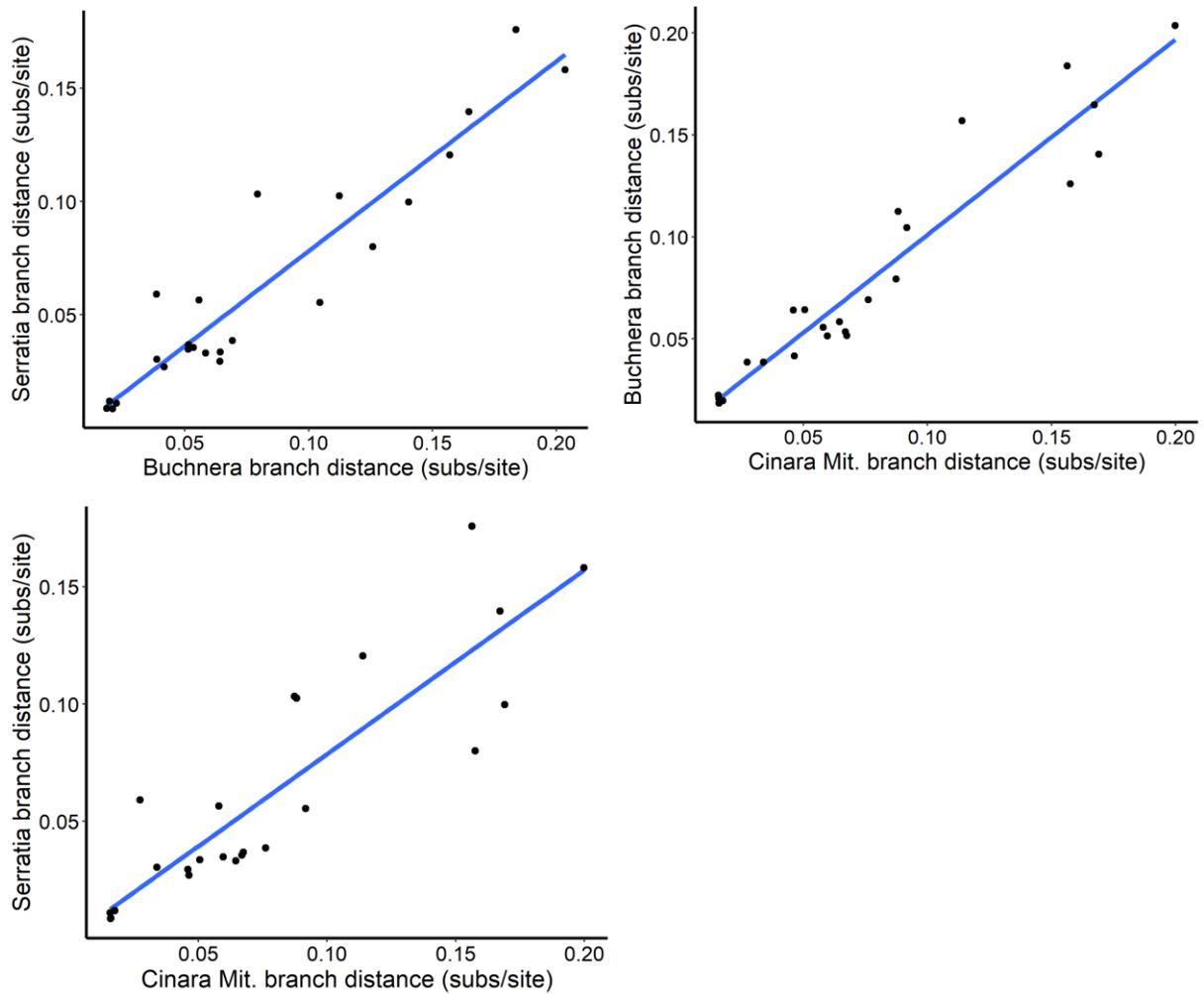


Figure S3 : Comparison of evolutionary rates of *Serratia* and *Buchnera* symbionts and their host *Cinara*. Top left correlation of branch distances in phylogenies of *Serratia* and *Buchnera*, inferred using MrBayes Bayesian analysis of CDS from each data set. Top right correlation of branch distance in phylogenies of *Cinara* mitochondrial genome and *Buchnera*, inferred using MrBayes Bayesian analysis of *Buchnera* CDS and mitochondrial genes data set. Bottom correlation of branch distance in phylogenies of *Cinara* mitochondrial genome and *Serratia* inferred using MrBayes Bayesian analysis of *Serratia* CDS and mitochondrial genes data set. (host-*Buchnera* : $R = 0.90$, $p = 7.1e-13$; host-*Serratia* : $R = 0.77$, $p = 1.7e-8$; *Serratia*-*Buchnera* : $R = 0.89$, $p = 8e-12$)

