



**HAL**  
open science

# Diversité génétique et admixture au sein du complexe d'espèces *Bemisia tabaci*: contributions des compartiments nucléaires et cytoplasmiques

Gabriel Terraz

► **To cite this version:**

Gabriel Terraz. Diversité génétique et admixture au sein du complexe d'espèces *Bemisia tabaci*: contributions des compartiments nucléaires et cytoplasmiques. Génétique des populations [q-bio.PE]. Université de Lyon, 2016. Français. NNT: 2016LYSE1112 . tel-01430425

**HAL Id: tel-01430425**

**<https://theses.hal.science/tel-01430425>**

Submitted on 9 Jan 2017

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

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



N° d'ordre NNT : 2016LYSE1112

**THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON**

opérée au sein de

**l'Université Claude Bernard Lyon 1**

**École Doctorale : Evolution, Ecosystèmes, Microbiologie et Modélisation**

**Spécialité de doctorat : Biologie**

Soutenue publiquement le 6 juillet 2016, par :

**Gabriel TERRAZ**

---

**Diversité génétique et admixture au sein du complexe  
d'espèces *Bemisia tabaci* : contributions des compartiments  
nucléaires et cytoplasmiques.**

---

Devant le jury composé de :

Olivier Duron,

Rapporteur

Xavier Fauvergue,

Rapporteur

Cristina Vieira,

Examinatrice

Denis Bourguet,

Examineur

Frédéric Fleury,

Directeur de thèse

Laurence Mouton,

Co-directrice de thèse





## UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université	M. le Professeur Frédéric FLEURY
Président du Conseil Académique	M. le Professeur Hamda BEN HADID
Vice-président du Conseil d'Administration	M. le Professeur Didier REVEL
Vice-président du Conseil Formation et Vie Universitaire	M. le Professeur Philippe CHEVALIER
Vice-président de la Commission Recherche	M. Fabrice VALLÉE
Directeur Général des Services	M. Alain HELLEU

### COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard	Directeur : M. le Professeur J. ETIENNE
Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux	Directeur : Mme la Professeure C. BURILLON
Faculté d'Odontologie	Directeur : M. le Professeur D. BOURGEOIS
Institut des Sciences Pharmaceutiques et Biologiques	Directeur : Mme la Professeure C. VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	Directeur : M. le Professeur Y. MATILLON
Département de formation et Centre de Recherche en Biologie Humaine	Directeur : Mme la Professeure A-M. SCHOTT

### COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies	Directeur : M. F. DE MARCHI
Département Biologie	Directeur : M. le Professeur F. THEVENARD
Département Chimie Biochimie	Directeur : Mme C. FELIX
Département GEP	Directeur : M. Hassan HAMMOURI
Département Informatique	Directeur : M. le Professeur S. AKKOUCHE
Département Mathématiques	Directeur : M. le Professeur G. TOMANOV
Département Mécanique	Directeur : M. le Professeur H. BEN HADID
Département Physique	Directeur : M. le Professeur J-C PLENET
UFR Sciences et Techniques des Activités Physiques et Sportives	Directeur : M. Y. VANPOULLE
Observatoire des Sciences de l'Univers de Lyon	Directeur : M. B. GUIDERDONI
Polytech Lyon	Directeur : M. le Professeur E. PERRIN
Ecole Supérieure de Chimie Physique Electronique	Directeur : M. G. PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. le Professeur C. VITON
Ecole Supérieure du Professorat et de l'Education	Directeur : M. le Professeur A. MOUGNIOTTE
Institut de Science Financière et d'Assurances	Directeur : M. N. LEBOISNE



## Résumé

Les invasions biologiques ont des conséquences écologiques importantes, notamment l'émergence de ravageurs. En situation invasive, les populations font face à de nouvelles conditions biotiques et abiotiques qu'elles doivent surmonter. Du fait de l'introductions de multiples populations et de leur hybridation, une augmentation de la variabilité génétique nucléaire peut-être observée, support d'une réponse adaptative plus rapide.

De plus, chez les insectes, les bactéries symbiotiques peuvent jouer un rôle important dans l'adaptation.

Le ravageur de cultures *Bemisia tabaci* est un complexe d'espèces dont les barrières reproductives sont peu connues et dont les différents biotypes sont associés à des symbiotes qui leur sont spécifiques. En France, nous avons une situation idéale car deux biotypes sont en sympatrie, et sur la même plante.

Dans le cadre de cette thèse, je me suis posé plusieurs questions relatives à l'influence des compartiments cytoplasmique et nucléaire sur la réussite de l'invasion.

- Quelle est la dynamique spatio-temporelle de ces biotypes ?
- Est-ce que l'admixture de ces biotypes peut engendrer une augmentation de la diversité génétique dans l'aire d'introduction ?
- Quelle est la dynamique des associations symbiotiques, notamment dans les cas de coexistence de différents biotypes ?

Grâce à des outils de génétique des populations nous avons montré l'hybridation de ces biotypes en populations naturelles, capacité à s'hybrider que nous avons confirmée en laboratoire.

Finalement, avec les profils d'infection des populations naturelles et des expérimentations contrôlées, nous nous sommes posé la question de la stabilité de ces associations de bactéries.



## Abstract

Biological invasions have important ecological consequences, including the emergence of pests. In invasive condition, populations face new biotic and abiotic conditions. Due to the introduction of multiple populations and their hybridization, an increase in the nuclear genetic variability may be observed, supporting a faster adaptive response.

Moreover, in insects, symbiotic bacteria play an important role in adaptation.

The pest *Bemisia tabaci* is a complex of species whose reproductive barriers are little-known and whose different biotypes are associated with symbiont specific to them. In France, we have an ideal situation as both biotypes are sympatric, often on the same plant.

As part of this thesis, I asked myself several questions regarding the influence of the cytoplasmic and nuclear compartments on the success of the invasion.

- What is the spatiotemporal dynamics of these biotypes?
- Does the admixture of these biotypes can increase the genetic diversity in the area of introduction?
- What is the dynamic of these symbiotic associations, especially in the case of coexistence of different biotypes?

Through population genetic tools we showed the hybridization of these biotypes in natural populations, ability to hybridize we have confirmed in the lab.

Finally, with infection profiles of natural populations and controlled experiments, we asked ourselves about the stability of these bacterial associations.



# Remerciements

Je voudrais tout d'abord remercier mes deux directeurs, Laurence et Fred, qui m'ont encadré durant ces quelques années (autant rester vague sur le nombre exact). Merci d'avoir porté ce projet et obtenu ce financement, grâce auquel j'ai été payé, et qui m'a permis de faire mes recherches.

Merci Laurence pour ta disponibilité au quotidien, ton aide précieuse dans la rédaction du manuscrit, toutes les remarques constructives que tu m'as faites et les éclairages que tu m'as apportés sur certains concepts pas toujours très clairs pour moi... Merci pour ta patience... et tes coups de gueules (c'est sincère !), et surtout merci d'avoir tenu jusqu'au bout. Fred, je te remercie pour toutes tes idées et remarques que tu es capable de sortir en quelques minutes de conversation, et les « yapluka » qui suivent ! Merci pour le recul que tu as apporté aux problématiques qu'on a abordées, mais attention à ne pas trop te mettre en recul par rapport à la science ! Il faut tout de même admettre que Président de l'Université Claude Bernard Lyon 1 ça en jette !

Je tiens aussi à remercier Patricia pour la dernière ligne droite (bien que sinieuse eut été un terme plus approprié), sans ton aide je ne serais probablement pas en train d'écrire ces lignes !

Merci aux membres de mon jury, qui ont accepté de lire et juger mon travail : Cristina, Denis, Xavier et Olivier ainsi qu'aux membres de mes comités de pilotage, Matthieu, Alexandre, Fabrice et Patrick.

Merci Fabrice pour tous les conseils et les explications, sur de nombreux concepts, que tu m'as donnés.

Merci Julien Vendeville pour ton implication dans l'envoi d'échantillons qui ont été nécessaire à mon travail, ainsi que pour m'avoir fait découvrir les serres de production. Tu remercieras aussi tes collègues que tu as impliqué dans ces collectes.

Merci Alain Ferre et Thomas Hebbinckuys pour le travail que nous avons fait ensemble, ainsi que Laurence Arvanitakis et Dominique Bordat.

Merci à toute l'équipe GEIHP, pour l'accueil, pour la bonne ambiance (de travail bien sûr), pour les débats passionnés couvrant une large gamme de thématiques allant du fonctionnement de la recherche en France au meilleur club de foot en passant par des discussions stériles sur la pertinence d'un mot ou d'une flèche sur une diapo. Merci pour les conseils et la disponibilité, et puis pour tous les moments de convivialité autour d'un (ou plusieurs) verre(s).



Merci à tous ceux qui ont apporté une aide technique à mon travail : Sandrine, Nicole, Sonia, Gladys, Élise, Denis et Anaël.

Merci à Nathalie, Laetitia Aline et Odile pour tous les aspects administratifs que j'affectionne tant, particulièrement dans l'urgence et bien sûr pour les autres moments plus rigolos. Merci encore !

Un gros coucou à tous les membres du LBBE !

Merci Taisiia pour ton travail lors de ton stage, mais aussi pour les à-côtés comme la véritable vodka que tu nous a fait goûter à plusieurs reprises ! Merci aussi Laurine pour toutes ces magnifiques PCRq que tu as faites, bien que tu aurais pu éviter de mettre en évidence des contaminations ! Et félicitations Thais pour continuer ces manips !

Merci à tous les autres galériens, qu'ils soient encore ici ou partis depuis quelque temps, Cricri, Olivier, David, Julien, Marie, Mathilde, Émilie, Paf, Anna, Julien, Lucie, Vincent, Mathieu... Vous tous avec votre nette tendance à l'alcoolisme, bien que certains ne soient pas définitivement perdus, je vous remercie pour tous les moments sympas, au labo et surtout dans les lieux de perdition.

Merci à tous les grands humoristes du quotidien, particulièrement mes collègues de bureau actuels, Dave, Oliver et Julien dont la subtilité des blagues n'a d'égal que leur finesse. Anaël, tu ne démérites pas non plus dans les jeux de mots, bien que tu sois parti depuis un petit moment !

Merci Dardon pour ces 6 mois, ta compagnie bisounours-hippie a été très agréable et tu as relevé le niveau de finesse de notre bureau qui a été assez bas pendant cette période !

Parce que certains préfèrent rester anonyme, je remercie aussi slither, jetpunk, 2048 et toute la bande de rigolos qui ont amélioré le quotidien et nous ont bien fait marrer ! Parce que des fois, on a vraiment rien d'intéressant à se dire, merci à l'auteur du générique de Tintin : ta musique vaut parfois largement mieux qu'une discussion dans notre bureau, ou même que 10 minutes de répliques d'un film, répétées chaque jour par certains amateurs passionnés qui se reconnaîtront. Juju, j'attends toujours de jouer au billard chez toi, bien que les fléchettes et le palet à Bourgoin soient pas mal non plus. Merci Oliver pour tes tentatives de soirées cinéma où la perte de contrôle n'est jamais très loin, et pour cette fameuse danse que tu affectionnes tant.

Parce que travailler en équipe c'est être poli au quotidien, je remercie tous ces Bonjours que nous avons échangés, particulièrement avec Dave parce qu'il sait apprécier les différentes manières de se saluer et qu'il est conscient que de dire bonjour le nez bouché peut mener aux grandes vacances ; mais je remercie aussi ceux échangés avec Marie, bien qu'amatrice moins éclairée.

Marie, tant que t'es là, merci d'avoir partagé l'aventure radseq avec Émilie et moi, ou plutôt la galère radseq, sport nautique qui se pratique sur une planche à repasser en fonte, sans rame et avec beaucoup de blocs de béton à transporter, par nuit noire et orageuse. Je te remercie Sylvain

pour cette chouette idée de sport avec toutes les règles que tu as ajoutées en cours de route, et bien sûr Hélène qui a maintenu cet équipage à flot de toute son énergie et sa bonne humeur disponible !

Hélène, en plus d'être ingénieure en navigation, tu es ingénieure en biomol, manip et plein d'autres disciplines et ça c'est carrément la classe ! Je te remercie pour toute ces connaissances, conseils et remarques que tu m'as apporté durant ce cours laps de temps qu'a été ma thèse.

Merci à t-i-b-o, qui m'a grandement aidé pour de nombreuses manips et qui a maintenu l'élevage, source de ces charmantes manips. Merci d'avoir bien voulu galérer à manipuler ces insectes avec bonne humeur. En plus d'être le technicien/esclave/ingénieur/assistant (prend ce que tu veux), t-i-b-o est le photographe animalier officiel de *Bemisia*, et donc l'auteur de clichés que vous trouverez dans ce manuscrit, et de plein d'autres clichés de superbes insectes dans la vie civile. En plus de tout ça on s'est bien marré !

Merci à la goub, pour avoir partagé cette passion pour la chute des aéronefs mais surtout pour l'émission qui nous les explique avec une parfaite maîtrise du documentaire et de la mise en scène. Émission qui nous fait vibrer à chaque fois qu'un épisode est diffusé sur ces fameuses chaînes que tu affectionnes tant, avec des sujets aussi passionnant que l'ouverture d'un coffre-fort, la construction de cabanes et j'en passe. Merci pour tous ces chouettes moments, mais aussi pour les discussion plus sérieuses, mais surtout pour avoir élevé la dégustation de bière dans ce labo, avec l'aide indéfectible de cricri. Merci à tous les deux ! Merci à toi cricri pour me rappeler chaque jour que l'abus de bbq + hk n'est pas bon pour le palais, plus sérieusement merci pour les dégustations que nous avons partagées et pour lesquelles on a rarement été d'accord !

Merci Pôle Emploi pour avoir financé cette dernière année, ou plutôt merci à toutes les personnes qui cotisent, c'est plutôt chouette.

Normalement on remercie chaudement sa famille et *gnagnagna...*, mais faut avouer que votre implication dans la thèse est proche du zéro absolu, non ? Ça change pas le fait que vous soyez chouette, ne vous inquiétez pas. Enfin, tout de même, merci pour l'apéro !

En ce qui concerne les amis, je ne vous remercierais pas parce que franchement « T'es toujours en thèse ? » « C'est quand que tu finis ? » et autres amabilités en tout genre, j'ai assez donné !

Enfin, je voudrais bien sûr remercier, ou plutôt féliciter Sarah, pour avoir tenu tout ce temps, et particulièrement cette dernière année qui n'a pas été la plus facile. Félicitations !

Encore merci à tout le monde, et particulièrement à ceux que j'ai oubliés.

Gabriel



# Sommaire

Résumé	v
Abstract	vii
Remerciements	ix
Sommaire	xiii
Introduction	1
Problématique générale	35
Le modèle d'étude <i>Bemisia tabaci</i>	37
1 Description et dynamique spatio-temporelle des cytotypes	61
2 Analyse du compartiment nucléaire des cytotypes	81
3 Émergence d'un nouveau cytotype	107
4 Développement et utilisation des marqueurs <i>RADSeq</i> pour le modèle <i>B. tabaci</i>	127
Discussion générale	161
Travaux publiés	169
Annexes	219
Références bibliographiques	243
Liste des figures	255
Liste des tableaux	257
Table des matières	263



# Introduction

1	La génétique des invasions biologiques	3
1.1	Définition et notion d'espèce invasive	3
1.1.1	Qu'est ce qu'une espèce indigène ?	3
1.1.2	Les conséquences des invasions peuvent être plus subtiles, voire positives	5
1.1.3	Des effets variables dans le temps et l'espace	7
1.2	Déroulement d'une invasion	8
1.3	Qu'entend on par adaptation ?	10
1.3.1	L'admixture, facteur clé dans les processus adaptatifs	12
1.3.2	Augmentation du potentiel adaptatif	13
1.4	Complexe d'espèces et invasions	14
1.4.1	Comment définit-on une espèce ?	14
1.4.2	Complexe d'espèces et invasion	15
1.5	Interaction avec le compartiment cytoplasmique	19
2	La symbiose, actrice de l'invasion ?	20
2.1	Définition, brève histoire du concept de symbiose	20
2.2	Transmission de ces endosymbiotes	22
2.3	Les symbiotes obligatoires	22
2.3.1	Localisation tissulaire	23
2.3.2	Diversité et rôles de ces symbiotes	23
2.3.3	Les symbioses primaires sont des associations anciennes	25
2.4	Les symbiotes secondaires	26
2.4.1	Les effets phénotypiques de ces symbiotes	27
2.4.2	La transmission horizontale des symbiotes secondaires	29
2.4.3	Les multi-infections	29
2.5	Les manipulateurs de la reproduction	31
2.6	Quel rôle peut jouer la symbiose dans l'invasion biologique ?	32
2.6.1	Résistance contre des moyens de lutte	32
2.6.2	Augmentation du potentiel invasif de l'hôte	32



*Ils renouvelèrent leurs tentatives sur des poules et un canard, sur un dogue et une truie, avec l'espoir qu'il en sortirait des monstres et ne comprenant rien à la question de l'espèce.*

*Ce mot désigne un groupe d'individus dont les descendants se reproduisent. Mais des animaux classés comme d'espèces différentes peuvent se reproduire, et d'autres compris dans la même en ont perdu la faculté.*

– Gustave Flaubert, *Bouvard et Pécuchet*

## La génétique des invasions biologiques

### Définition et notion d'espèce invasive

On attribue le terme d'invasion biologique à Charles Elton qui publia *Ecologie des invasions par les animaux et les plantes*<sup>1</sup> (Elton 1958), mais il ne définit pas directement le concept d'invasion biologique, c'est ainsi que naît la biologie des invasions, qui ne connaîtra son véritable essor que dans les années 1980.

### Qu'est ce qu'une espèce indigène ?

Les espèces dites invasives sont définies comme des espèces non indigènes (*native* dans la littérature anglophone) s'établissant et se diffusant largement hors de leur aire d'origine. Le premier problème est de définir ce qu'est une espèce indigène. Ce concept a d'abord été esquissé par un botaniste (John Henslow) en 1835, puis par ses confrères dans les années qui suivirent pour distinguer la « vraie » flore anglaise des autres, dites exotiques car elles ont bénéficié de l'aide de l'homme (Davis *et al.* 2011). Cette distinction s'inscrit dans notre héritage culturel qui place l'homme en dehors de la nature, et n'est pas neutre puisque les espèces indigènes sont implicitement déclarées légitimes au contraire des autres qui sont intruses, puis avec le temps elles deviendront même indésirables (Tassin 2014). En plus, ce critère n'a pas de fondement biologique mais est avant tout culturel, les espèces invasives sont celles introduites après 1500 en Europe et aux États-Unis, mais c'est 1778 qui est la date retenue en Nouvelle-Zélande. Dans ce même pays, des espèces introduites par les polynésiens ne seront pas considérées comme invasives car le « polynésien » est considéré comme indigène et fait partie de la nature<sup>2</sup>. Une définition souvent rencontrée dans la littérature est celle de Pysek (1995) qui donne le statut d'indigène à toute espèce présente dans une région sans l'aide des activités humaines. Les espèces introduites par l'homme avant le néolithique doivent être considérées comme indigènes puisqu'à cette époque l'homme était partie intégrante de la nature (Webb 1985).

1. Le livre commence par ces termes : *Nowadays we live in a very explosive world, and while we may not know where or when the next outburst will be, we might hope to find ways of stopping it or at any rate damping down its force. It is not just nuclear bombs and wars that threaten us, though these rank very high on the list at the moment : there are other sorts of explosions, and this book is about ecological explosions.*

2. On trouve ici le mythe du bon sauvage (Tassin 2014, chap. 2)



## Introduction

Il est utile de mentionner le fait qu'opposer les espèces selon leur origine n'a pas vraiment de sens. Une espèce indigène n'est pas le signe d'une valeur sélective supérieure et n'est pas non plus synonyme d'effets bénéfiques : le coléoptère *Dendroctonus ponderosae* est l'insecte qui tue le plus d'arbres en Amérique du nord malgré le fait qu'il soit indigène de la région (Davis *et al.* 2011). Le fait que nous qualifions des espèces d'invasives sous entend qu'elles ne sont pas légitimes. Par exemple, personne ne remet en cause la légitimité de la présence des caméléons qui ont envahis de nombreuses régions dans l'océan Indien à partir de Madagascar (Rieppel 2002) ; par contre nous considérons comme invasives les populations d'écureuils gris en Europe ! Les nombreux cas de recolonisation post-glaciaire, par exemple la truffe noire *Tuber melanosporum* (Murat *et al.* 2004) ou le carabe *Carabus solieri* (Garnier *et al.* 2004), nous apparaissent normaux et à aucun moment nous envisagerions ces taxons comme invasifs.

Bien sûr certaines espèces introduites par l'homme ont eu des effets dramatiques et il ne s'agit pas de nier le phénomène mais de l'aborder de manière plus rationnelle. Tous les grands groupes taxonomiques sont représentés par des espèces invasives : plantes, mammifères, oiseaux, reptiles, amphibiens, poissons, arthropodes, mollusques, microbes ; ainsi que tous les continents et tous les climats (Pimentel *et al.* 2000). Les conséquences de ces invasions peuvent être multiples (Vitousek & D'Antonio 1997) et leur impact économique se chiffrer en milliards de dollars (Pimentel *et al.* 2000), notamment à cause de l'émergence de nouveaux ravageurs et de pathogènes de plantes ou d'animaux. La moule *Dreissena polymorpha* introduite accidentellement en Amérique du nord depuis la Russie dans les années 1980 a causé des dégâts très importants en bouchant les canalisations de distribution d'eau ou de l'industrie de l'énergie. En terme écologique, le paludisme aviaire introduit avec son vecteur *Culex quinquefasciatus* à Hawaï au début du vingtième siècle est impliqué dans le déclin de la faune aviaire de l'île (Lapointe *et al.* 2012). Il y a aussi le cas des ravageurs de cultures, dont de nombreux insectes : le phylloxéra, le doryphore, les criquets, les cochenilles, charançons, drosophiles et autres aleurodes, mais aussi les rats, lièvres, passereaux, nématodes et mollusques (Voir la base de données DAISIE<sup>1</sup>). Les dégâts occasionnés à l'agriculture par exemple sont considérables (14 milliards de dollars par an aux États-Unis, Pimentel *et al.* 2000), du fait de plantes adventices comme le paspalum (*Paspalum distichum*), de ravageurs de culture comme *Ceratitidis capitata* (mangeur de fruits) ou *Bemisia tabaci* (vecteur de virus phytopathogènes)... Un autre exemple est celui de l'ambrosie, plante originaire d'Amérique dont les pollens sont très allergènes, occasionnant un coût en terme de santé et donc en terme économique (Genton *et al.* 2005). Les espèces invasives ont des conséquences sur les écosystèmes où tous les processus écologiques peuvent être altérés : production primaire, prédation, décomposition, hydrologie (Vitousek & D'Antonio 1997) ; ainsi que la biodiversité par compétition avec les espèces autochtones ou l'apport de pathogènes pour les espèces locales, ceci pouvant mettre en péril leur survie (Williamson &

---

1. DAISIE (Delivering alien invasive species in Europe) est une base de données européenne sur les espèces invasives, financée par la commission européenne au titre du sixième programme-cadre pour la recherche et le développement technologique (2002-2006)

Fitter 1996). Les échanges commerciaux jouent un grand rôle dans ces invasions. En effet, les marchandises transportées par avions et bateaux sont les voies majeures pour les introductions. La modification des habitats (urbanisation, agriculture, canaux de navigation) les rend aussi plus susceptibles (Vitousek & D'Antonio 1997).

### **Les conséquences des invasions peuvent être plus subtiles, voire positives**

Souvent les conséquences des invasions ne sont pas perçues de la même manière par les différents acteurs de la société civile (Simberloff *et al.* 2013). Par exemple l'introduction de la crevette *Marsupenaeus japonicus* en méditerranée depuis la mer rouge a été perçue comme positive par les pêcheurs qui augmentent leurs prises mais négative par les défenseurs de l'environnement car elle entraîne le déclin d'une autre espèce *Melicertus kerathurus* (Galil 2007). De plus, certaines espèces peuvent avoir des effets bénéfiques et négatifs : le poisson *Gambusia affinis* a été largement introduit comme prédateur de larves de moustiques alors qu'il a un effet négatif sur des espèces d'insectes, poissons et amphibiens (Goodsell & Kats 1999).

Alors que les invasions biologiques sont souvent accusées d'éroder la biodiversité, Schlaepfer *et al.* (2011) citent de nombreux effets induits par des espèces non indigènes introduites accidentellement qui contribuent au maintien ou à la restauration de la biodiversité : habitat, abri, ressource, catalyseur de la restauration d'écosystèmes, pollinisation... En fait il existe peu de données précises pour quantifier les effets négatifs induits par des espèces non indigènes sur la biodiversité locale. Gurevitch & Padilla (2004) montrent que les espèces non indigènes ne représentent pas un danger particulier pour la biodiversité, bien que des fortes réductions de densité des populations indigènes puissent être observées (Davis 2009). Ceci n'est par contre pas vrai pour les milieux insulaires et les lacs (Wilcove *et al.* 1998) qui sont beaucoup plus susceptibles. Les espèces invasives sont souvent accusées d'éroder la biodiversité car les extinctions peuvent coïncider avec l'apparition des non indigènes. Pourtant quand les données sont examinées plus précisément leur rôle direct dans l'extinction peut être remis en question. Gurevitch & Padilla (2004) citent deux exemples classiques d'espèces indigènes qui sont peut-être accusées à tort de nuire à la biodiversité. La perche du Nil *Lates niloticus* introduite dans le lac Victoria durant les années 1960 est accusée d'être impliquée dans la disparition des cichlidés locaux, mais ce déclin a probablement commencé 40 ans auparavant avec l'érosion des rivages due au développement économique. Ce déclin s'est accentué dans les années 1970 avec l'eutrophisation du lac du fait de l'urbanisation et les nombreux épisodes d'anoxie qui ont suivis. Pour finir, l'eutrophisation a favorisé l'installation de la jacinthe d'eau *Eichhornia crassipes* [Pontederiaceae] qui dégrade les frayères des poissons. Leur deuxième exemple est celui de la moule zébrée *Dreissena polymorpha* considérée comme étant une menace majeure pour les bivalves unionoïdes d'eau douce en Amérique du nord (Ricciardi *et al.* 1998). Pour se développer la moule zébrée a besoin d'un substrat dur, rare dans la plupart des lacs. Elle s'attache donc aux coquilles des bivalves locaux, ce qui perturbe leur développement. Malgré tout, le rôle de la moule zébrée dans le déclin des bivalves est incertain puisque qu'il a commencé bien avant l'introduction de

## Introduction

la moule. L'eutrophisation, les pesticides, l'industrie de la nacre et la disparition des poissons hôtes des bivalves unionoïdes<sup>1</sup> joueraient en fait un rôle clé dans la disparition des bivalves.

Voici deux exemples qui montrent l'utilité que peuvent avoir certaines espèces non indigènes. A Puerto-Rico les paysages anciennement occupés par l'agriculture sont facilement colonisés par des espèces d'arbres non indigènes telles que le tulipier du Gabon *Spathodea campanulata* [Bignoniaceae], mais très peu par la flore locale à cause des températures élevées, du faible taux d'humidité ainsi que des sols compacts et peu fertiles. L'établissement des arbres indigènes facilite le développement d'autres espèces, tant indigènes qu'allochtones ce qui crée une communauté écologique originale à long terme à même de restaurer les sols et la biodiversité de paysages dégradés (Lugo 2004). En Pennsylvanie, l'abondance des oiseaux frugivores indigènes est fortement corrélée à la disponibilité en fruits des arbustes du genre *Lonicera* [Caprifoliaceae], invasifs dans cette région. Ainsi, les auteurs de l'étude suggèrent d'être prudent dans la conduite des programmes d'éradication de ces arbustes (Gleditsch & Carlo 2011).

Dans une méta-analyse, Stromberg *et al.* (2009) montrent comment la communauté scientifique peut jouer un rôle primordiale dans la perception d'une espèce non indigène, en l'occurrence le tamaris. Initialement introduit aux États-Unis pour contrôler l'érosion, le tamaris est devenu ensuite *persona non grata*, accusé de consommer beaucoup d'eau et d'être à l'origine de divers problèmes environnementaux. Des programmes d'éradication très coûteux ont été mis en place jusqu'à ce que les recherches récentes montrent son importance comme membre à part entière des communautés écologiques ripariennes, et l'intérêt de l'intégrer dans les programmes de restauration des écosystèmes. Clairement, la communauté scientifique a maintenu une image négative de cet arbre en citant des travaux dépassés, en déduisant des relations de causes à effets à partir de simples corrélations, en tirant des conclusions qui dépassent le domaine d'étude et plus généralement en présentant l'espèce comme un agent non naturel des changements en cours et en utilisant un vocabulaire à connotation guerrière (Stromberg *et al.* 2009).

Une meilleure objectivité dans l'approche que nous avons des invasions biologiques pourrait être envisageable en parlant d'écologie du mouvement (Nathan *et al.* 2008), discipline qui a maintenant sa revue (Nathan & Giuggioli 2013). Plutôt que biologie de l'invasion, on supprimerait ainsi les nombreux *a priori* négatifs sur la question amenés par l'emploi du mot invasif. Ainsi, classifier les biotopes selon nos standards culturels ou les bénéfiques qu'ils apportent (ou non) n'a pas de sens dans une démarche scientifique et dans notre compréhension des mécanismes évolutifs et du fonctionnement des écosystèmes. Il faut donc être rationnel et ne pas juger une espèce juste par sa provenance et considérer qu'un milieu était mieux avant par principe (Davis *et al.* 2011).

---

1. Le cycle de vie de ces bivalves unionoïdes impliquerait une relation mutualiste avec certains poissons : leur stade larvaire serait « parasite » des branchies des poissons, mais à l'état adulte ces bivalves protégeraient les œufs de ces poissons (Wisniewski *et al.* 2013).

### Des effets variables dans le temps et l'espace

Un problème posé par le concept d'espèce invasive est le mot même d'« espèce » puisque ce ne sont que certaines populations d'une espèce qui sont invasives (Simberloff *et al.* 2013). Nous aurons l'occasion de revenir sur la notion d'espèce, pour le moment nous ne ferons que rappeler la définition de Mayr qui stipule qu'une espèce est la plus grande unité de populations au sein de laquelle le flux génétique est possible. Cette définition est suffisante pour appréhender ce qui suit.

Certaines espèces peuvent être considérées comme nuisibles dans certaines régions où elles ont peu d'impacts, simplement par le fait qu'elles sont considérées comme nuisibles ailleurs (Colautti & Macisaac 2004). En effet, toute espèce peut devenir invasive si elle rencontre des conditions favorables. Par exemple le lapin de garenne est en fort déclin dans la péninsule ibérique (entraînant avec lui le déclin des populations de lynx (Moreno *et al.* 2007), mais il est considéré comme nuisible localement, ailleurs en Europe et particulièrement en Australie. Un autre exemple est celui du crapaud de la cane *Rhinella marina* originaire d'Amérique du sud où il ne pose pas de problème particulier. Par contre il a été introduit dans de nombreuses régions du monde pour lutter contre divers ravageurs comme les coléoptères, où il est devenu invasif, comme en Australie. Finalement il s'est avéré que les écosystèmes furent beaucoup plus résilients que prévu et que le crapaud peut être bénéfique pour diverses espèces (Shine 2010). On peut citer aussi les cas de deux insectes : *Drosophila suzukii* qui est devenue un ravageur très problématique en Europe et aux États-Unis car dévastant de nombreuses cultures (fraises, cerises, pêches, abricots, prunes...) mais n'est pas un problème dans son aire d'origine (Chine et Corée du Sud, Asplen *et al.* 2015). Un autre exemple est celui du coléoptère xylophage *Anoplophora chinensis* [Cerambycidae] s'attaquant à de nombreux arbres, notamment les *Citrus*, et qui ne pose pas de problèmes particulier en Chine d'où il est originaire (Haack *et al.* 2010).