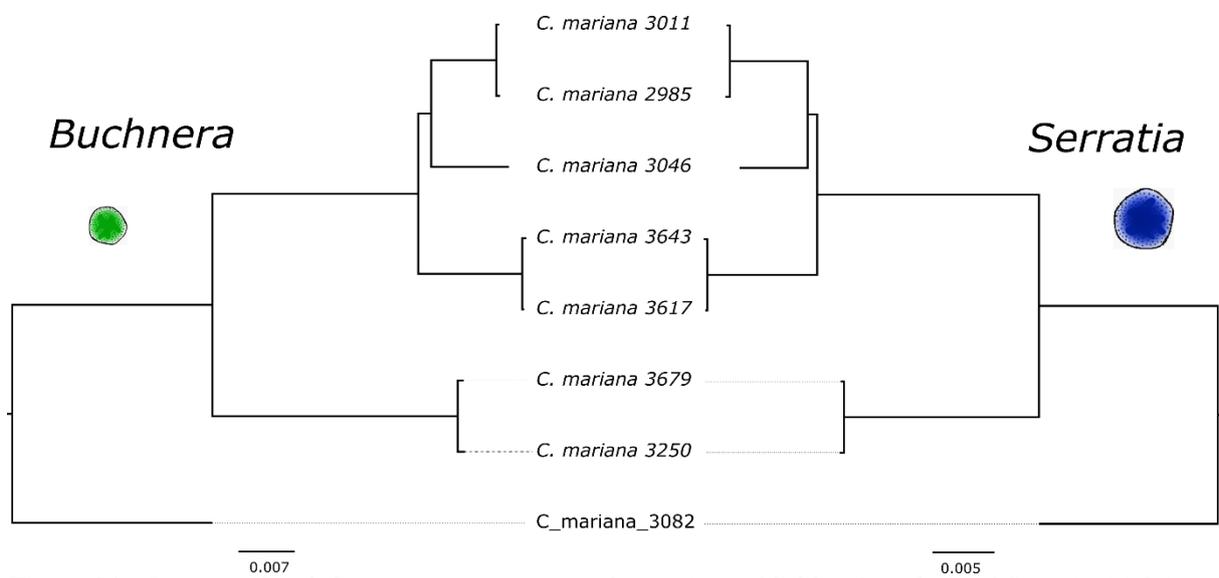


Figure S4 : Intraspecific phylogenetic reconstruction by maximum-likelihood analysis of *Serratia* (right) and *Buchnera* (left) of *Cinara mariana*.

Table S6 : Mutational patterns in genomes of *Buchnera* (up) and *Serratia* (Down) of 7 *Cinara mariana*. Nsyn. = number of non-synonymous SNPs, coding = coding sequences, Ncoding = non-coding sequences.

Buchnera										
	length (pb)	SNPs	Nsyn.	dN/dS	SNPs/kb	1pb indels	2bp indels	total indels	indels/kb	SNP/indels
coding	369243	36320	11995	0.33	98.36	30	12	89	0.24	408.09
Ncoding	78838	11374	NA	NA	144.27	938	112	1146	14.54	9.92
total	448081	47694	NA	NA	106.44	968	124	1235	2.76	38.62

Serratia										
	length (pb)	SNPs	Nsyn.	dN/dS	SNPs/kb	1pb indel	2bp indel	total indels	indels/kb	SNP/indels
coding	626723	44554	15051	0.34	71.09	41	15	88	0.14	506.3
Ncoding	556968	84760	NA	NA	152.18	14336	1865	24041	43.16	3.53
total	1183691	129314	NA	NA	109.25	14377	1880	24129	20.38	5.36

Table S7 : Mutational patterns in genomes of *Buchnera* (up) and *Serratia* (Down) of 7 *Cinara formacula*. Nsyn. = number of non-synonymous SNPs, coding = coding sequences, Ncoding = non-coding sequences.

Buchnera										
	length	SNPs	Nsyn.	dn/ds	SNPs/kb	1bp indels	2bp indels	total indels	indels/kb	SNP/indel
coding	370866	8587	2996	0.35	23.21	18	3	23	0.06	373.35
Ncoding	76807	3465	NA	NA	46.82	346	30	486	6.39	7.13
total	447673	12052	NA	NA	26.96	364	33	509	1.14	23.68

Serratia										
	length	SNPs	Nsyn.	dn/ds	SNPs/kb	1bp indels	2bp indels	total indels	indels/kb	SNP/indel
coding	636090	10368	3936	0.38	16.30	27	3	58	0.09	178.76
Ncoding	524728	20595	NA	NA	39.30	2377	286	3408	6.50	6.04
total	1160818	30963	NA	NA	26.69	2404	289	3466	2.99	8.93

Table S8 : Substitution rates reported in the literature for the primary aphid symbiont, *Buchnera aphidicola*. For each study, the dataset and methods used to estimate these average substitution rates are indicated. Subst=substitutions, My= millions years, nb.=number.

Study	substitution rate (subst/site/My)	data set	Calibrations	Methods
Ochman et al. 1999	0.0082	Genes : 16S rRNA on 2 strains of <i>B. aphidicola</i>	Approximate date of divergence of the two lineages	Estimation of the total number of substitutions on the 16S fragment
Tamas et al. 2002	0.0016 subst non syn. 0.009-subst. Syn	2 genomes of <i>B. aphidicola</i> (<i>Schizaphis graminum</i> and <i>Acyrtosiphon pisum</i>), estimation on 348 orthologous genes	Approximate date of divergence of the two lineages estimated at 50My	Estimation of the total number of substitutions, relative to the size of the genomes and to 50My
Gomez-Valero et al. 2007	0.0043-0.0067	Calibration on 2 introns and 1 pseudogene of <i>B. aphidicola</i> associated to 5 <i>Rhopalosiphum</i> sp. , <i>A. pisum</i> and <i>S.gramimum</i>	Divergence between Macrosiphini and Aphidini (between 50 and 70My ago)	Estimation using a strict molecular clock with BEAST
Moran et al. 2009	0.11 - 0.08	9 genomes of <i>B. aphidicola</i> associated to <i>A. pisum</i>	Date of introduction of the pea aphid in the USA (130 years)	Estimation of total number of substitutions / total tree length in y / genome size
Perez-Brocal et al. 2011	0.004- 0.007- subst non syn. 0.0015-0.0018 subst. syn	4 genomes of <i>B. aphidicola</i> (<i>A. pisum</i> , <i>S.gramimum</i> , <i>C. cedri</i> , <i>B. pistacia</i>), 348 to 350 orthologous genes	100My for aphids of different subfamilies; 50My for Macrosiphini/Aphidini	Estimation of nb. of substitutions between genome pairs/approximate divergence dates of genomes in My
Clark et al. 1999	0.00258 subst non syn. 0.0143-subst. syn	26 genes in 4 genomes of <i>B. aphidicola</i> , <i>E. coli</i> and <i>S. typhimurin</i>	50-70My for <i>Buchnera</i> and 100-150My for the other two	Estimated nb. of substitutions between genome pairs
Van Ham et al. 2002	0.00072 subst non syn. 0.0032 subst. syn	3 genomes of <i>B. aphidicola</i>	50-70My	Estimated nb. of substitutions between genome pairs