Notons encore que nous manquons de recul sur ces déplacements de populations, on sait aujourd'hui que des espèces qui étaient fortement invasives ne le sont plus aujourd'hui, probablement car le milieu s'est stabilisé après des fortes perturbations anthropiques. Pour une synthèse sur ces invasions qui « disparaissent », voir Simberloff & Gibbons (2004). Tassin (2014) cite deux exemples : l'élodée du Canada, *Elodea canadensis* [Hydrocharitaceae] et la caulerpe, *Caulerpa taxifolia* [Caulerpaceae]. La première, introduite dans les années 1830 en Irlande depuis l'Amérique du nord, a envahi de très nombreux cours d'eau en Europe mais elle est aujourd'hui en régression. La deuxième, surnommée la « peste verte » ou encore « l'algue tueuse » (noter le vocabulaire utilisé), est apparue en 1984 sur les côtes méditerranéennes françaises (échappée de l'aquarium de Monaco) et est devenue rapidement envahissante. En 1990, la superficie colonisée par l'algue atteint 15 000 hectares en France, Espagne, Italie, Croatie et Tunisie, mais aujourd'hui elle a presque disparu de ces régions. Les effets d'une espèce peuvent ainsi varier dans le temps et l'espace, et des espèces extérieures qui provoquent des dégâts aujourd'hui n'en provoqueront pas forcément demain.

## Introduction

Les espèces non indigènes sont en fait des *populations* d'espèces, non indigènes. Pour définir plus clairement le concept, de façon neutre, sans vocabulaire connoté négativement tel que « invasif », Colautti & Macisaac (2004) ont proposé un cadre qui s'appuie sur les différentes étapes d'un processus invasif (les auteurs eux-même ne réussissent pas à se débarrasser du terme « invasif » dans leur article), ainsi que sur l'abondance et la distribution des populations étudiées. Dans ce cadre, une population non indigène peut être restreinte à une région et à faible densité, répandue mais rare, localisée et dominante ou encore répandue et dominante. Ainsi on fait référence à des populations individuelles plutôt qu'à des espèces. Le même changement de paradigme a eu lieu en biologie de la conservation, des espèces vers les populations, ce qui modifie profondément les pratiques de gestion et les questions théoriques sous-jacentes (Luck *et al.* 2003). Bien sûr, le niveau spécifique a permis de nombreux travaux puisque plusieurs populations envahissantes peuvent appartenir à la même espèce, et sont des échantillons non aléatoires de celle-ci (Kolar & Lodge 2001). Il est donc primordial de se concentrer sur des populations, car elles peuvent avoir des dynamiques différentes au sein d'une même espèce suivant l'environnement dans lequel elles évoluent. Les processus qui mènent à l'établissement d'une population dans une nouvelle aire ne sont pas forcément les mêmes qui ont mené à l'établissement dans une autre aire (Thompson *et al.* 1995).

Il n'existe donc pas de définition unique de l'invasion biologique, on considérera qu'une invasion résulte de la conjonction d'une espèce, d'un milieu et d'une histoire (Tassin 2014).

## Déroulement d'une invasion

Une invasion biologique est classiquement décomposée en plusieurs étapes (voir figure 1). Après une phase de transport depuis l'aire d'origine vers l'aire d'introduction (naturelle, ou plus souvent dans le sillage des activités humaines), les individus arrivent dans leur nouvel environnement. Les premiers arrivants doivent initier une nouvelle population viable et « autosuffisante » (Sakai *et al.* 2001). Beaucoup de traits liés à l'auto-fertilisation pourrait-être facilitant comme chez les plantes, ou certains insectes capables de stocker le sperme ou encore capables de parthénogenèse, car un seul individu est alors nécessaire dans la création de la nouvelle population (Sakai *et al.* 2001)

Les arrivants doivent alors s'adapter à des conditions environnementales qui leurs sont nouvelles. Notons que cette hypothèse n'est pas forcément vérifiée : peut-on dire que des insectes que l'on retrouve exclusivement en serre dans les pays tempérés tels que les aleurodes ou les thrips par exemple font face à des conditions nouvelles dans toutes les exploitations où on les retrouve ? Mais même si l'environnement est favorable, pour de nombreuses raisons (absence de pathogènes ou prédateurs, niche disponible, climat...), celui-ci est toujours un peu différent et nécessite donc un ajustement : qu'il soit directement potentialisé (plasticité) ou plus indirectement par adaptation génétique.

La plasticité phénotypique est la capacité d'un organisme à exprimer différents phénotypes



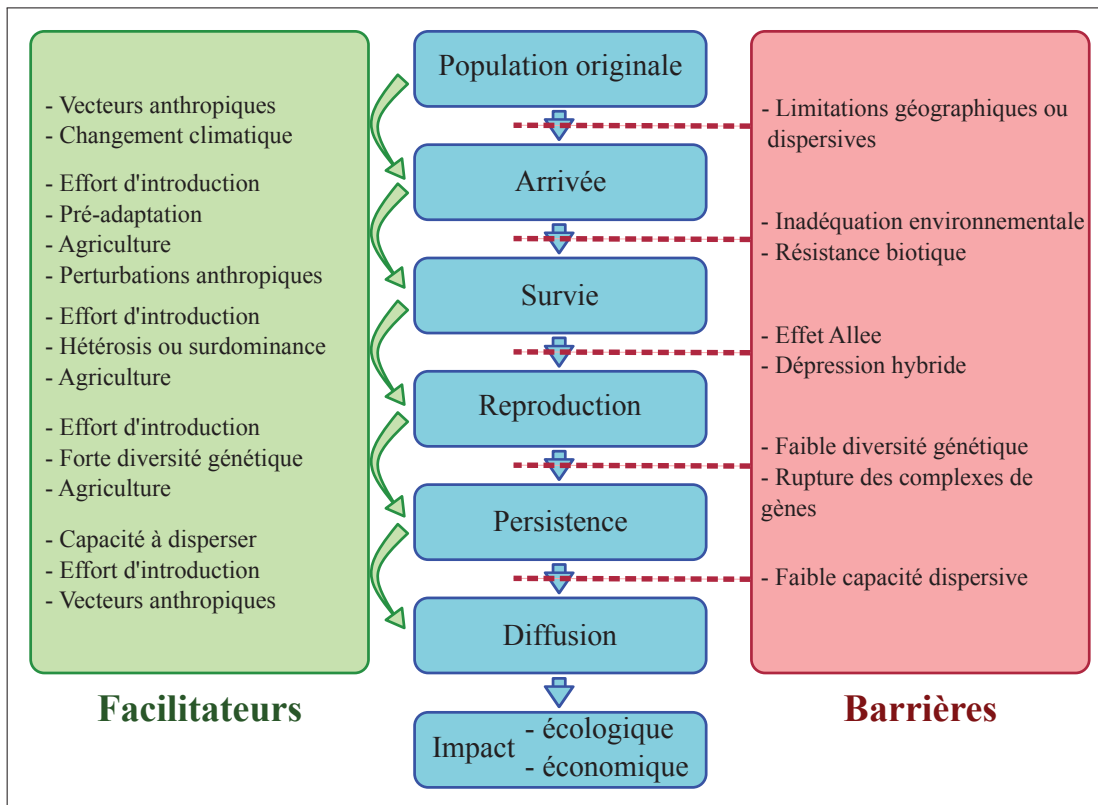
à partir d'un génotype donné selon les conditions environnementales. La norme de réaction décrit la gamme des phénotypes produits par ce même génotype. Par exemple la flèche d'eau *Sagittaria sagittifolia* [Alismataceae] présente trois types de feuilles : des feuilles dressées hors de l'eau en forme de flèche aiguë, des feuilles nageantes arrondies en forme de cœur, et des feuilles immergées allongées en forme de ruban. La plasticité phénotypique est bien sûr un atout, car l'individu plastique peut potentiellement s'adapter à de nombreux environnements. À l'inverse, l'adaptation génétique désigne les processus évolutifs qui font varier la composition génétique d'une population et augmente sa valeur sélective. Nous allons revenir sur cette notion par la suite.

Dans un deuxième temps, une phase de latence entre l'introduction et l'expansion géographique permet l'adaptation aux nouvelles conditions (qu'elles soient biotiques ou abiotiques), cette phase est probablement critique. Selon le schéma classique, cette étape est d'autant plus longue que le changement adaptatif est important et sera fonction de la variabilité génétique disponible. Durant cette période, l'effort d'introduction, c'est-à-dire la quantité d'individus introduits (qui est fonction du nombre d'évènements d'introduction et du nombre d'individus de chaque évènement) est critique, par l'apport de variabilité génétique. En parallèle, les allèles délétères peuvent être purgés. Cette phase était auparavant interprétée comme le temps nécessaire pour atteindre une taille de population suffisante (Ellstrand & Schierenbeck 2000) car l'accroissement de la population suit une courbe exponentielle.

Enfin, les populations introduites voient leur nouvelle aire de répartition s'agrandir, c'est l'invasion au sens classique. Les contraintes génétiques qui ont influencé l'étape précédente sont toujours présentes car en s'éloignant de son point d'introduction, la population invasive peut rencontrer de nouvelles contraintes nécessitant de l'adaptation. La vitesse d'expansion de l'aire ainsi que les limites finales de celle-ci sont fonction de la capacité des individus à s'adapter dans ces nouveaux environnements (Sakai *et al.* 2001).

Ce schéma classique en différentes étapes (Introduction, Adaptation, Invasion) peut être aujourd'hui grandement revisité. Plusieurs scénarios peuvent mener à une invasion mais il y a toujours une étape d'adaptation, ce qui varie d'un scénario à l'autre est le lieu où se déroule l'étape d'adaptation. Trois possibilités sont donc envisagées dans la littérature :

- Le scénario en deux étapes, ou adaptation préalable (« two step » dans la littérature) : les changements évolutifs clés ont lieu dans l'aire d'origine des populations, avant l'introduction. Voir Bossdorf *et al.* (2008) ; Hufbauer *et al.* (2012) ; Rey *et al.* (2012)
- Les changements ont lieu après le déplacement, c'est une réponse aux nouvelles conditions biotiques et abiotiques (Sakai *et al.* 2001), ce qui nécessite un changement adaptatif pour chaque aire d'introduction. C'est le scénario classique, qui est le moins parcimonieux d'un point de vue adaptatif puisqu'il nécessite plusieurs changements adaptatifs. Voir l'exemple du copépode *Eurytemora affinis*, (Lee 1999).
- Le scénario tête de pont, où une première population devient invasive dans une aire géographique, puis toutes les autres aires géographiques envahies le sont à partir de celle-ci.



**FIGURE 1** – Schéma de la décomposition classique de l’invasion biologique en différentes étapes. Les facteurs favorisant l’implantation des espèces non indigènes sont dans le cadre vert à gauche, les facteurs s’y opposant sont à droite. Adapté de Sakai *et al.* (2001) ; Rius & Darling (2014).

Cela ne nécessite qu’un seul changement évolutif (Lombaert *et al.* 2010 ; Keller *et al.* 2012).

- Un scénario proposé plus récemment, appelé MISS (Multiple Introduction from Single Source) qui, comme son nom l’indique, implique l’invasion d’une aire géographique depuis une même source par vagues successives (Benazzo *et al.* 2015).

## Qu’entend on par adaptation ?

Pour Dobzhansky (1968) l’adaptation est le processus évolutif par lequel la capacité d’un organisme à vivre et se reproduire dans un habitat donné est augmentée. La définition que nous retiendrons est qu’une adaptation est la modification d’un caractère anatomique, d’un processus physiologique ou d’un trait comportemental dans une population d’individus sous l’effet de la sélection naturelle, le nouvel état de ce caractère améliorant la valeur sélective des individus qui en sont porteurs.

En faisant l’hypothèse que l’environnement d’introduction est sub-optimal pour les individus, semblable à une vallée dans l’espace topographique de leur valeur sélective, les populations non adaptées ne pourront persister dans ce nouvel environnement que sous certaines conditions. Il faut d’abord distinguer l’adaptation physiologique de l’adaptation génétique (Dobzhansky

1968). En fait, la première se réfère à la plasticité phénotypique, tous les génotypes ont une norme de réaction plus ou moins grande : ce sont les phénotypes que le génotype peut produire dans toute une gamme d'environnement. Ainsi, tous les traits phénotypiques d'un individu sont le fruit de cette plasticité (et de sa trajectoire individuelle), que ces traits soient avantageux ou non, morphologiques, physiologiques ou encore comportementaux. La valeur sélective d'un génotype est donc une fonction de sa norme de réaction et de la gamme des environnements dans lesquels il évolue. Un génotype qui a la capacité de réagir à tous les environnements qu'il rencontre possédera ainsi un fort potentiel adaptatif. Il est certain que la plasticité phénotypique joue un grand rôle dans les processus invasifs. Certaines plantes monoclonales ont ainsi une énorme plasticité et peuvent être extrêmement invasives, conduisant au concept de super-génotype lors de l'étude de *Pennisetum setaceum* (Le Roux *et al.* 2007).

L'adaptabilité génétique par contre permet de changer la norme de réaction possible à un environnement donné et ainsi améliorer la valeur sélective. Un tenant de la biologie évolutive est que le changement évolutif en réponse à la sélection naturelle est proportionnel à la quantité de variance génétique additive présente (Fisher, 1930). C'est-à-dire que pour pouvoir s'adapter, une population doit avoir la quantité de ressources génétiques la plus élevée possible. La quantité d'allèles disponibles est donc primordiale pour l'adaptation.

Durant la colonisation, l'effet de fondation peut réduire la diversité génétique dans les populations nouvellement installées, d'autant plus que le nombre d'individus introduits est faible (ce qu'on appelle l'effort d'introduction : le produit du nombre d'individus à chaque événement d'introduction par la fréquence de ceux-ci (Dlugosch & Parker 2008). Historiquement, on pensait que les populations invasives, *a fortiori* très adaptées à leur nouvel environnement, étaient génétiquement pauvres consécutivement à l'effet de fondation, ce qui avait amené à parler de « paradoxe génétique ». De nombreux travaux ont montré que ce n'était pas le cas, observant une richesse allélique parfois plus grande dans l'aire invasive (Kolbe *et al.* 2004 ; Gillis *et al.* 2009 ; Whitfield *et al.* 2006 ; Ellstrand & Schierenbeck 2000). On sait aujourd'hui que l'admixture contribue à ce phénomène et plus généralement contribue à la réussite de l'expansion géographique (Kolbe *et al.* 2004 ; Whitfield *et al.* 2006). L'admixture (On pourra parler aussi tout simplement d'hybridation, même si ce dernier terme est souvent associé à des croisement interspécifiques ou de rang supérieur) a lieu lorsque des populations précédemment isolées se retrouvent en contact et se croisent, recombinant leurs génomes, comme lors de l'invasion du silène *Silene vulgaris* en Amérique du Nord (Keller & Taylor 2010). De façon plus large on parlera d'hybridation dans le cas d'individus provenant de populations différenciables sur la base de caractères héréditaires. Petit *et al.* (2003) ont montré que ce même phénomène était à l'origine de la reconquête des habitats après la dernière glaciation comme dans le cas du silène (Taylor & Keller 2007) ou du tremble (De Carvalho *et al.* 2010). L'adaptation des populations sera fonction de la variabilité génétique disponible, d'où l'idée que les introductions multiples d'individus peuvent faciliter cette phase grâce à l'accumulation de matériel génétique induit par l'hybridation de différentes populations, aux patrimoines génétiques variés.



### L'admixture, facteur clé dans les processus adaptatifs

La plupart des nouvelles combinaisons créées lors de l'hybridation seront non adaptées à la plupart des environnements, seule une infime partie de celles-ci le sera mieux que ses parents et pourra potentiellement devenir invasive.

En mélangeant du matériel génétique de différentes sources, l'admixture va augmenter la variabilité génétique et créer des occasions de combinaisons alléliques inédites, un fait qui a été observé et est reconnu comme étant un mécanisme majeur de la réussite des invasions biologiques (Kolbe *et al.* 2004 ; Gillis *et al.* 2009 ; Whitfield *et al.* 2006 ; Ellstrand & Schierenbeck 2000). Ainsi les papillons du genre *Heliconius* et plus précisément *H. melpomene*, *H. timarata* et *H. elevatus* ont échangé des gènes dans deux régions du génome fortement impliquées dans le mimétisme. L'hybridation joue donc un rôle important dans la radiation adaptative de ce groupe (Dasmahapatra *et al.* 2012). Un autre cas documenté d'échange de matériel génétique est celui des allèles de résistance aux insecticides chez le moustique. Chez deux espèces de moustiques *Anopheles gambiae* and *A. coluzzii*, un bref épisode d'hybridation au Mali a permis à ce dernier d'hériter d'un morceau de chromosome qui comprend des allèles de résistance aux organophosphorés et aux pyréthrinoides (Norris *et al.* 2015). De plus, l'introgression a coïncidé avec une campagne de distribution de moustiquaires traitées aux insecticides. L'hybridation est ainsi un outil d'acquisition de nouveautés très efficace car des gènes entiers peuvent être échangés à une fréquence qui peut-être bien supérieure à des taux de mutation spontanée (Mallet *et al.* 2015).