Annexes 3 : Matériel supplémentaire du chapitre 3

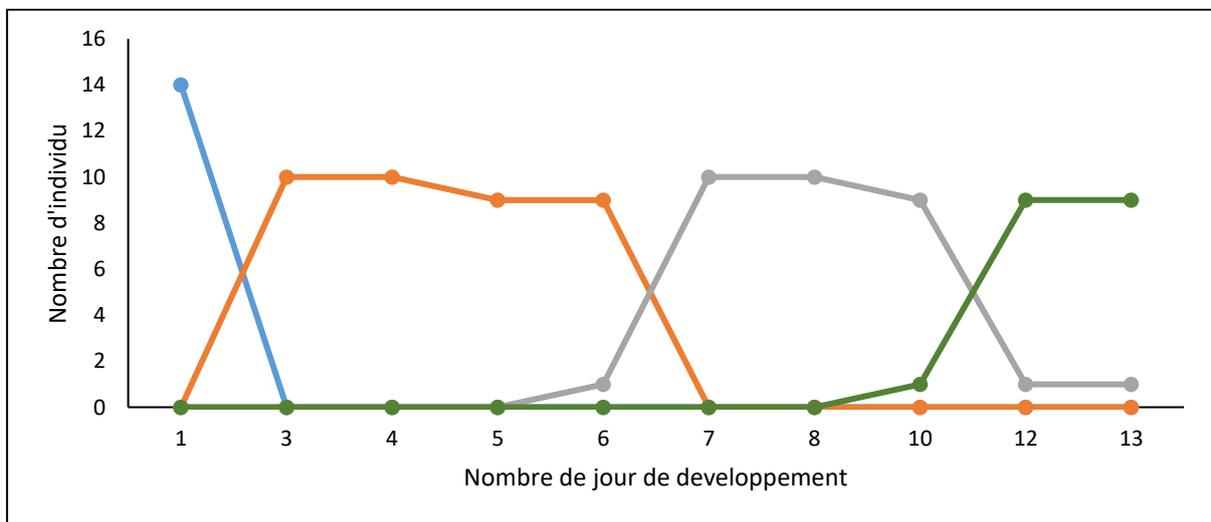


Figure S1 : Nombre d'individu de chaque stade observé chaque jour après la naissance pour l'espèce *C. tujafilina*. En bleu les stades 1, en orange les stades 2 en gris les stades 3 et en vert les adultes.

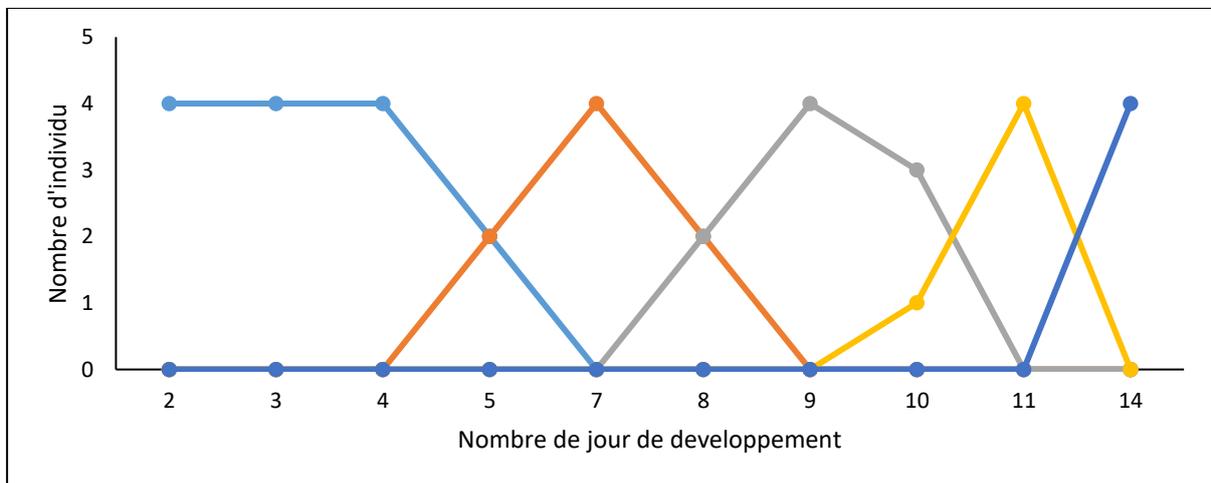


Figure S2 : Nombre d'individu de chaque stade observé chaque jour après la naissance pour l'espèce *C. palaestinis*. En bleu les stades 1, en orange les stades 2 en gris les stades 3, en jaunes les stades 4 et en vert les adultes.

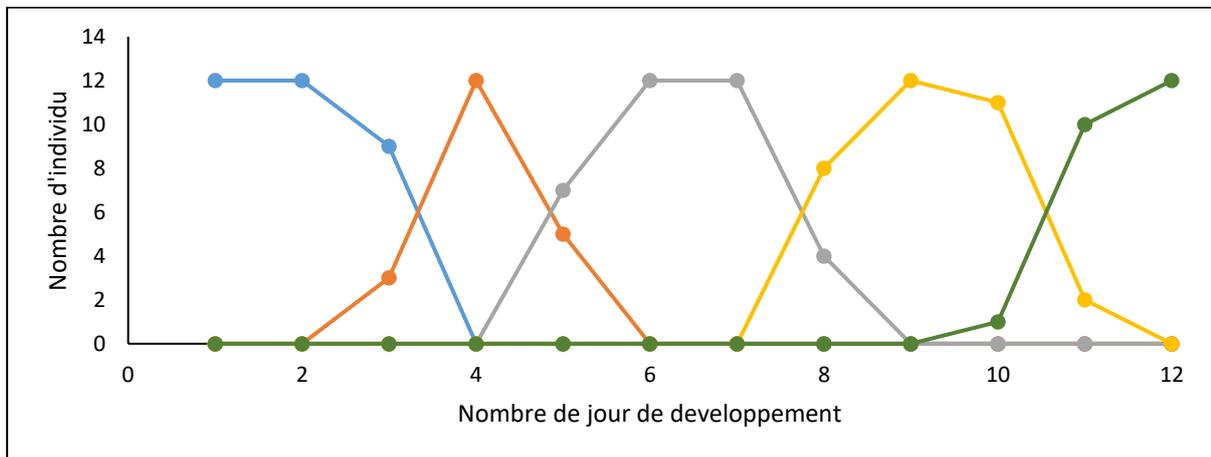


Figure S3 : Nombre d'individu de chaque stade observé chaque jour après la naissance pour l'espèce *C. cuneomaculata*. En bleu les stades 1, en orange les stades 2 en gris les stades 3, en jaunes les stades 4 et en vert les adultes.

Annexes 4 : Matériel supplémentaire du chapitre 4

Table S1. Collection data and accession numbers for hosts of novel APSE phages and those extracted from public databases. * was isolated from an unidentified symbiont.

voucher ID/strain	species	latitude	longitude	collection date	Host plant	cox1 accession	host/APSE accession	citation (doi)
2628	<i>Cinara cuneomaculata</i>	44.90333333	6.455555556	20-Aug-2008	<i>Larix decidua</i>	KF694061.1 (cob)	LR794149.1	This study
2671	<i>Perophyllus testudinaceus</i>	44.098404	7.506451	24-Jun-2009	<i>Acer</i> sp.	NA	LR794150.1	This study
2801	<i>Cinara confinis</i>	43.6085968	3.873177528	07-May-2010	<i>Abies cephalonica</i>	KF649386.1	LR794147.1	This study
2836	<i>Cinara pinimaritimae</i>	43.75619888	-1.328822255	08-Jun-2010	<i>Pinus pinaster</i>	KF694103.1 (cob)	LR794146.1	This study
3046	<i>Cinara</i> sp.	45.426021	-117.269102	23-Jun-2011	<i>Picea</i> sp.	KF649515.1	TBA	This study
3293	<i>Cinara watanabei</i>	37.766208	128.611036	14-Aug-2012	<i>Pinus</i> sp.	KP869229.1	LR794148.1	This study
3692	<i>Eriosoma grossulariae</i>	43.87782	3.703014	03-Jun-2016	<i>Ulmus</i> sp.	NA	LR794152.1	This study
3702	<i>Drepanosiphum platanoidis</i>	43.92628706	3.66216859	23-Jun-2016	<i>Acer</i> sp.	NA	LR794151.1	This study
5AT	<i>Acyrtosiphon pisum</i>	NA	NA	2000	NA	NA	CP001277.1	10.1073/pnas.0900194106
NY26	<i>Acyrtosiphon pisum</i>	NA	NA	NA	NA	NA	CP017605.1	10.1093/gbe/evy036
5D	<i>Acyrtosiphon pisum</i>	NA	NA	NA	NA	NA	CP021663.1	10.1093/gbe/evz252
ZA17	<i>Acyrtosiphon pisum</i>	NA	NA	NA	NA	NA	CP017613.1	10.1093/gbe/evy036
*APSE1	<i>Acyrtosiphon pisum</i>	NA	NA	NA	NA	NA	AF157835.1	10.1006/viro.1999.9902
MI47	<i>Acyrtosiphon pisum</i>	NA	NA	NA	NA	NA	CP022932.1	NA
AS3	<i>Acyrtosiphon pisum</i>	NA	NA	NA	NA	NA	CP017610.1	10.1093/gbe/evy036
MEAM1	<i>Bemisia tabaci</i>	NA	NA	NA	NA	PRJNA352527 (genome)	CP016303.1	10.1186/s12915-016-0321-y
MED-Q1	<i>Bemisia tabaci</i>	NA	NA	Jun-2009	cucumber	DQ473394.1	AJLH02000001.1	10.1186/s12864-015-1379-6

Table S2. Accession numbers for proteins used for BLASTX scaffold binning

species	strain	type	accessions
<i>Acyrtosiphon pisum</i>	LSR1	nuclear	ABLF00000000.2
<i>Acyrtosiphon pisum</i>	5A	mitochondrion	FJ411411.1
<i>Arsenophonus nasoniae</i>	DSM 15247	chromosome + plasmid	AUCC00000000.1
<i>Buchnera aphidicola</i>	APS	chromosome + plasmid	BA000003.2, AP001071.1, AP001070.1
<i>Buchnera aphidicola</i>	Bp	chromosome + plasmid	AE016826.1, AF492591.1
<i>Buchnera aphidicola</i>	BCc	chromosome + plasmid	CP000263.1, AY438025.1, EU660486.1
<i>Buchnera aphidicola</i>	Sg	chromosome + plasmid	AE013218.1, AF041836.1, Z21938.1
<i>Buchnera aphidicola</i>	BCt	chromosome + plasmid	CP001817.1, AY438024.1
<i>Buchnera aphidicola</i>	BTs	plasmid	Y11966.1
<i>Buchnera aphidicola</i>	BTc	plasmid	AJ012333.1
<i>Buchnera aphidicola</i>	BPS	plasmid	AJ404864.1
<i>Cordyceps brongniartii</i>	RCEF 3172	all	AZHA00000000.1
<i>Cordyceps militaris</i>	CM01	all	AEVU00000000.1
<i>Erwinia amylovora</i>	ATCC 49946	chromosome + plasmid	FN666575.1, FN666576.1, FN666577.1
<i>Erwinia pyrifoliae</i>	Ep1/96	chromosome + plasmid	FP236842.1, FP236827.1, FP236828.1, FP928999.1, FP236829.1
<i>Candidatus Erwinia</i> sp.	ErCipseudotaxifoliae	chromosome + plasmid	LT670851.1, LT670852.1
<i>Candidatus Sulcia muelleri</i>	PSPU	chromosome	AP013293.1
<i>Flavobacterium johnsoniae</i>	UW101	chromosome	CP000685.1
<i>Blattabacterium</i> sp.	Bge	chromosome + plasmid	CP001487.1, CP002849.1
<i>Candidatus Fukatsuia symbiotica</i>	FuCiconfinis	all	FQSP00000000.1
<i>Saccharibacter</i> sp.	AM169	all	CBLY00000000.1
<i>Gluconobacter oxydans</i>	621H	chromosome + plasmid	CP000009.1, CP000004.1, CP000005.1, CP000006.1, CP000007.1, CP000008.1
<i>Candidatus Hamiltonella defensa</i>	5AT	chromosome + plasmid	CP001277.1, CP001278.1
<i>Candidatus Schmidhempelia bombi</i>	Bimp	all	AWGA00000000.1
<i>Gilliamella apicola</i>	wkB1	chromosome	CP007445.1
<i>Frischella perrara</i>	PEB0191	chromosome	CP009056.1
<i>Candidatus Regiella insecticola</i>	LSR1	chromosome + plasmid	ACYF00000000.1, CM000957.1
<i>Rickettsia bellii</i>	OSU 85-389	chromosome	CP000849.1
<i>Rickettsia bellii</i>	RML369-C	chromosome	CP000087.1
<i>Rickettsia prowazekii</i>	Madrid E	chromosome	AJ235269.1
<i>Rickettsiella grylli</i>	NA	all	AAQJ00000000.2
<i>Candidatus Rickettsiella isopodorum</i>	NA	all	LUKY00000000.1
<i>Serratia symbiotica</i>	CWBI-2.3	all	CCES00000000.1
<i>Serratia symbiotica</i>	Tucson	all	AENX00000000.1
<i>Serratia symbiotica</i>	SCt-VLC	all	GCA_900002265.1
<i>Serratia symbiotica</i>	Cinara cedri	chromosome	CP002295.1
<i>Serratia symbiotica</i>	STs	chromosome	LN890288.1
<i>Serratia marcescens</i>	Db11	chromosome	HG326223.1
<i>Sodalis glossinidius</i>	morsitans	chromosome + plasmid	AP008232.1, AP008233.1, AP008234.1, AP008235.1
<i>Sodalis praecaptivus</i>	HS1	chromosome + plasmid	CP006569.1, CP006570.1
<i>Candidatus Sodalis pierantonius</i>	SOPE	chromosome	CP006568.1
<i>Spiroplasma citri</i>	R8-A2	chromosome + plasmid	CP013197.1, CP013199.1, CP013198.1
<i>Wolbachia</i> sp.	wPip	chromosome	AM999887.1
<i>Wolbachia</i> sp.	wCle	chromosome	AP013028.1
<i>Yersinia pestis</i>	CO92	chromosome + plasmid	AL590842.1, AL117189.1, AL117211.1, AL109969.1