Un exemple qui montre l'importance de l'hybridation dans l'expansion géographique est celui de l'adaptation à l'hypoxie liée à la haute altitude chez les populations humaines du plateau tibétain. Cette adaptation serait possible grâce à un allèle du gène *EPAS1* issu de l'introgression de matériel génétique de populations apparentés aux Denisoviens<sup>1</sup> (Huerta-Sánchez *et al.* 2014).

Il existerait une forte corrélation entre l'effort d'introduction et le succès d'invasion (Lockwood *et al.* 2005). Ainsi on aurait bien une accumulation d'allèles, chaque individu introduit apportant sa contribution, augmentant d'autant le potentiel adaptatif des populations introduites. Même si cette hypothèse paraît cohérente, il reste très difficile de mettre en évidence des effets directs de l'hybridation sur la valeur sélective des individus de manière empirique. En effet, mesurer les conséquences de l'admixture sur la valeur sélective est difficile car elle est le résultat de complexes interactions historiques, génétiques et écologiques. Il est compliqué de séparer les effets démographiques dus aux multiples introductions elles-mêmes de ceux dus aux conséquences des effets de l'augmentation de la variance génétique (Kolbe *et al.* 2004).

---

1. L'homme de Denisova est une espèce d'hominidé éteinte qui aurait vécu entre 1 million d'années et 40 000 ans. Il a été identifié grâce à un doigt, un orteil et deux dents dans la grotte de Denisova, dans les montagnes de l'Altai au sud de la Sibérie.

### Augmentation du potentiel adaptatif

En 1966, Lewontin & Birch (1966) montrent en laboratoire que l'hybridation entre deux espèces de mouches *Bactrocera tryoni* et *B. neohumeralis* (hybridation ayant également lieu en milieu naturel) permet une évolution adaptative relativement rapide vers une tolérance pour des températures stressantes. L'hybridation *per se* est pourtant désavantageuse, car durant la première année de l'expérience, les hybrides affichent des performances inférieures (se pose alors la question d'une possible contre sélection en milieu naturel). Ce n'est qu'au bout d'une année supplémentaire que les performances (fécondité, longévité, viabilité larvaire et taux d'accroissement) deviennent supérieures pour les hybrides à une température normalement stressante pour les non hybrides, et identiques à des températures inférieures. Ce résultat est très intéressant car il illustre bien l'avantage de l'hybridation par augmentation de la variance génétique sans effet hétérosis<sup>1</sup>, et que les performances sont supérieures seulement dans un environnement donné, qui n'est pas celui des espèces parentales, en tout cas dans les conditions de l'expérience. Les auteurs font même l'hypothèse que cette hybridation a permis la colonisation de nouveaux espaces par *B. tryoni*.

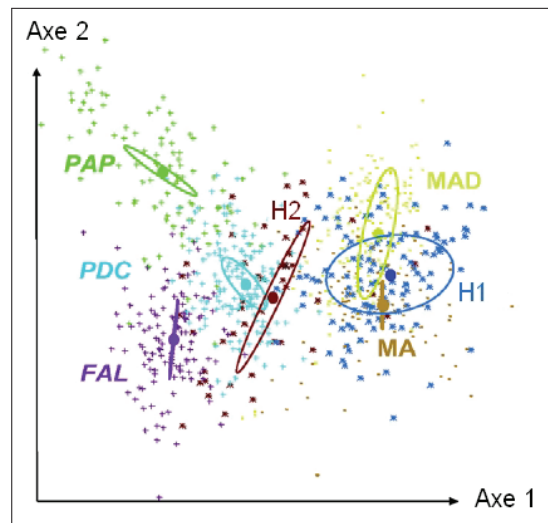
En conditions naturelles, plusieurs études ont observé les effets bénéfiques de l'admixture. Chez la moule *Mytella charruana*, (Gillis *et al.* 2009) les populations d'origines montrent un grand nombre d'haplotypes proches, alors que les populations invasives ont une distribution multimodale de la fréquence des haplotypes, patron reflétant l'admixture (Slatkin 2008), ainsi qu'une plus grande diversité génétique. Ce résultat s'explique par l'admixture entre deux populations différentes ayant subies un goulot d'étranglement. Les auteurs vont plus loin en prédisant l'invasion par cette moule hybride : les ballasts des bateaux amènent sans cesse de nouveaux individus, la population accumule donc du matériel génétique nécessaire à son adaptation (Sakai *et al.* 2001 ; Holt *et al.* 2003) et grâce à la large tolérance à la salinité de cette espèce (plasticité phénotypique), elle devrait finir par se répandre largement dans l'environnement. Ce rôle de la plasticité interagissant avec l'admixture est appuyé par les résultats sur la baldingère (plante de la famille des poacés) où ces deux phénomènes joueraient un rôle majeur dans la capacité d'invasion (Lavergne & Molofsky 2007). Cette dernière étude ainsi que d'autres sur les silènes (Keller & Taylor 2010) et sur l'écrevisse (Perry *et al.* 2001), montrent clairement le rôle de l'admixture dans l'invasion grâce à des mesures de valeur sélective des différentes populations ; ce qui n'était pas le cas de l'étude sur la moule qui ne montrait que la présence d'admixture via des marqueurs neutres. Une autre approche a permis de montrer l'impact de l'admixture sur le potentiel adaptatif. Facon *et al.* (2008) ont montré chez *Melanoides tuberculata* (mollusque gastéropode) que l'admixture de différentes populations avait augmenté la diversité intra-population pour plusieurs traits phénotypiques par rapport à leurs parents et que les individus pouvaient avoir des valeurs phénotypiques plus extrêmes que leurs parents. Pour cela, ils ont réalisé une analyse multivariée de cinq traits d'histoire de vie des hybrides naturels ainsi que des popula-

1. L'effet hétérosis désigne l'augmentation des capacités et ou de la vigueur d'un hybride par rapport aux lignées parentales.

## Introduction

tions parentales. Les résultats montrent que les hybrides sont différents des autres populations sur le plan factoriel de l'analyse, ce qui est la preuve d'un potentiel différent par rapport à celui des parents (figure 2).

**FIGURE 2** – Analyse multivariée de 5 traits d'histoire de vie de différentes populations de mollusque *Melanoides tuberculata*. Les hybrides naturels sont les populations H1 et H2. Leur populations parentales sont respectivement FAL x PAP et FAL x PDC.



Ces données montrent l'importance de l'histoire des populations. En effet l'admixture a plus d'impact si elle mélange des patrimoines génétiques fortement différenciés, notamment dans le cas de populations fortement structurées génétiquement dans l'aire d'origine. Ses conséquences évolutives sont différentes si les populations d'origine sont séparées depuis longtemps ou non et si elles ont subi un goulot d'étranglement qui réduirait la variabilité génétique disponible. Dans le cas de populations structurées, l'invasion depuis une seule aire sera non représentative de la diversité génétique et l'admixture aura un fort rôle potentiel dans l'adaptation. Dans le cas de populations non structurées, l'invasion sera représentative de la diversité totale des populations et l'admixture aura potentiellement moins d'impact sur le potentiel adaptatif (Taylor & Keller 2007).

## Complexe d'espèces et invasions

### Comment définit-on une espèce ?

L'espèce est le niveau taxonomique de base dans notre systématique, il en existe des dizaines de définitions (De Queiroz 2007). Nous retiendrons la définition de Mayr, dite définition « biologique » : *Species are groups of actually or potentially interbreeding natural populations, which are isolated from other such groups.* Ainsi, comme nous l'avons brièvement évoqué, l'espèce est la plus grande unité de populations au sein de laquelle le flux génétique est possible et les individus d'une même espèce sont donc génétiquement isolés d'autres ensembles équivalents du point de vue reproductif. Dans tous les cas cela reste un concept car la réalité biologique est celle des populations, nous humains, ne faisons qu'en définir des limites en tant qu'observateur. Cela nous permet de classer les choses, pouvoir les ranger dans des boîtes bien identifiées à un temps donné, la classification binomiale en est l'archétype.

Ainsi, pour certaines espèces, l'isolement reproducteur apparaît de manière évidente (on voit mal une vache et un cheval se reproduire<sup>1</sup>) mais dans le cas d'espèces étroitement apparentées, les barrières sont beaucoup moins claires. Le concept biologique de l'espèce possède donc certaines limites. L'isolement reproducteur ne peut pas être déterminé dans le cas des organismes asexués (par exemple, les bactéries, les organismes parthénogénétiques, apomictiques ou encore certaines plantes et leur reproduction végétative). Il y a aussi beaucoup d'espèces qui peuvent s'hybrider dans la nature mais sans que les taxonomistes les rassemblent, particulièrement chez les plantes (Arnold 1997, chap. 3). Nous citerons un dernier problème qui peut-être soulevé, celui de la transitivité : le concept biologique de l'espèce suppose que si A est interféconde avec B et que B l'est avec C, alors A et C sont interfécondes. Comme cette chaîne peut être beaucoup plus longue, les deux extrêmes ne sont plus forcément interfécondes, c'est le cas classique des espèces en anneaux. Le pouillot verdâtre *Phylloscopus trochiloides* en est l'exemple type : deux populations de l'ouest (*P. t. viridanus*) et de l'est (*P. t. plumbeitarsus*) de la Sibérie coexistent en Sibérie centrale mais ne sont pas interfécondes. Pourtant, ces deux formes sont connectées par une chaîne de populations interfécondes, dont les traits phénotypiques changent graduellement, qui passe par le sud du haut plateau tibétain, non habité par ces oiseaux. Un exemple assez similaire est celui des passereaux du genre *Alophoixus* qui forment un anneau autour des plaines thaïlandaises (Fuchs *et al.* 2015)

### Complexe d'espèces et invasion

Dans la littérature le concept d'espèces cryptiques est souvent synonyme du concept d'espèces sœurs (Sáez & Lozano 2005), tandis que pour certains chercheurs ce dernier concept implique une relation de parenté directe (Knowlton 1986), ce qui est en fait assez logique (Bickford *et al.* 2007). Pour compliquer la tâche, d'autres utilisent ce terme pour désigner des espèces capables de camouflage (Claridge *et al.* 2004). Une définition classique est celle de Bickford *et al.* (2007) : plusieurs espèces sont cryptiques si elles sont ou ont été classées par erreur comme une seule espèce, parce qu'elles étaient au moins en partie morphologiquement indissociables. D'autres critères peuvent être nécessaires selon les auteurs, comme le fait que la divergence soit récente, que les espèces soient dissociables seulement par les marqueurs moléculaires, qu'elles soient en sympatrie ou encore isolées génétiquement (Bickford *et al.* 2007). Ce dernier point est important car des espèces peuvent être morphologiquement indissociables mais complètement isolées génétiquement, il « suffit » donc de trouver le marqueur moléculaire adéquat pour lever toute ambiguïté. Ce qui est beaucoup plus intéressant est le cas d'espèces dont les barrières reproductives sont peu connues soit par ignorance de la science, soit parce que l'isolement reproducteur n'est pas complet en populations naturelles, permettant des événements d'hybridation. Nous allons donc utiliser une définition qui implique le fait que les barrières reproductives du clade (l'ensemble des espèces désignées comme cryptiques) sont mal connues. En effet, ame-

---

1. Le jumart, mâle ou femelle, était, selon une croyance répandue au XVIII<sup>ème</sup> siècle, un hybride issu de l'âne ou de l'étalon avec une vache ou d'un taureau avec une ânesse ou une jument.

## Introduction

ner une population au rang d'espèce (c'est à dire fractionner une espèce en plusieurs espèces) repose sur la mise en évidence de barrières reproductives, les différences morphologiques ne sont qu'un sous produit du processus de la divergence génétique résultant de l'isolement reproducteur.

La difficulté à distinguer des espèces n'est pas restreinte aux petits organismes comme on se serait tenté de le croire, par exemple le crocodile du nil est formé de deux taxons qui diffèrent notamment par leur caryotype et dont l'une des espèces s'avèrent être phylogénétiquement plus proche des crocodiles américains (Hekkala *et al.* 2011). Il est probable que nous ayons du mal à discriminer ces espèces car nous sommes plus sensibles aux informations visuelles, et qu'ainsi les caractères chimiques et auditifs sont peu utilisés dans notre système de classification. Le manque de caractères permettant de reconnaître ces espèces n'implique pas forcément une spéciation récente (le temps aurait manqué pour que ces caractères évoluent). Bien que cela puisse être vrai, certains exemples montrent une spéciation très ancienne comme chez les poissons du genre *Albula* qui ont divergé il y a environ 4 à 20 millions d'années (Colborn *et al.* 2001). Par extension, un ensemble d'espèces cryptiques sera appelé complexe d'espèces, sous entendues cryptiques. On trouve aussi abondamment dans la littérature les expressions groupes d'espèces ou super-espèce.

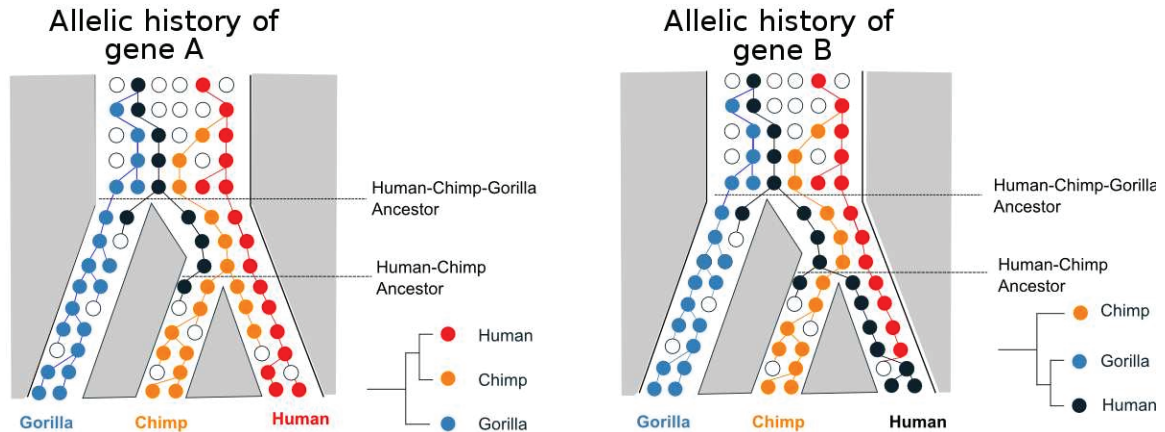
Ce concept de complexe d'espèces est indissociable de la définition de l'espèce car l'expression « barrière reproductrice » pour définir une espèce est peut-être mal choisie, il s'agit plutôt d'un continuum dans bien des cas, de l'interfécondité totale à la séparation génétique complète en passant par tous les degrés d'hybridation possible. Cette possibilité d'hybridation brouille les signaux génétiques et phylogénétiques nécessaires à l'identification des espèces : tous les locus d'un génome ne racontent pas la même histoire. Les deux principales causes de discordances sont le tri incomplet des lignées<sup>1</sup> ainsi que l'introgession. Le tri incomplet des lignées est lié à la persistance du polymorphisme entre des évènements de spéciation, pour une même histoire des espèces et lorsque les spéciations sont temporellement proches, les histoires des différentes populations d'allèles peuvent être différentes. Ainsi la généalogie d'un gène ou d'une région du génome ne reflètent pas les vraies relations phylogénétiques des espèces. Par exemple, 15 % de nos gènes sont plus apparentés à leurs homologues chez le gorille plutôt qu'à ceux du chimpanzé, notre plus proche parent (voir figure 3). Un autre exemple est celui des considérables discordances observées à la base de l'arbre des oiseaux modernes, probablement provoquées par du tri incomplet de lignées lors de leur radiation (Suh *et al.* 2015), mais sûrement aussi par de nombreux évènements d'hybridation. L'introgession a lieu lorsque des backcross ont lieu entre des hybrides appartenant à deux espèces différentes, transférant du matériel d'une espèce à l'autre. Par exemple, le genre *Heliconius* (papillon) est constitué d'environ 15 espèces. La plupart d'entre elles sont bien délimitées, se trouvent en sympatrie dans certaines régions et peuvent être interfertiles. De rares hybrides et des individus introgressés ont été décrits dans ce groupe, suggérant un flux continu de gène dans le complexe (Mallet *et al.* 2007). Ainsi, lo-

---

1. Dans la littérature, ILS : *Incomplete Lineage Sorting*



calement des individus d'une population de *H. melpomene* peuvent être plus apparentés à des individus sympatriques de l'espèce *H. cydno* que de ses conspécifiques sur plus de 40 % du génome (Martin *et al.* 2013).



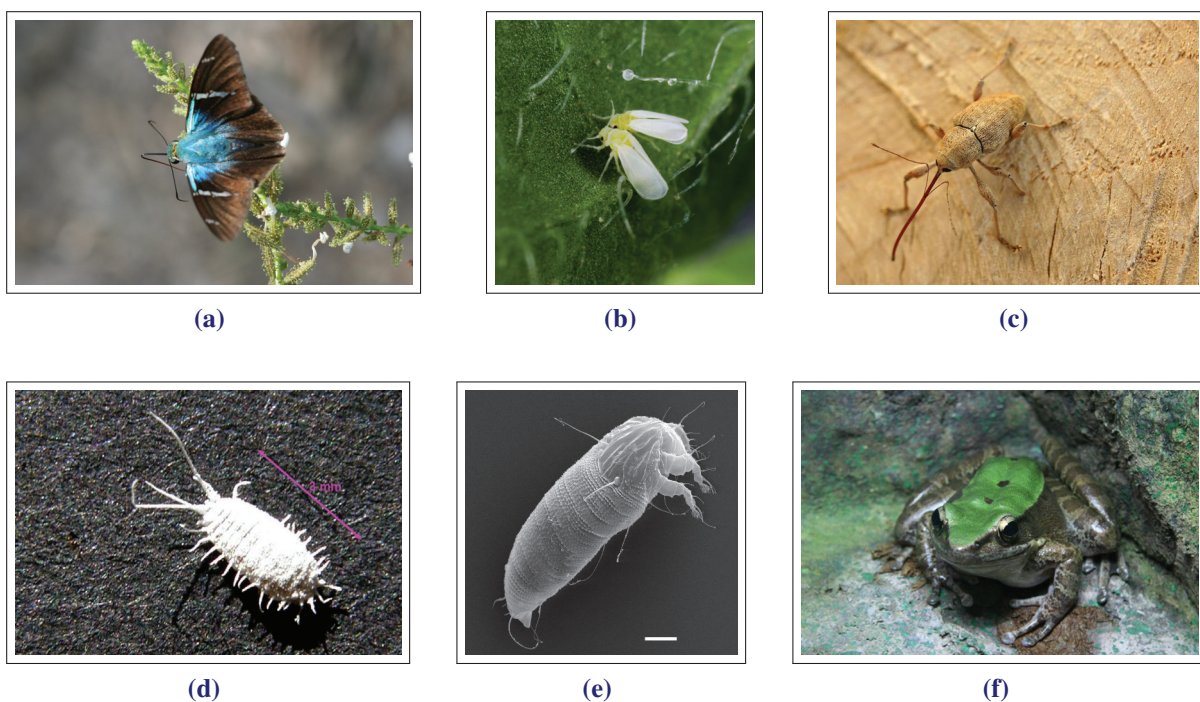
**FIGURE 3** – Schéma explicatif du tri incomplet des lignées. Du polymorphisme est présent chez l'ancêtre commun aux trois espèces de singe et dans le cas du gène A (à gauche), l'histoire des allèles est identique à celle des espèces, le chimpanzé est donc correctement regroupé avec l'être humain. Dans le cas du gène B (à droite), le polymorphisme persiste et l'allèle noir est présent au final chez l'être humain, alors que cet allèle est plus proche parent de celui retrouvé chez le gorille. L'humain est incorrectement rapproché du gorille.

La littérature est riche en exemples d'espèces cryptiques et la description de nouveaux cas est en augmentation constante, probablement grâce à l'apparition des techniques de séquençage (séquençage Sanger, puis séquençage haut-débit), même si ces nouveaux outils, bien que très puissants, amènent de nombreuses questions (Bickford *et al.* 2007). En effet, il peut être difficile en pratique de déterminer si des populations qui n'ont pas échangées de gènes dans un passé récent peuvent effectivement le faire aujourd'hui. Les études de génétiques de populations pour évaluer les flux de gènes, qui se font sur des marqueurs supposés neutres (en tout cas c'est qui est souhaité), permettent de détecter l'absence d'hybridation entre populations. Par contre cela ne dit pas si ces populations pourraient potentiellement le faire si l'occasion s'en présentait car par exemple les barrières géographiques les en empêchent jusqu'à présent. Pour vérifier la capacité à s'hybrider, cela nécessite des études comportementales en laboratoire, ce qui n'est pas forcément facile. Nous verrons par exemple dans le cas de *Bemisia tabaci* que certaines espèces de ravageurs sont soumises à réglementation sur leur importation et leur détention. Pour d'autres il peut-être difficile en pratique de les maintenir en laboratoire (taille, mode de vie...).

Par contre, beaucoup d'études ont démontré la présence d'espèces cryptiques isolées génétiquement alors qu'elles sont en sympatrie, suggérant indirectement que ces entités n'échangent pas de gènes (Hebert *et al.* 2004 ; Stuart *et al.* 2006 ; Mouton *et al.* 2015). La plupart des grands clades du vivant sont concernés par ce phénomène d'espèces cryptiques, mais avec

## Introduction

un fort biais en faveur des animaux (les plantes sont très peu représentées). Une méta-analyse de la littérature (771 931 articles publiés entre 1978 et 2006) a permis de mettre en évidence 2 207 groupes d'espèces cryptiques (Pfenninger & Schwenk 2007), qui sont réparties uniformément parmi les nombreux taxons de métazoaires testés : mammifères, amphibiens, reptiles, oiseaux, crustacés, nématodes, annélides, éponges, céphalopodes, insectes, échinodermes, chordés, gastéropodes, bivalves, arachnides... La région biogéographique<sup>1</sup> ne semble pas non plus influencer la fréquences des complexes d'espèces. Parmi les milliers d'exemple d'espèces cryptiques (voir les illustrations en figure 4) on peut citer des papillons (Hebert *et al.* 2004), l'acarien *Aceria tosichella* (Skoracka *et al.* 2012), les cochenilles des genres *Pseudococcus* et *Planococcus* (Malausa *et al.* 2011), et bien sûr le complexe *Bemisia tabaci*. Ce phénomène n'est pas l'apanage des arthropodes puisqu'on a découvert que la chauve-souris *Rhinolophus rouxii* est en réalité composée de deux espèces génétiquement différenciées (8 % de divergence) et également distinguables par leur cris (Chattopadhyay *et al.* 2012), ou encore les grenouilles des genres *Odorrana* et *Rana* (Stuart *et al.* 2006) qui sont composées de plusieurs espèces.



**FIGURE 4** – Illustrations de différentes espèces cryptiques. (a) : *Astraptes fulgerator*, (b) : *Bemisia tabaci*, (c) : *Curculio elephas*, (d) *Pseudococcus affinis*, (e) *Aceria tosichella*, (f) *Odorrana livida*.

Dans la biologie des invasions, la notion de complexe d'espèces et de barrières reproductives prend toute son importance car — en caricaturant — dans le cas de deux espèces distinctes en contact dans une région envahie, il y aura potentiellement compétition entre les deux, alors que dans le cas de deux entités inter-fécondes il y aura admixture. Donc dans un complexe d'espèces, formé de multiples entités dont les limites reproductives sont peu connues, on peut

1. Une zone biogéographique désigne en biogéographie une zone géographique climatiquement et écologiquement relativement homogène du point de vue des formations végétales et des températures.

s'attendre à l'émergence de nombreuses combinaisons de matériels génétiques différents selon les populations qui s'hybrideront.

### **Interaction avec le compartiment cytoplasmique**

Pour le moment nous avons implicitement considéré le compartiment nucléaire seulement : les gènes soumis à recombinaison lors de la reproduction sexuée. Mais une cellule eucaryote a deux compartiments porteurs d'ADN qui sont le noyau avec les chromosomes et le cytoplasme avec les organites telles que les mitochondries et les bactéries symbiotiques. C'est le compartiment nucléaire qui reste le plus étudié dans la biologie de l'invasion, alors qu'une cellule eucaryote comprend beaucoup d'autres choses, et un organisme pluricellulaire encore plus, comme toute la flore digestive !

Lors de la reproduction sexuée le compartiment cytoplasmique maternelle se retrouve en contact des gènes provenant de la lignée paternelle. Dans une population panmictique les gènes de la lignée paternelle ne sont *a priori* pas différent de ceux de la lignée maternelle. Dans le cas de l'hybridation de plusieurs populations invasives, ou d'une population invasive avec une autochtone (toutes les combinaisons sont possibles), le compartiment cytoplasmique se retrouve en contact d'autres gènes ou allèles provenant de la lignée paternelle, ce qui peut créer de nouvelles combinaisons nucléo-cytoplasmiques. C'est ce phénomène qui est à l'origine de l'introgession mitochondriale, c'est-à-dire la sélection d'une combinaison nouvelle entre mitochondries et un fond génétique nucléaire. Ce phénomène est extrêmement répandu dans la nature et est détecté lorsque les phylogénies mitochondriales et nucléaires sont discordantes (pour une revue voir Toews & Brelsford 2012). Au même titre que l'admixture de matériel génétique nucléaire, cela va créer de nouvelles combinaisons nucléo-cytoplasmiques.

Pourtant, ces deux compartiments, cytoplasmiques et nucléaires, sont rarement appréhendés conjointement dans le domaine de la biologie de l'invasion. Il y a bien sûr de nombreuses études qui ont montré la nécessité d'utiliser simultanément des marqueurs mitochondriaux et nucléaires dans l'étude des populations invasives, puisque ceux-ci ne racontent pas forcément la même histoire évolutive. On observe alors des arbres phylogénétiques discordants entre les deux marqueurs (Pons *et al.* 2014 ; Shaw 2002). Néanmoins ces approches se restreignent souvent à la mitochondrie, et n'utilisent souvent qu'un seul gène comme témoin de l'histoire maternelle. Ce qui n'a pas vraiment été étudié pour le moment est le rôle des bactéries symbiotiques dans l'invasion. Ces bactéries sont intracellulaires, transmises verticalement de mère à descendants, et jouent de multiples rôles (manipulation de la reproduction, spécialisation à une plante hôte, résistances aux insecticides...) et sont capables d'envahir très rapidement les populations d'insectes et ainsi modifient considérablement leur biologie et leur écologie. En plus de ces effets phénotypiques bien définis apportés par les bactéries, un mécanisme particulièrement intéressant a lieu lorsque différentes populations provenant de différentes sources, avec des communautés symbiotiques différentes vont entrer en contact. L'hybridation va créer de nouvelles com-



binaisons de traits, comme pour l'admixture nucléaire classique : une population apporte l'allèle  $a$ , une deuxième l'allèle  $b$ , l'hybride résultante présentera la combinaison  $ab$ , jusque là inconnue, avec toutes les conséquences génétiques possibles. Mais que les allèles soient portés par le génome nucléaire ou symbiotique ne change pas fondamentalement les choses, l'admixture est donc un phénomène qui concerne autant les allèles nucléaires que cytoplasmiques. Mais en plus, lorsque le symbiote apporte à un fond nucléaire donné un trait totalement nouveau, la population hybride aura un trait phénotypique en plus, le saut est qualitatif.

L'intérêt de l'étude des complexes d'espèces est la multitude de génotypes nucléaires comme nous l'avons dit, mais aussi les différentes bactéries qui pourront être trouvées avec toutes les différences de prévalence d'infection possibles. Pour cela, nous allons définir ce qu'est la symbiose pour essayer de comprendre le rôle qu'elle peut jouer dans la biologie de l'invasion.

## La symbiose, actrice de l'invasion ?

### Définition, brève histoire du concept de symbiose

La symbiose est littéralement le fait de vivre ensemble qu'on pourrait définir comme l'association durable entre deux organismes (ou plus) n'appartenant pas à la même espèce. L'utilisation du terme dans ce contexte <sup>1</sup> remonte à la fin du XIX<sup>e</sup> siècle et est due à Frank (1876) qui parlait alors de *symbiotismus*, terme popularisé sous la forme *symbiosis* par De Bary en 1879.

Le terme symbiose n'implique pas nécessairement une relation bénéfique mutuelle entre les parties, l'étymologie du terme est d'ailleurs très claire puisqu'elle signifie « vivre avec » et non « vivre en harmonie », contrairement à ce que le sens populaire pourrait laisser penser. Le parasitisme (un des deux partenaires tire profit de l'autre qui est donc pénalisé) et le commensalisme (une espèce tire profit de l'autre sans la pénaliser) sont tout autant des symbioses que le mutualisme où les deux partenaires tirent profit de la relation, pouvant même provoquer une inter-dépendance entre les organismes. Ces deux organismes peuvent aussi se retrouver en compétition sans que l'un ou l'autre en tire profit, la symbiose peut donc ainsi nuire aux deux partenaires, mais alors l'association ne sera pas durable.

La symbiose est à l'origine de très nombreuses innovations dans l'histoire de la vie dont l'exemple le plus connu est probablement celui de la cellule eucaryote constituée de deux partenaires procaryotes dont la mitochondrie, une alphaprotéobactérie (Margulis 1970 ; Lang *et al.* 1999). Une deuxième endosymbiose, avec cette fois des cyanobactéries, conduira à l'émergence de la lignée verte (les végétaux).

Le phénomène de la symbiose est très répandu dans la nature : en 1953, Buchner décrivait des centaines d'associations différentes entre des insectes et des microorganismes (Buchner 1953). Les bactéries peuvent permettre de coloniser de nouvelles niches grâce à la spécialisation à de

---

1. Le terme existe dans la langue anglaise depuis 1622 et avait le sens de « vie en communauté » et était propre aux relations humaines

nouveaux milieux nutritionnels (Douglas 2009). Elles peuvent fournir des nutriments essentiels ou sécréter des inhibiteurs de champignons antagonistes (Cardoza *et al.* 2006). Il existe des exemples de bactéries symbiotiques qui jouent un rôle dans la défense contre des pathogènes ou des parasitoïdes (Oliver *et al.* 2003 ; Hedges *et al.* 2008 ; Teixeira *et al.* 2008 ; Kaltenpoth 2009). D'autres peuvent jouer un rôle dans l'hygiène ou la conservation des proies (Currie *et al.* 1999 ; Kaltenpoth *et al.* 2005 ; Kaltenpoth 2009). Un dernier exemple est la thermotolérance apportée par certaines bactéries endosymbiotiques à leur hôte (Dunbar *et al.* 2007).

La symbiose n'est pas l'apanage des insectes et des bactéries puisque l'ensemble des organismes vivants est concerné. Les mollusques peuvent établir des symbioses avec des bactéries (Fiala-Medioni 1988) ; le termite *Mastotermes darwiniensis* héberge un flagellé (*Mixotricha paradoxa*) dans son tube digestif, lui-même associé à plusieurs bactéries. Cet exemple est d'autant plus fascinant qu'il y a ici plusieurs niveaux de symbioses emboîtés, on parle alors de plurisymbiose.

Les organismes en association peuvent être liés de deux façons : soit l'un se situe à l'intérieur de l'autre, on parle alors d'endosymbiose. Dans le cas contraire on parle d'ectosymbiose comme par exemple les lichens ou l'association entre le poisson clown et l'anémone de mer.

Dans le cas qui nous intéresse, l'endosymbiose, le symbiote bénéficie d'une protection contre l'environnement extérieur et l'hôte lui fournit des nutriments ainsi qu'un milieu stable. Ces caractéristiques sont encore plus marquées pour les endocytobiotés que nous verrons un peu plus bas. L'endosymbiose permet d'incorporer dans un individu un organisme entier avec tout son répertoire de gènes et ses voies métaboliques, une manière très rapide d'acquérir de nouveaux traits, en comparaison d'autres phénomènes tels que la mutation ou même le transfert horizontal de gènes.