Table S4. Phipack recombination test results and amino acid substitution model used for each gene in the partition dataset.

<i>orthogroup</i>	<i>Phi_normal</i>	<i>Phi_permutation_N1000</i>	<i>NSS_permutation_N1000</i>	<i>Max_Chi_squarepermutation_N1000</i>	<i>model</i>
APSEcp_002	NA	1.00e+00	5.36e-01	1.00e+00	VT+I
APSEcp_008	5.90e-01	6.61e-01	0.00e+00	0.00e+00	WAG+I+G4
APSEcp_011	NA	6.22e-01	6.10e-02	0.00e+00	DAYHOFF+I
APSEcp_012	4.58e-01	4.92e-01	0.00e+00	0.00e+00	WAG+I+G4
APSEcp_013	8.31e-01	8.67e-01	2.55e-01	7.55e-01	VT+G4
APSEcp_014	7.90e-02	1.71e-01	5.19e-01	6.00e-03	CpRev+I+G4
APSEcp_016	5.84e-02	7.10e-02	0.00e+00	1.13e-01	WAG+I+G4
APSEvp_002	NA	3.62E-01	3.83E-01	6.46E-01	CpRev+G4
APSEcp_021	3.70e-01	4.02e-01	3.00e-02	4.90e-02	WAG+I+G4
APSEcp_022	1.00e+00	1.00e+00	1.00e+00	4.69e-01	RtRev
APSEcp_023	6.42e-02	1.78e-01	5.10e-02	8.00e-02	WAG+I+G4
APSEcp_025	NA	2.99e-01	6.20e-01	8.83e-01	WAG+I
APSEcp_026	6.05e-02	7.50e-02	1.00e-03	6.00e-03	WAG+I+G4
APSEcp_027	NA	5.10e-01	5.00e-02	4.60e-02	WAG+G4
APSEcp_028	NA	1.00e+00	1.76e-01	1.05e-01	CpRev+I+G4
APSEcp_029	7.28e-01	1.00e+00	2.77e-01	9.28e-01	CpRev+I+G4
APSEcp_030	1.15e-01	3.29e-01	2.81e-01	4.33e-01	VT+I+G4
APSEcp_031	8.83e-02	1.34e-01	7.85e-01	9.23e-01	WAG+I+G4

Table S5. Codon usage (per thousand) calculated by the cousin web server for the two lysine codons for both APSE and *Hamiltonella* hosts.

AAA	AAG	APSE tRNA	strain	APSE type
55.5	11.4	<i>CUU/UUU</i>	<i>MEAMI</i>	NA
53.8	16.5	<i>CUU</i>	<i>MEAMI APSE</i>	7
57.3	12.5	<i>CUU/UUU(pseudogene)</i>	<i>5AT</i>	NA
53.2	16.2	<i>UUU</i>	<i>5AT APSE</i>	2
57.4	12.7	<i>CUU/UUU</i>	<i>NY26</i>	NA
53.4	16.1	<i>UUU</i>	<i>NY26 APSE</i>	2
57.2	12.6	<i>CUU/UUU</i>	<i>5D</i>	NA
53.8	15.9	<i>pseudogene</i>	<i>5D APSE</i>	2
56.1	12.3	<i>CUU/UUU</i>	<i>ZA17</i>	NA
53.7	15.8	<i>pseudogene</i>	<i>ZA17 APSE</i>	2
57.2	12.8	<i>CUU/UUU</i>	<i>MI47</i>	NA
52.8	17.1	<i>UUU</i>	<i>MI47 APSE</i>	1
56.3	12.5	<i>CUU/UUU</i>	<i>AS3</i>	NA
51.2	16.0	<i>UUU</i>	<i>AS3 APSE</i>	3