L'endosymbiose est très répandue, c'est le cas par exemple des termites (Bandi *et al.* 1995) et des ruminants qui hébergent de nombreux microorganismes dans leur tube digestif, jouant un rôle très important dans l'assimilation de nutriments à partir de molécules complexes qu'ils ne pourraient assimiler seuls. L'homme est bien sûr concerné par ce type de symbiose puisqu'il héberge environ 100 000 milliards de bactéries et eucaryotes unicellulaires dans son intestin<sup>1</sup> (c'est environ 10 fois plus que ses propres cellules) qui participent à la fermentation de substrats, à la synthèse de vitamines et jouent le rôle de barrière contre les pathogènes (Eberl 2015). Dans le tube digestif on peut retrouver une association plus inattendue : l'organe lumineux de certains poissons comme le perciforme *Siphamia versicolor* est le résultat d'une endosymbiose avec la bactérie *Photobacterium leiognathi* (Haygood & Distel 1993). Notons tout de même que le contenu du tube digestif reste à l'extérieur de l'organisme puisque le tube digestif s'ouvre sur l'environnement à ses deux extrémités. De ce point de vue, les endosymbiotes du tube digestif ne sont alors que de vulgaires symbiotes de surface, au même titre que les bactéries que nous retrouvons sur notre peau.

---

1. L'intestin n'est pas le seul organe colonisé chez l'homme, on retrouve des bactéries dans le système respiratoire, l'estomac et l'appareil uro-génital.

## Introduction

Ici nous allons nous intéresser plus précisément aux endocytobioses qui sont des endosymbioses que l'on pourrait considérer comme plus poussées car le symbiote vit à l'intérieur des cellules de son hôte, on parle aussi de symbiose intra-cellulaire. Le symbiote est alors appelé endocytobiotte, mais le terme endosymbiote est très souvent employé pour le désigner dans la littérature scientifique, ce que nous allons faire également dans ce manuscrit. Il existe de très nombreux exemples de telles associations : virus-drosophile (Thomas-Orillard 1996), bactérie-mollusque (Fiala-Medioni 1988), bactérie-cloporte (Rigaud *et al.* 1991), bactérie-insecte (nous allons voir de nombreux exemples par la suite) ou encore levure-coléoptère (Davis 2015). Bien que ces endosymbioses soient obligatoires du point de vue de la bactérie (la bactérie ne peut vivre sans son hôte), elles ne le sont pas forcément du point de vue de l'hôte. C'est pourquoi elles sont classiquement divisées en trois catégories (Moran *et al.* 2008) : les symbioses obligatoires, les symbioses facultatives et enfin les symbioses qui impliquent la manipulation de la reproduction.

## Transmission de ces endosymbiotes

Une propriété de tous ces symbiotes est qu'ils se transmettent essentiellement de façon verticale, le plus souvent maternellement, *via* le cytoplasme des ovocytes. Les symbiotes secondaires peuvent se transmettre horizontalement, mais c'est un phénomène dont la fréquence est beaucoup plus rare et doit être envisagé à l'échelle des temps évolutifs.

La transmission se fait lors de la reproduction sexuée, soit parce que le symbiote est contenu dans les gamètes (de manière analogue aux mitochondries et plastes que nous avons cités plus haut), soit par une colonisation post-fécondation. L'hérédité biparentale est possible (Xu & Wang 2015), quoique rare car elle peut engendrer des conflits entre symbiotes non apparentés, ou des conflits nucléo-cytoplasmiques. C'est probablement pour cette raison que la transmission monoparentale a été plus souvent sélectionnée notamment dans le cas de la cellule eucaryote et ses organelles (Birky 1995).

## Les symbiotes obligatoires

La symbiose obligatoire, ou primaire, résulte d'une longue histoire co-évolutive entre les deux partenaires, une association facultative qui s'est muée en dépendance pour l'hôte, l'hôte ne peut alors survivre sans la bactérie. Lorsque le symbiote a été acquis ancestralement, il devait jouer un rôle majeur dans l'écologie de son hôte notamment *via* l'exploitation de nouvelles niches (Ferrari & Vavre 2011). On peut considérer la symbiose primaire comme une relation mutualiste puisque le symbiote est obligatoire pour la survie ou la reproduction de l'hôte et inversement.

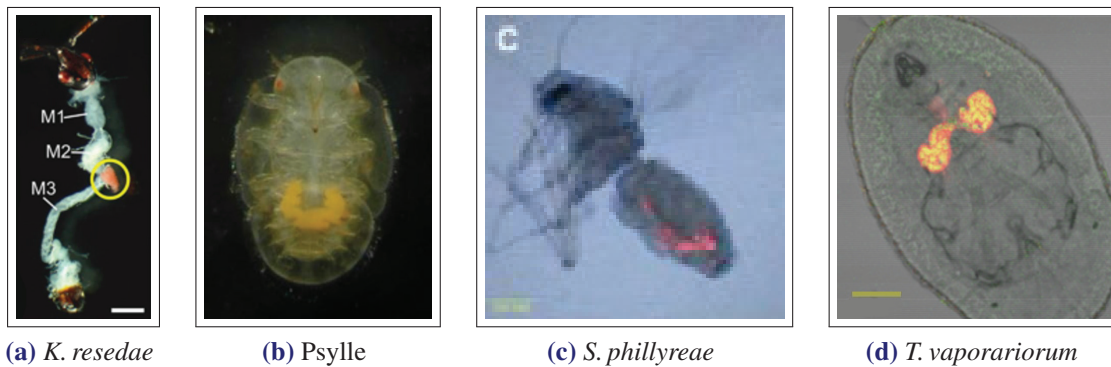
### Localisation tissulaire

Le symbiote primaire est bien souvent logé dans des bactériocytes, des cellules spécialisées de l'hôte. Dans certaines espèces, ces cellules sont regroupées dans une structure appelée bactériome, organe propre à ces symbioses (voir figure 5). Selon l'hôte, les bactériocytes peuvent être des cellules adipeuses, des cellules de la paroi intestinale ou encore des cellules hautement spécialisées qui se développent dès le stade embryonnaire (Braendle *et al.* 2003). Le bactériome peut agir de vecteur pour la transmission verticale. Par exemple, chez le puceron comme chez la plupart des insectes, le symbiote primaire quitte à un moment donné les cellules spécialisés qui l'hébergent et migre vers les cellules germinales *via* l'hémolymphe pour être transmis à la génération suivante (Koga *et al.* 2012). Dans d'autres cas, comme les aleurodes et les blattes, ce sont les bactériocytes qui servent de vecteur de transmission à la génération suivante en colonisant les ovocytes (Costa *et al.* 1996 ; Szklarzewicz & Moskal 2001).

Bien que l'endosymbiote soit bénéfique et nécessaire à son hôte, l'association permanente entre l'hôte et les bactéries implique l'existence de mécanismes de contrôle qui régulent la population de symbiotes pour éviter leur prolifération dans les autres tissus de l'hôte et les dommages qu'ils pourraient ainsi lui causer. Le système immunitaire de l'insecte est donc toujours actif contre les symbiotes. Par exemple, il a été montré que *Blochmannia floridanus*, symbiote obligatoire de la fourmi *Camponotus floridanus*, est toujours reconnu et attaqué par le système immunitaire lorsqu'il est injecté dans l'hémolymphe de l'insecte (Ratzka *et al.* 2011). L'hémolymphe de l'insecte n'est pas le tissu habituel du symbiote, sa détection dans ce tissu n'est donc peut-être pas si inattendue. En revanche, Login *et al.* (2011) ont montré que des peptides antimicrobiens ciblent spécifiquement le symbiote primaire dans les bactériocytes chez les charançons du genre *Sitophilus*, régulant leur croissance et inhibant leur division cellulaire. Chez certains insectes, cas plus rare, le symbiote primaire se situe dans des compartiments spécialisés du tube digestif. Même si cette existence est extracellulaire, la transmission verticale à la génération suivante s'effectue grâce à la contamination de la surface des œufs ou des fèces à proximité de ces derniers (Kaltenpoth 2009).

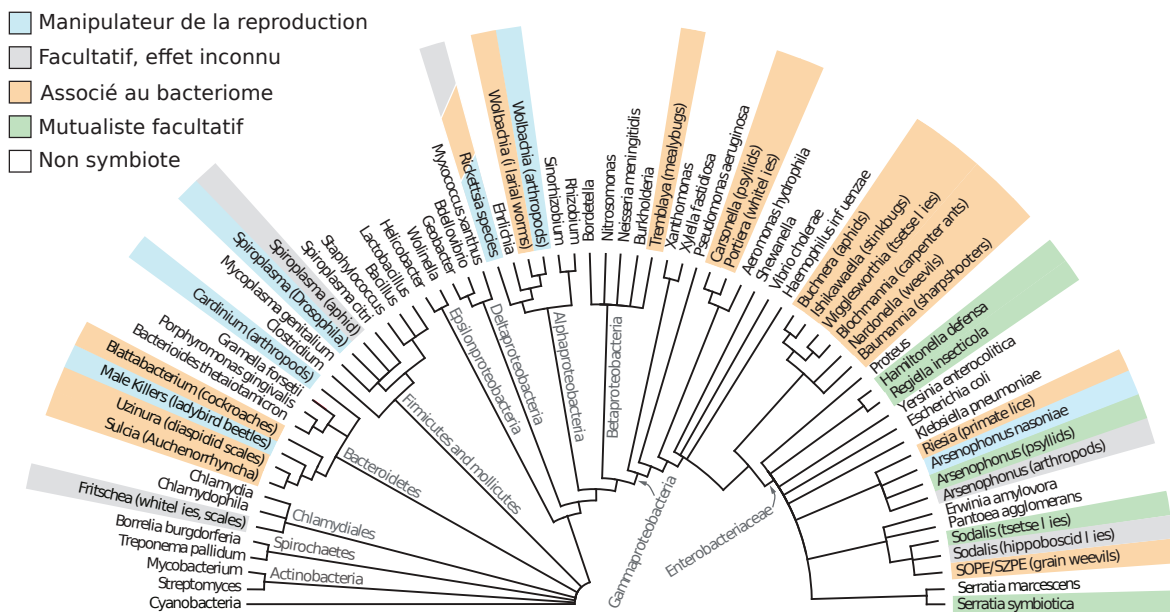
### Diversité et rôles de ces symbiotes

Si on regarde un arbre phylogénétique des symbiotes primaires et de leurs hôtes, on remarque que la symbiose obligatoire est apparue à plusieurs reprises dans différents phylums de bactéries, et plusieurs fois à l'intérieur de ces différents phylums (Moran *et al.* 2008). Bien que les groupes bactériens concernés soient divers, certains groupes sont très représentés comme les Bacteroidetes et les Gammaprotéobactéries, particulièrement la famille des Entérobactéries (Voir figure 6). Ces symbiotes ont évolué indépendamment dans diverses lignées d'un état facultatif à obligatoire pour l'hôte. Par exemple le groupe frère des Entérobactéries (voir la figure 6) infecte des hôtes aussi divers que des fourmis, des pucerons, des mouches tsétsé, etc (Moran *et al.* 2008). Et à l'inverse, les hémiptères ont établi des symbioses obligatoires avec des



**FIGURE 5** – Illustrations de quatre bactériomes. (a) : Système digestif de la punaise *Kleidocercy resedae* [Heteroptera : Lygaeidae] avec les deux bactériomes (cercles jaunes) hébergeant la bactéries *Kleidocercy schneideri* (Küchler *et al.* 2010). (b) : bactériomes de psylle contenant *Carsonella ruddii*. (c) : bactériomes (en rouge) de l’aleurode *Siphoninus phillyreae* contenant *Portiera aleyrodidarum* (Skaljac *et al.* 2013). (d) : bactériomes de l’aleurode *Trialeurodes vaporariorum* au stade pupa, avec la bactérie *Portiera aleyrodidarum* en rouge et *Hamiltonella* en jaune (Skaljac *et al.* 2010).

Bacteroidetes, des Gammaprotéobactéries et des Bétaprotéobactéries (Moran *et al.* 2008).



**FIGURE 6** – Arbre phylogénétique des endosymbiotes retrouvés chez les insectes, tiré de Moran *et al.* (2008)

Les insectes sont un groupe extrêmement divers, peut-être parce que les symbioses obligatoires ont joué un rôle dans l’apparition de taxons, puisque des associations symbiotiques primaires datent de l’origine de ces taxons (Moran *et al.* 2005). Les bactéries endosymbiotiques primaires (ou obligatoires) ont permis d’exploiter de nombreuses niches écologiques car elles apportent des nutriments essentiels tels que des vitamines ou des acides aminés, que leur milieu (bois, sang et sève par exemple) ne leur apporte pas car il est carencé (Moran *et al.* 2008). Par exemple, les insectes hématophages ont des bactéries qui leur apportent de la vitamine B (Akman *et al.*



2002) et les insectes phloémophages (qui se nourrissent de sève phloémienne, aussi appelée sève élaborée) hébergent des bactéries qui leur fournissent des acides aminés (Baumann 2005), etc.

Bien que la plupart des symbioses obligatoires sont des symbioses nutritives, ils existent quelques cas de symbiotes obligatoires jouant un rôle différent. Le champignon *Streptomyces philanthi* secrète des antibiotiques qui augmentent très fortement (d'un facteur douze environ) la survie larvaire de la guêpe *Philanthus triangulum*. Ces champignons sont très probablement transmis verticalement lors de l'émergence de la larve (Kaltenpoth *et al.* 2005). Supprimer le symbiote *Wigglesworthia* (obligatoire pour la nutrition) des mouches tsé-tsé les rend non fertiles (ce qui est attendu), mais aussi plus susceptibles à l'infection par le trypanosome (Pais *et al.* 2008). Toutes les fourmis du genre *Camponotus* possèdent le symbiote *Blochmannia* qui leur apporte des acides aminés. En plus de ce rôle, cette bactérie améliore la capacité de ses hôtes à encapsuler ce qui les rend moins susceptibles aux attaques des mouches parasitoïdes de la famille des Phoridae (De Souza *et al.* 2009).

### **Les symbioses primaires sont des associations anciennes**

Ces associations sont anciennes et résultent d'un événement d'infection ancestral, puis l'endosymbiote a été transmis exclusivement verticalement *via* la lignée germinale maternelle à tous ses descendants. Si on exclue les mitochondries et les plastes, la plus vieille connue est l'association entre la bactérie *Sulcia muelleri* et les Auchenorrhynches (cicadelles, membracides, cercopes, fulgores...) qui remonte à 270 millions d'années (Moran *et al.* 2005). On peut citer de nombreuses associations insectes-bactéries qui révèlent cette longue coévolution entre les deux partenaires : les pucerons (*Buchnera aphidicola*, Munson *et al.* 1991), les mouches tsé-tsé (*Wigglesworthia glossinidia*, Chen *et al.* 1999), les blattes (*Blattabacterium*, Lo *et al.* 2003), les aleurodes (*Portiera aleyrodidarum*, Thao & Baumann 2004a), les psylles, (*Carsonella*, Thao *et al.* 2000), les cochenilles (*Tremblaya princeps*, Baumann & Baumann 2005), les fourmis du genre *Camponotus* (*Blochmannia*, Degan *et al.* 2004). Tous ces symbiotes sont visibles sur la figure 6.

Les nombreux génomes complets d'endosymbiotes primaires nous permettent maintenant de décrire leurs caractéristiques, résultat de leur « mode de vie », et d'appréhender leur évolution. Les similarités entre les différents génomes de symbiotes primaires nous indique que ce sont les mêmes forces évolutives qui sont à l'oeuvre. Le symbiote primaire typique a un génome très réduit, un contenu en base très biaisé en faveur des bases A (Adénine) et T (Thymine) ainsi qu'un taux d'évolution très rapide qu'on retrouve dans le taux de substitution en acides aminés (Moran *et al.* 2008). La taille des génomes connus varie de 160 (*Carsonella*) à 800 (*Blochmannia*) kilobases et le taux de GC de 16 à 33 %, contrairement aux bactéries libres qui ont un génome 5 à 10 fois plus grand, et un taux en GC beaucoup plus proche de 50 % (Moran *et al.* 2008). La plupart des fonctions cellulaires sont touchées par cette réduction du génome (c'est-à-dire la disparition de certains gènes), par exemple les gènes de réparation de l'ADN sont

## Introduction

globalement moins nombreux, ce qui explique probablement le biais de composition en base. L'hypothèse de l'absence de gènes de réparation comme facteur expliquant le biais est appuyée par le fait que le plus faible biais est observé chez *Baumannia*, symbiote primaire qui a conservé le plus de gènes de réparations, à l'inverse de *Carsonella* et *Sulcia* (Moran *et al.* 2008). Le biais de composition en base pourrait aussi résulter de la compétition pour les ressources (Rocha & Danchin 2002). Les symbiotes subissent une dérive génétique élevée ce qui diminue d'autant l'efficacité de la sélection, même sur des gènes *a priori* importants. Cet effet de dérive est dû à leur taille de population réduite et au goulot d'étranglement lors du transfert à la génération suivante (environ 36 000 cellules de la bactérie *Buchnera* sont présentes dans un individu pour *Acyrtosiphon pisum*, et seulement 850 bactéries sont transmises à chaque œuf chez *Nasonovia*, Mira & Moran 2002), au contraire des bactéries libres dont les tailles de populations immenses permettent à la sélection d'agir sur de très faibles différences de valeur sélective.