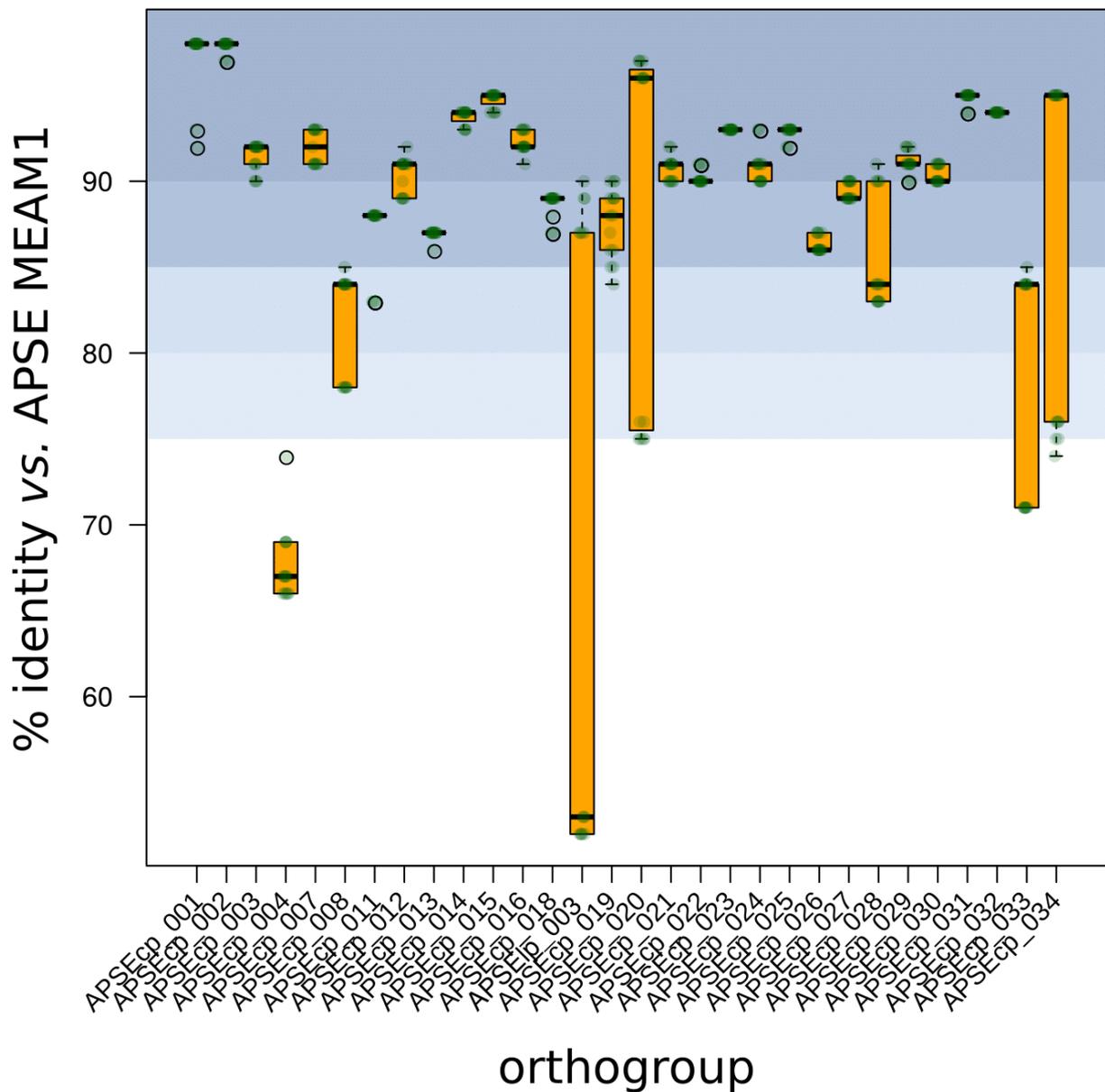


Fig. S1. Identity of conserved proteins across APSE vs. MEAM1. Example of identity across single-copy shared proteins across APSE phages vs. MEAM1 as calculated in ClustalX. A file showing identity tables across all 30 proteins can be found in file S3.