## Les symbiotes secondaires

Les symbioses secondaires (ou facultatives) sont le résultat d'une coévolution beaucoup plus courte que les symbioses primaires, à l'échelle d'une seule espèce. A la différence des symbiotes primaires, les symbiotes secondaires sont facultatifs pour l'hôte, mais pas pour la bactérie qui ne peut vivre sans lui. Il existe toutefois des cas de culture de ces symbiotes dans des lignées cellulaires (Darby *et al.* 2005) voire même de culture en milieu artificiel (Dale & Moran 2006). Comme les symbiotes primaires, les symbiotes secondaires sont transmis verticalement et occasionnellement de manière horizontale, au sein d'une espèce ou même entre différentes espèces, ce qui leur permet de coloniser des hôtes naïfs (sous entendu non infectés), ce que nous verrons par la suite. La transmission verticale est maternelle, (comme les symbiotes primaires, de la mère aux descendants) bien qu'il existe aussi au moins un cas de transmission paternelle, chez le puceron (Moran & Dunbar 2006). Les symbiotes secondaires peuvent être bénéfiques pour l'hôte, mais aussi délétères et ceci en fonction de l'environnement.

Ces symbiotes peuvent se retrouver dans les bactériomes, en plus du symbiote primaire comme *Arsenophonus* chez l'aleurode *Dialeurolonga mallewaramensis* (Pandey & Rajagopal 2015), mais peuvent être présents aussi dans tout l'hémolymphe de l'insecte, extracellulairement, comme *Rickettsia* chez *B. tabaci* (Gottlieb *et al.* 2008) ou *Spiroplasma* chez la drosophile (Anbutsu & Fukatsu 2006). On peut aussi les retrouver dans d'autres tissus comme le tissu adipeux, les muscles, le système nerveux, le tube digestif et les organes reproducteurs (Moran *et al.* 2008).

Le symbiote facultatif n'est pas forcément présent en plus d'un symbiote obligatoire, par exemple les drosophiles peuvent être infectées par *Spiroplasma* et *Wolbachia* mais n'ont pas de symbiote primaire (Mateos *et al.* 2006).

### Les effets phénotypiques de ces symbiotes

Les bénéfices apportés par les symbiotes peuvent être de plusieurs types tels que la protection contre différents stress, l'accroissement de la survie ou la protection contre divers ennemis naturels. Pour des revues sur les effets phénotypiques de ces symbiotes, voir Janson *et al.* (2008) ; Moran *et al.* (2008) ; Oliver *et al.* (2010). Nous allons décrire quelques exemples pour illustrer cette diversité d'effets. La bactérie *Serratia symbiotica* améliore la résistance au stress thermique (d'un facteur trois) chez le puceron *Acyrtosiphon pisum* (Montllor *et al.* 2002), effet que l'on peut retrouver également chez l'aleurode *B. tabaci* grâce à la bactérie *Rickettsia* (Brumin *et al.* 2011). Plusieurs cas de résistance à divers ennemis naturels sont documentés : la bactérie *Regiella insecticola* permet de lutter contre un champignon parasite et une guêpe parasitoïde, respectivement chez les pucerons *A. pisum* et *A. fabae* (Scarborough *et al.* 2005 ; Vorburget *et al.* 2010), tandis que certaines *Wolbachia* permettent à *Drosophila melanogaster* de lutter contre des virus à ARN (Hedges *et al.* 2008). Chez *Drosophila neotestacea*, la bactérie *Spiroplasma* confère la résistance à un nématode parasite (Jaenike *et al.* 2010).

Dans la partie précédente traitant des symbiotes primaires, nous avons vu que ceux-ci jouent souvent un rôle dans la nutrition, mais certains symbiotes facultatifs peuvent également jouer ce rôle. Par exemple chez *Drosophila melanogaster*, la bactérie *Wolbachia pipientis* est impliquée dans le métabolisme du fer : la fécondité de la drosophile est plus élevée lorsqu'elle est infectée, que ce soit en présence de basse ou haute concentration en fer dans le milieu (Brownlie *et al.* 2009). Un exemple plus classique et plus documenté est le cas du puceron *A. pisum* qui est en fait un complexe de populations génétiquement différenciées, spécialisées sur diverses plantes hôtes. Ces populations ont des combinaisons de symbiotes secondaires différentes, suggérant un rôle de ceux-ci dans l'exploitation de la ressource, bien que les bactéries ne soient pas complètement exclusives des différentes populations (Simon *et al.* 2003 ; Ferrari *et al.* 2004 ; Frantz *et al.* 2009), voir la figure 7 qui montre la différence de cortège symbiotique d'une plante à l'autre.

Dans un registre légèrement différent, la bactérie *Rickettsiella* change la couleur du puceron *A. pisum*, ce qui pourrait lui permettre indirectement d'échapper à des prédateurs (Tsuchida *et al.* 2010).

Certains symbiotes peuvent avoir des effets indirects dramatiques sur les plantes. Par exemple, il a été montré que les bactéries symbiotiques pouvaient favoriser la transmission de virus phytopathogènes (les symptômes provoqués sont présentés sur la figure 8) par des insectes phytophages (pucerons et aleurodes). En effet, les particules virales se lient aux protéines GroEL, produites par le symbiote, qui circulent dans l'hémolymphe (le virus PLRV<sup>1</sup> et *Myzus persicae* par exemple, Heuvel *et al.* 1994), leur permettant probablement d'échapper au système immunitaire. Il a été montré qu'*Hamiltonella* pouvait faciliter la transmission du TYLCV<sup>2</sup> par

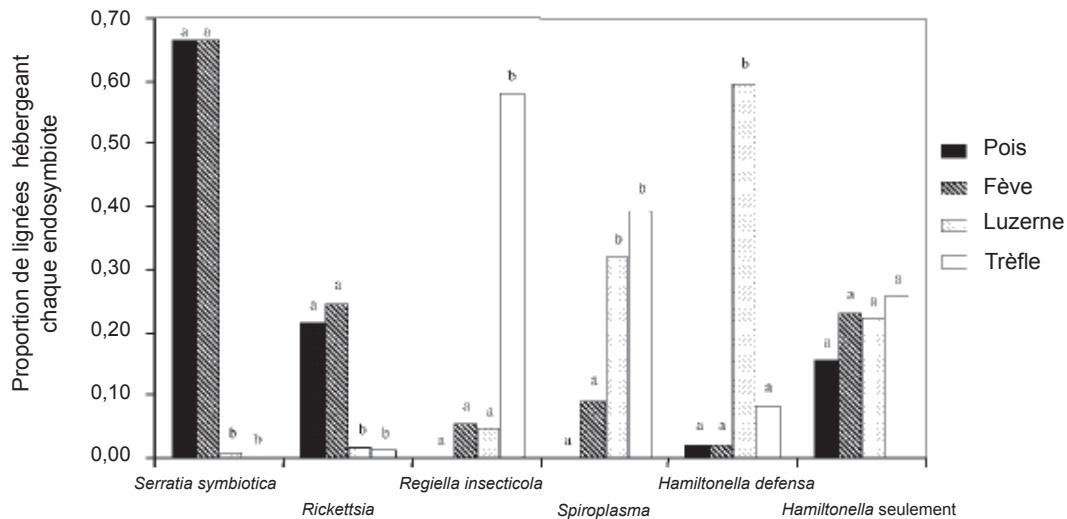
---

1. PLRV (Potato leaf roll virus) : virus du genre *Poleovirus* [Luteoviridae] responsable de l'enroulement de la pomme de terre.

2. TYLCV (Tomato Yellow Leaf Curl Virus) : virus du genre *Begomovirus* [Geminiviridae] responsable des feuilles jaunes en cuillère de la tomate.

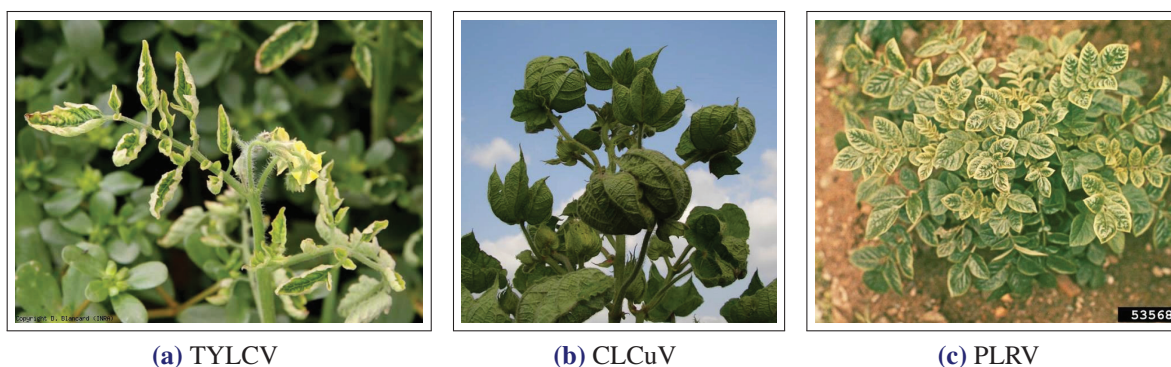


## Introduction



**FIGURE 7** – Prévalence de chaque symbiote en fonction de la plante. Les barres avec la même lettre indiquent que la prévalence d’une bactérie n’est pas différente entre les populations ( $P > 0,05$ ). Tiré de (Frantz *et al.* 2009)

*B. tabaci* (Gottlieb *et al.* 2010) et que la bactérie *Arsenophonus* pouvait faciliter la transmission du CLCuV<sup>1</sup> par ce même vecteur (Rana *et al.* 2012).



**FIGURE 8** – Symptômes provoqués par différents virus phytopathogènes. (a) : Pied de tomate atteint par le TYLCV. (b) : Plant de coton touché par le CLCuV. (c) : Plant de pomme de terre atteint par le PLRV.

Certains effets phénotypiques apparaissent comme délétères au premier abord, comme dans le cas de *Drosophila simulans* dont la capacité à encapsuler est réduite lorsqu’elle héberge la bactérie *Wolbachia*, au bénéfice de la guêpe *Leptopilina heterotoma* (Fytrou *et al.* 2006). Mais puisque le symbiote se maintient dans les populations, il apporte nécessairement un avantage sélectif que l’on ne perçoit peut-être pas directement car l’effet est environnement dépendant.

Un exemple classique est celui d’*Hamiltonella defensa* qui protège le puceron de ses parasitoïdes, la fréquence de la bactérie augmente alors dans la population (preuve de l’avantage sélectif), mais en son absence la pression de sélection sur les non infectés diminue, la fréquence

1. CLCuV (Cotton leaf curl virus) : virus du genre *Begomovirus* [Geminiviridae] est l’agent de la cloque du coton.

de la bactérie diminue aussi (Oliver *et al.* 2003). Ce même endosymbiote ne semble pas procurer ce bénéfice chez *B. tabaci* qui héberge aussi cette bactérie.

En Californie, la fréquence de la bactérie *Arsenophonus* varie de 0 à 75 % chez le psylle *Glycaspis brimblecombei*. En fait cette fréquence est corrélée à la pression de parasitisme par l'hyménoptère *Psyllaphaegus bliteus* suggérant un effet protecteur du symbiote (Hansen *et al.* 2007).

### **La transmission horizontale des symbiotes secondaires**

Plusieurs observations montrent que les symbiotes secondaires se transmettent horizontalement entre individus de la même espèce et entre différentes espèces. Bien que de nombreux cas soient documentés il existe peu d'informations sur la fréquence de ces transferts.

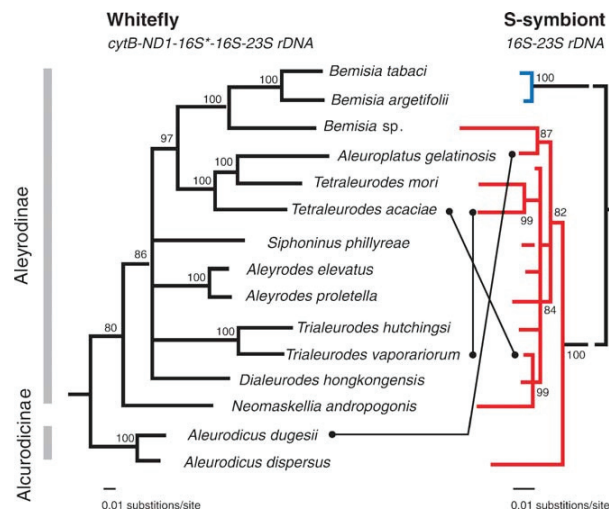
Ce mode de transmission explique pourquoi certains symbiotes sont retrouvés dans des taxons très divers. En effet, des phylogénies discordantes entre symbiotes et hôtes montrent que les infections sont des phénomènes récents et ne sont pas ancestrales à l'échelle de l'évolution des clades. De nombreux cas de tels transferts sont décrits dans la littérature et impliquent des bactéries telles que *Arsenophonus* ou *Wolbachia* (Thao & Baumann 2004a ; Vavre *et al.* 1999 ; Baldo *et al.* 2008 ; Mouton *et al.* 2012). La figure 9 montre le transfert du symbiote secondaire à l'intérieur du clade des aleurodes grâce aux phylogénies incongruentes entre les hôtes et leur symbiote secondaire.

Plusieurs mécanismes de transferts horizontaux ont été mis en évidence : lors de l'accouplement, *via* des parasites ou encore par l'intermédiaire de la plante hôte (Moran & Degnan 2006 ; Ahmed *et al.* 2015 ; Caspi-Fluger *et al.* 2012). Ces mécanismes de transferts horizontaux n'impliquent pas une distribution aléatoire de ces bactéries, certains groupes d'insectes ont ainsi une fréquence d'infection élevée pour certains symbiotes. Par exemple, *Arsenophonus* est très présent chez les aleurodes (Mouton *et al.* 2012) mais rare chez les pucerons (Russell *et al.* 2003), *Wolbachia* est présente chez les termites (Lo & Evans 2007) mais rare chez les pucerons (Russell *et al.* 2003).

### **Les multi-infections**

La transmission horizontale des symbiotes permet l'émergence d'individus multi-infectés, lorsqu'une bactérie est transmise à un individu déjà infecté. Nous avons déjà parlé du cas de symbiotes secondaires présents en plus du symbiote primaire comme chez le puceron ou la mouche tsétsé. Plus que deux symbiotes, un hôte peut héberger une multitude de symbiotes secondaires, que ce soit à l'échelle de l'espèce (chaque individu n'a pas forcément tous les symbiotes) où à l'échelle individuelle. Les données sur les multi-infections ne sont pas assez conséquentes pour estimer l'incidence de ce phénomène, puisqu'on a déjà du mal à déterminer le nombre d'espèces qui hébergent au moins un symbiote donné (Duron *et al.* 2008 ; Weinert *et al.* 2015). Néanmoins il y a de nombreux cas documentés.