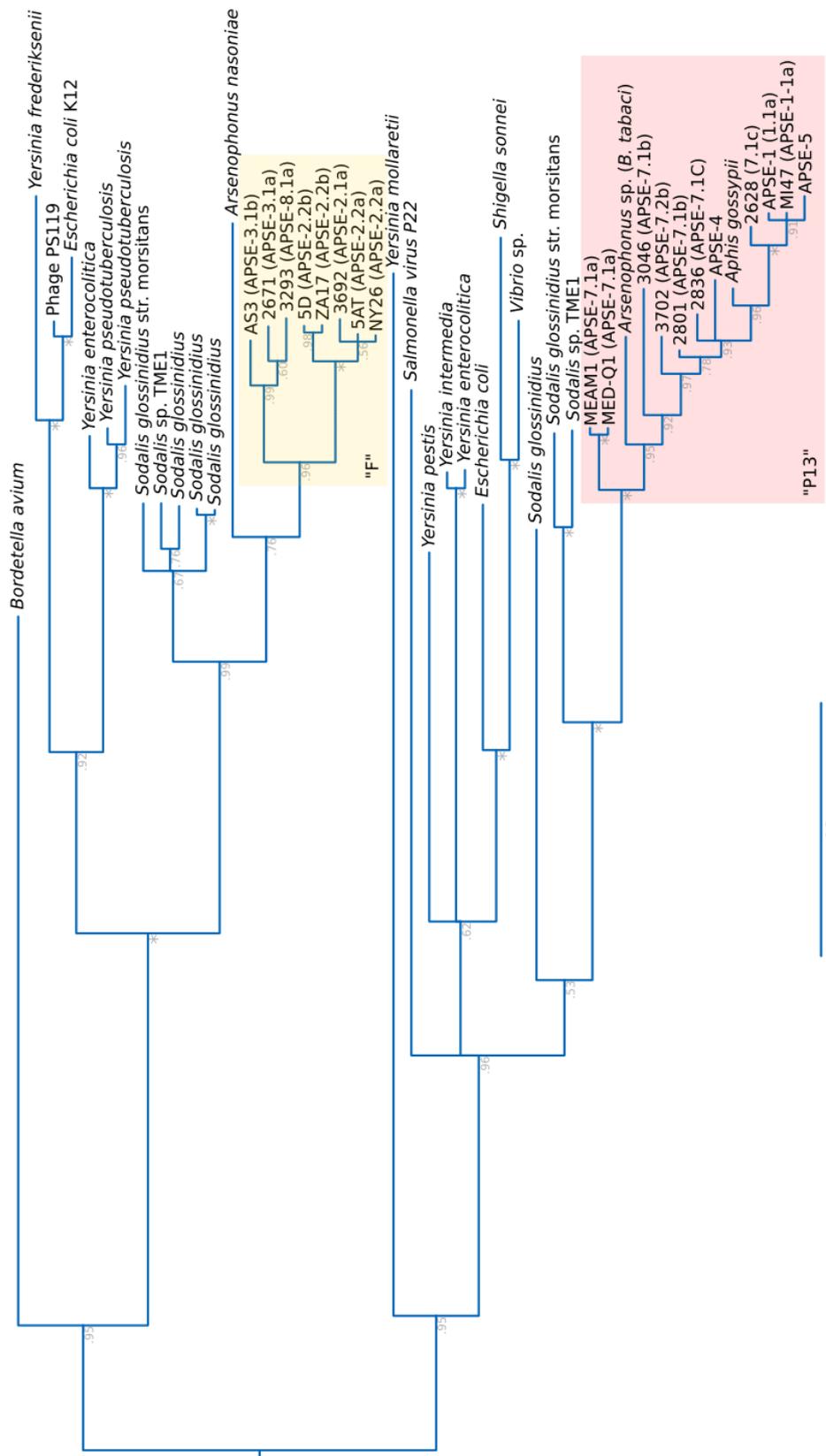


Fig. S2. Bayesian phylogeny of lysozyme genes from APSE phages. Amino acid Bayesian phylogeny reconstructed in MrBayes v3.2.7 (JTT+I+G4; 300,000 generations) highlighting the two lysozyme phylogroups: the so-called "F" (yellow) and "P13" (red). Phylogeny was midpoint rooted.

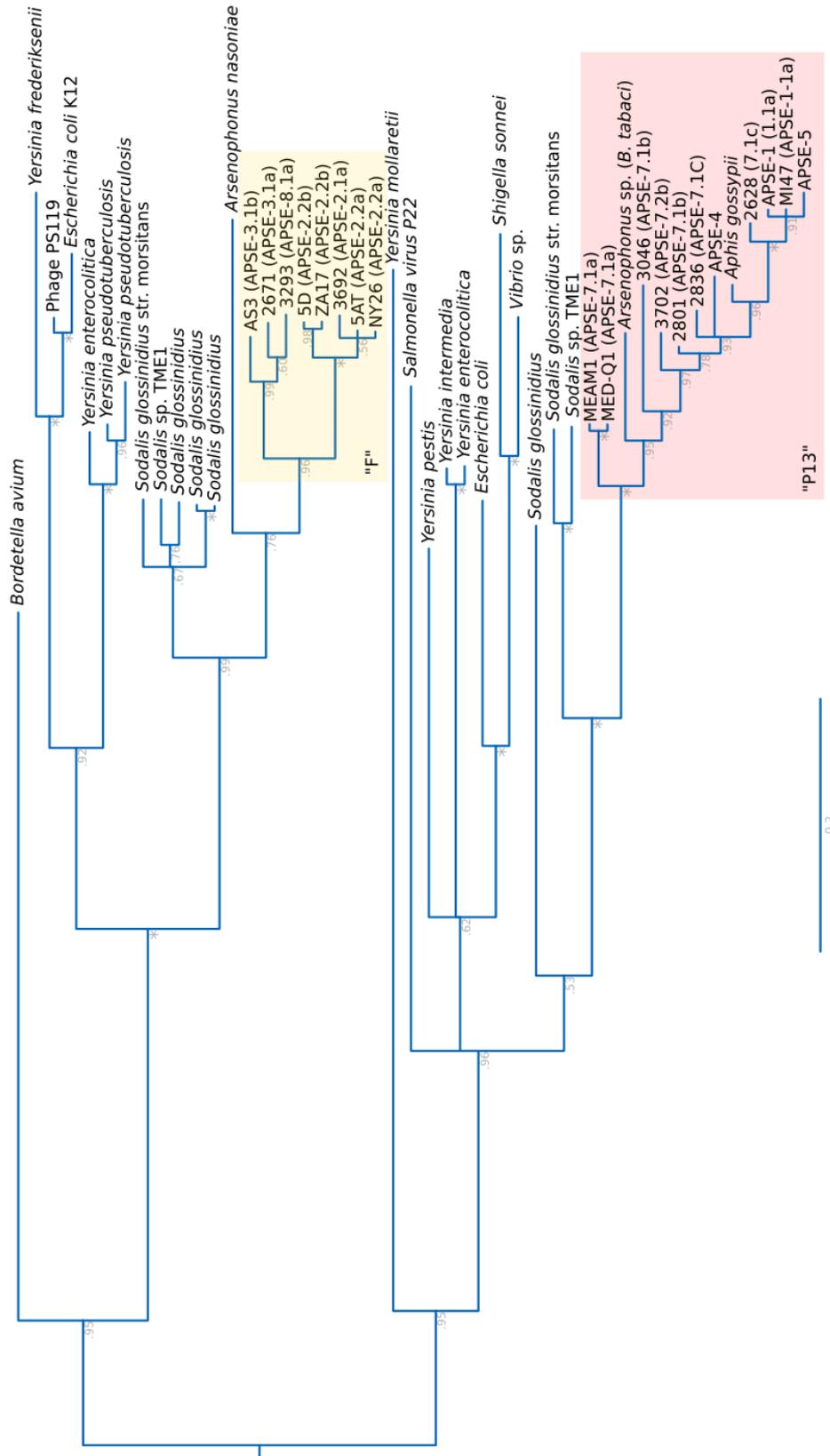


Fig. S2. Bayesian phylogeny of lyozyme genes from APSE phages. Amino acid Bayesian phylogeny reconstructed in MrBayes v3.2.7 (JTT+I+G4; 300,000 generations) highlighting the two lyozyme phylogroups: the so-called "F" (yellow) and "P13" (red). Phylogeny was midpoint rooted.

Évolution et dynamique du système di-symbiotique chez les pucerons du genre *Cinara*

Résumé : Des systèmes pluri-symbiotiques ont été mis en évidence chez plusieurs espèces de pucerons (Hemiptera : Aphididae) et plus particulièrement chez les espèces du genre *Cinara*. La caractérisation taxonomique des symbiotes présents chez une soixantaine d'espèces de *Cinara* a permis de déterminer que *Serratia symbiotica* est la bactérie la plus fréquemment retrouvée comme symbiote secondaire obligatoire. De plus, la reconstruction de l'histoire évolutive des associations symbiotiques chez *Cinara* semble indiquer une acquisition ancestrale de *Serratia symbiotica* comme co-symbiote obligatoire, même si la présence d'autres bactéries chez certaines espèces indique que des événements de remplacement ont eu lieu. L'analyse des génomes de souches de *Serratia symbiotica* ont montré que ces symbiotes possèdent des caractéristiques génomiques différentes et des variations de leur forme et de leur localisation dans des bactériocytes. Ces données semblent montrer un gradient de modifications dans le passage d'une bactérie facultative à une bactérie endosymbiote fixée sur lequel se trouve l'ensemble des co-symbiotes obligatoires des *Cinara*. Au cours de cette thèse, à l'aide d'approches de phylogénomique, nous avons décrit l'histoire évolutive de *Serratia symbiotica* et mis en évidence de multiples acquisitions indépendantes de cette bactérie comme co-symbiote obligatoire des *Cinara*. Ces transitions d'un mode de vie facultatif à obligatoire sont associées à des réductions de génome et de taux de G-C plus ou moins importantes selon les lignées, révélant l'ancienneté relative de l'association. Parmi les événements d'acquisitions, nous avons montré que l'un d'entre eux avait conduit à une co-spéciation durant environ 20Ma des trois partenaires de la symbiose nutritionnelle (*Cinara*, *Buchnera* et *Serratia*). Dans ce clade, nous avons observé des structures cellulaires et des caractéristiques génomiques stables pour les deux co-symbiotes. De plus, l'analyse des taux de substitutions de chacun des partenaires a permis de décrire pour la première fois une évolution parallèle entre les deux symbiotes et leur hôte. L'analyse détaillée de la composition en gènes des bactéries co-symbiotiques de ce clade a mis en évidence une complémentation forte pour la production des nutriments nécessaires à l'hôte. De plus, l'analyse de l'ensemble de leurs gènes a révélé qu'une forte sélection purifiante s'exerce sur eux. Dans une autre partie, une approche expérimentale de quantification des symbiotes au cours de développement sur trois espèces de *Cinara* suggère une différence de capacité de régulation du co-symbiote *Serratia*. Cette différence est observée entre deux espèces de *Cinara* pour lesquelles les *Serratia* sont issues d'acquisition indépendantes et dans lesquelles sa localisation est différente. Finalement, une analyse de comparaison génomique sur des phages APSE (bactériophage intégré) présents dans les co-symbiotes de *Cinara* et procurant une défense contre les parasitoïdes a permis de décrire la diversité des APSE et de l'évolution de leur génome chez les pucerons. Cette analyse a aussi montré que les éléments mobiles pourraient jouer un rôle dans l'acquisition de nouveaux gènes dans la cassette de la toxine défensive. L'ensemble de ces travaux révèle un système pluri-symbiotique dynamique dans lequel le nouveau partenaire bactérien est régulièrement remplacé au cours de l'évolution.

Evolution and dynamic of di-symbiotic system in *Cinara* Aphids

Abstract : The existence of multi-symbiotic systems has been demonstrated in several aphid species (Hemiptera: Aphididae) and more particularly in species of the genus *Cinara*. The taxonomic characterisation of the symbionts present in about sixty species of *Cinara* has made it possible to determine that *Serratia symbiotica* is the most frequently found bacterium as an obligatory secondary symbiont. Furthermore, the reconstruction of the evolutionary history of symbiotic associations in *Cinara* species suggests an ancestral acquisition of *Serratia symbiotica* as an obligate co-symbiont, although the presence of other bacteria in some species indicates that replacement events have occurred. More recently, genome analysis of *Serratia symbiotica* strains associated have shown that these symbionts exhibit different genomic characteristics ranging and variations in their shape and localisation in bacteriocytes. These data seem to show a gradient of changes in the transition from a facultative life-style to obligate endosymbiont in which we observe all of the (now obligate) co-symbionts of *Cinara*. During this phd, using phylogenomics, we inferred the evolutionary history of *Serratia symbiotica* within *Cinara* and highlighted multiple independent acquisitions of this bacterium as a co-obligate symbiont of *Cinara*. These transitions from facultative to obligate life-style are associated with genome reductions and low G-C content that are more or less important depending on the lineage, revealing the relative age of the symbiotic association with *Serratia*. Among the inferred acquisition events, one in particular led to co-speciation over about 20 Ma of the three partners of the nutritional symbiosis (*Cinara*, *Buchnera* and *Serratia*). In this clade, we observed stable cellular and genomic structures with highly reduced genomes and low G-C content for both co-symbionts. Furthermore, the analysis of the substitution rates of each partner allowed us to describe for the first time a parallel evolution between the two symbionts and their host. Further analysis of the gene composition of the co-symbiotic bacteria in this clade revealed strong complementation for the production of nutrients required by the host. Furthermore, the analysis of their set of genes revealed that they are under strong purifying selection. In addition an experimental approach aimed at quantifying the symbionts during development on three *Cinara* species suggest that aphid species differ in their capacity to regulate the co-symbiont *Serratia*. This difference is observed between two *Cinara* species for which the *Serratia* originated from independent acquisitions and in which its localisation is different. Finally, a genomic comparison analysis on APSE phages (integrated bacteriophage) present in *Cinara* co-symbionts and providing defence against parasitoids allowed a description the diversity of APSEs and their genome evolution in aphids. This analysis also showed that mobile elements could play a role in the acquisition of new genes in the defensive toxin cassette. Altogether, this study reveals a dynamic multi-partner endosymbiosis where the new partner is recurrently replace.