Les plus connus, car les plus extrêmes, sont le puceron *A. pisum* et l'aleurode *B. tabaci*. A



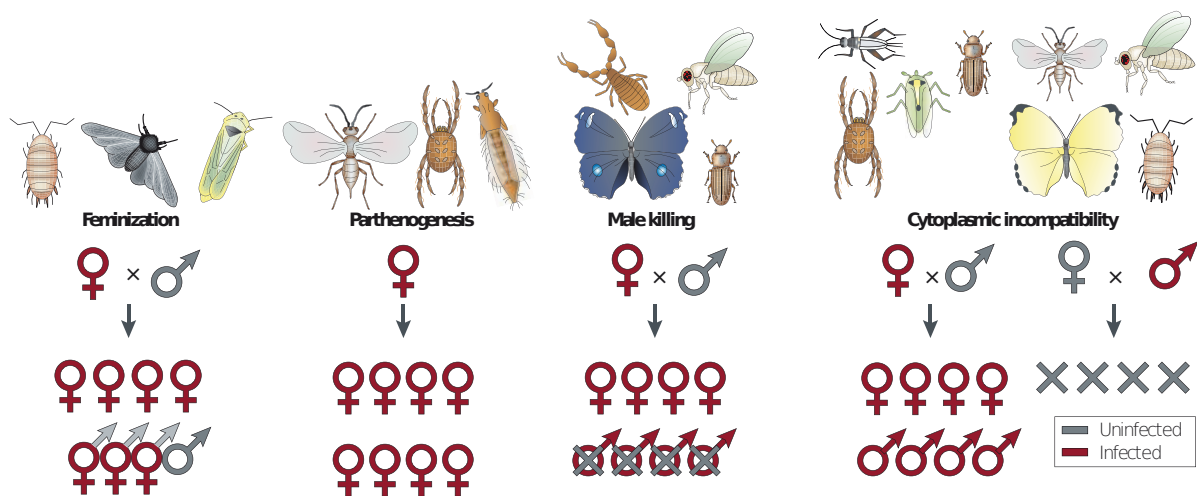
**FIGURE 9** – Comparaison entre la phylogénie des aleurodes réalisée en combinant un marqueur mitochondrial et un gène du symbiote primaire, et la phylogénie de leur symbiote secondaire (Baumann 2005).

l'échelle de l'espèce le premier peut héberger six symbiotes secondaires en plus du primaire (Darby *et al.* 2001 ; Sandström *et al.* 2001 ; Russell *et al.* 2003 ; Frantz *et al.* 2009), et le second sept (Jeyaprakash & Hoy 2000 ; Zchori-Fein & Brown 2002 ; Weeks *et al.* 2003 ; Thao *et al.* 2003 ; Gottlieb *et al.* 2006). Ce nombre de symbiotes retrouvés dans une espèce n'implique pas que tous les individus sont infectés par tous ces symbiotes en même temps. Ainsi les données de Frantz *et al.* (2009) sur le puceron montrent qu'environ la moitié des individus n'ont qu'un symbiote facultatif, un quart en ont deux et seulement un sur les 313 individus de l'étude a trois symbiotes. Chez *B. tabaci*, Gueguen *et al.* (2010) montrent que 59 % des individus ont deux symbiotes facultatifs et 6 % en ont plus de deux. On peut citer aussi le cas du balanin (Merville *et al.* 2013) dont les individus peuvent héberger deux symbiotes secondaires. Weinert *et al.* (2007) ont montré lors d'une recherche systématique de symbiotes chez des coccinelles que, parmi 21 espèces, 11 étaient infectées par au moins un symbiote et deux (*Rhizobius litura* et *Coccidula rufa*) avaient des individus infectés par deux symbiotes, ce qui fait environ 10 % de multi-infections (bien sûr on ne tirera pas de conclusions quantitatives sur ce seul exemple).

Pour rendre la chose plus compliquée, et qui fait que nous sous-estimons encore plus ce phénomène de multi-infection, de nombreuses espèces de ces symbiotes présentent une certaine variabilité, identifiée par des souches, augmentant d'autant le nombre de symbiotes. Par exemple chez *Curculio venosus* on retrouve deux souches de *Wolbachia* (Merville *et al.* 2013) comme chez *B. tabaci*, chez qui on retrouve aussi plusieurs souches d'*Arsenophonus* (Gueguen *et al.* 2010). Un autre exemple se trouve chez le puceron qui peut être infecté par plusieurs souches d'*Hamiltonella* (Degnan & Moran 2008), ou encore la guêpe parasitoïde *Leptopilina heterotoma* qui peut être infectée simultanément par trois souches de *Wolbachia* (Vavre *et al.* 1999).

## Les manipulateurs de la reproduction

Les manipulateurs de la reproduction, toujours à transmission verticale, vont favoriser les lignées maternelles qui sont le support de leur transmission, même si cela se fait au dépend de la valeur sélective des mâles. Il y a deux grandes stratégies : 1) Réduire le nombre de descendants femelles par les femelles non infectées grâce à l'incompatibilité cytoplasmique. 2) Déséquilibrer le sexe-ratio en faveur des femelles dans la descendance des femelles infectées, à travers trois mécanismes : le *male-killing*, la féminisation des mâles génétiques ou encore l'induction de la parthénogenèse thélytoque. L'incompatibilité cytoplasmique se caractérise par une réduction totale ou partielle du nombre de descendants viables lors du croisement entre un mâle infecté et une femelle non infectée. Cet effet tend à diminuer la capacité de reproduction des femelles non-infectées, ce qui permet l'invasion des populations. Dans certains cas, le symbiote manipulateur peut entraîner la dégénérescence des embryons mâles c'est le *male killing*, dans d'autres cas le symbiote peut féminiser les embryons en transformant les mâles génétiques en femelles fonctionnelles. Dans le cas de la parthénogenèse thélytoque, le symbiote manipulateur induit la diploïdisation des œufs non fécondés qui vont ainsi se développer en femelles. Les femelles infectées ne produisent donc plus que des femelles (2 fois plus que les femelles non infectées) qui peuvent à leur tour se reproduire sans accouplement. Tous ces mécanismes sont résumés sur la figure 10. *Wolbachia* est la bactérie la plus connue pour manipuler la reproduction, mais elle n'est pas la seule : *Arsenophonus*, *Spiroplasma Cardinium* et *Rickettsia* sont aussi capables de manipuler la reproduction. L'incidence de ces bactéries est élevée chez les arthropodes, mais difficile à estimer (l'estimation varie beaucoup selon la fréquence minimum choisie pour dire qu'une espèce est infectée), et est probablement sous estimée (Duron *et al.* 2008 ; Werren *et al.* 2008 ; Weinert *et al.* 2015).



**FIGURE 10** – Manipulation de la reproduction par la bactérie *Wolbachia*. L'illustration est extraite de Werren *et al.* (2008)

## Quel rôle peut jouer la symbiose dans l'invasion biologique ?

Nous avons vu que les symbiotes peuvent jouer de très nombreux rôles dans l'écologie de leur hôte, notamment en permettant l'exploitation de nouvelles niches (que ce soient les symbiotes primaire ou secondaires) et ainsi potentiellement augmenter le potentiel invasif des populations. Mais en plus ces populations peuvent être vectrices de pathogènes qui peuvent interagir avec les symbiotes de ces populations. Enfin, les méthodes de lutte contre certains ravageurs peuvent être perturbées par certains symbiotes. Nous allons détailler quelques exemples pour chacun de ces cas qui montrent la nécessité de prendre en compte le compartiment symbiotique dans la biologie de l'invasion, jusqu'à présent plutôt négligé (Zindell *et al.* 2011).

### Résistance contre des moyens de lutte

*Aphidius colemani* est utilisé pour contrôler les populations de pucerons sur différentes cultures. Or *Regiella insecticola* est responsable de résistances des pucerons envers ce parasitoïde (Vorburger *et al.* 2010). Un deuxième exemple est celui de l'hyménoptère *Psyllaephagus biteus* qui a été introduit en Californie pour contrôler les populations naturelles de *Glycaspis brimblecombei* qui se nourrit sur *Eucalyptus camaldulensis*. Un symbiote secondaire a été mis en évidence chez *G. brimblecombei*, il permet la résistance à son parasite et est retrouvé en plus forte fréquence dans les populations où *P. biteus* est présent (Hansen *et al.* 2007). A l'inverse, *Rickettsia* augmente la susceptibilité de l'aleurode *B. tabaci* à certains insecticides (Kontsedalov *et al.* 2008). Ainsi, connaître les bactéries symbiotiques présentes chez un ravageur pourrait permettre d'améliorer les moyens de lutte biologique ou d'orienter les traitements insecticides mis en place.

### Augmentation du potentiel invasif de l'hôte

On sait que les champignons qui forment des associations symbiotiques avec des plantes peuvent aider celles-ci en améliorant leur valeur sélective. Par exemple l'invasion par des plantes non indigènes peut-être facilitée par des champignons mycorhiziens (Richardson *et al.* 2000). L'invasion par la plante *Festuca arundinacea*, introduite en Amérique du Nord depuis l'Europe, pourrait avoir été facilitée par la présence du champignon endophyte *Epichloë coenophialum*, en induisant une plus forte résistance, une toxicité contre les herbivores et une résistance à la sécheresse que n'ont pas les herbes indigènes (Rudgers *et al.* 2005). Bien que le rôle de ces associations serait faible, il serait significatif (Pour une revue voir Bennett 2013).

Aujourd'hui, peu d'études ont pu relier des traits phénotypiques apportés par les endosymbiotes à un rôle dans l'invasion, malgré le nombre élevé d'insectes phytophages infectés, dont nombres sont invasifs (Douglas 2009). En tout cas, il a été montré que la mineuse *Phyllonorycter blancardella* est infectée par une *Wolbachia* qui manipule le niveau de cytokine de la plante (Kaiser *et al.* 2010). Ceci augmente la valeur sélective de *P. blancardella* en retardant la sénescence de

## 2. La symbiose, actrice de l'invasion ?

la feuille et l'obligant à envoyer des nutriments dans les zones où la mineuse se nourrit (Giron *et al.* 2013).





# Problématique générale

Le projet s'inscrit dans le cadre d'un questionnement sur les mécanismes fondamentaux impliqués dans le succès des invasions biologiques, en considérant à la fois le compartiment cytoplasmique (bactéries symbiotiques et mitochondries) et le compartiment nucléaire (gènes de l'insecte). On sait que l'admixture joue un rôle dans les invasions biologiques, en augmentant le potentiel adaptatif par l'apport d'allèles et la création de nouvelles combinaisons nucléaires, supports de cette adaptation.

D'autre part, nous avons vu que les symbiotes peuvent jouer de très nombreux rôles dans la biologie et l'écologie de leur hôte, notamment en apportant des nutriments permettant l'exploitation de nouvelles ressources, en protégeant leur hôte contre divers ennemis naturels ou des insecticides, voir même manipuler leur reproduction. Ces effets leurs permettent d'envahir très rapidement les populations d'insectes et ainsi modifient considérablement leur biologie et leur écologie. Il apparait donc évident que la symbiose puisse être un facteur influent dans un contexte d'invasion, et probablement aussi dans la dynamique des ravageurs. La prise en compte des bactéries symbiotiques a été plutôt négligés dans la littérature (Zindel *et al.* 2011 ; Mcneill *et al.* 2014).

L'admixture met en contact des compartiments nucléaires et cytoplasmiques différents, donc en plus de créer des nouvelles combinaisons nucléaires comme cela est classiquement étudié, de nouvelles combinaisons nucléo-cytoplasmiques vont se former. De plus, les bactéries symbiotiques peuvent transférer d'un individu à l'autre, créant de nouvelles combinaisons.

Les interactions entre bactéries peuvent conduire à l'exclusion de l'une par l'autre du fait de la compétition intra-hôte, mais aussi diminuer la valeur sélective de l'hôte du fait de la compétition au niveau populationnel. A l'inverse, les bactéries peuvent cohabiter et la nouvelle association être sélectionnée au niveau populationnel. Un deuxième effet d'interaction provient de la mise en contact de fonds nucléaires et cytoplasmiques réciproquement inconnus avant ces transferts. En effets, chaque bactérie a un effet individuel, mais que peut-on dire des effets additifs entre compartiment nucléaire et cytoplasmique, ou même, des interactions (négative ou positives) potentielles sur le phénotype de l'hôte ? On peut supposer que de nouvelles combinaisons nucléo-cytoplasmiques plus performantes vont être créées — sorte de super génotype issu de la sélection de la meilleure combinaison possible parmi les éléments nucléaires et cytoplasmiques mis en contact —, mais ce n'est pas forcément vrai. Des associations préalablement coadaptées — du fait d'une histoire évolutive commune — vont être brisées du fait de la re-



combinaison et des associations moins performantes créées, l'admixture jouant alors un rôle de frein à l'adaptation.

Les complexes d'espèces invasifs nous permettent d'étudier ces phénomènes pour plusieurs raisons. 1) Par définition un complexe d'espèces est composé de multiples entités dont les barrières reproductives sont peu connues 2) Étant invasives, les populations sont en mouvement permanent, ce qui donne autant de possibilités d'hybridations de fonds nucléaires différents, tant autochtones que allochtones. 3) De manière analogue au compartiment nucléaire, de nouvelles combinaisons cytoplasmiques vont pouvoir émerger par l'intermédiaire de transferts horizontaux entre entités du complexe d'espèces.

Ce travail vise à appréhender simultanément les compartiments cytoplasmique et nucléaire d'un complexe d'espèces pour comprendre leur rôle respectif sur les processus d'invasion, d'adaptation et d'explosion démographique d'un ravageur. *Bemisia tabaci* est un modèle biologique très particulier puisqu'il est composé de nombreuses entités génétiques dont les limites reproductives sont mal définies, probablement parce que la radiation du clade est récente (Santos-Garcia *et al.* 2015). Le fait que ces entités soient différentiellement infectées par certaines bactéries donnent l'occasion d'étudier les phénomènes d'admixture par l'étude simultanée des compartiments cytoplasmique et nucléaire en contexte invasif.

*B. tabaci* est très probablement soumis à des introductions multiples du fait des échanges commerciaux <sup>1</sup> qui sont autant d'occasions d'admixture. Pourtant, la plupart des populations observées sont homogènes (sur le plan cytoplasmique) et il semble donc que la compétition induite lors du mélange de populations conduise au remplacement de l'une par l'autre. La situation en France est particulière car depuis quelques années deux biotypes (entités différenciées par leurs caractéristiques symbiotiques, génétiques ou écologiques) se retrouvent en sympatrie, voire dans des populations mixtes. Cette situation originale va nous permettre d'étudier les conséquences nucléaire et cytoplasmique de l'hybridation sur la diversité génétique observée dans l'aire d'introduction

Avant de passer au travail de recherche proprement dit, nous allons maintenant décrire ce modèle biologique et montrer en quoi il est intéressant pour répondre à ce questionnement.

---

1. Entre 2010 et 2011, des plantes interceptées par le PHSI (Plant Health and Seeds Inspectorate) au Royaume-Uni et infectées par *B. tabaci* provenaient de 12 pays d'Europe, d'Afrique d'Asie et d'Amérique. (Powell *et al.* 2012)

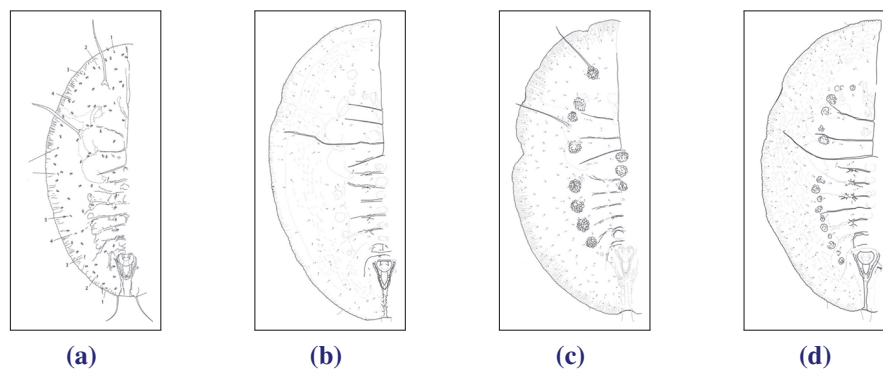
# Le modèle d'étude *Bemisia tabaci*

1	Description écologique	39
2	Des races d'hôtes au complexe d'espèces	41
3	Les associations symbiotiques chez <i>Bemisia tabaci</i>	48
3.1	Le symbiote primaire	48
3.2	Les symbiotes secondaires	49
3.2.1	Généralités	49
3.2.2	Description des partenaires symbiotiques secondaires	51
4	Matériel et méthodes général	56
4.1	Élevage des insectes	57
4.2	Détermination du cytoype	58



## Description écologique

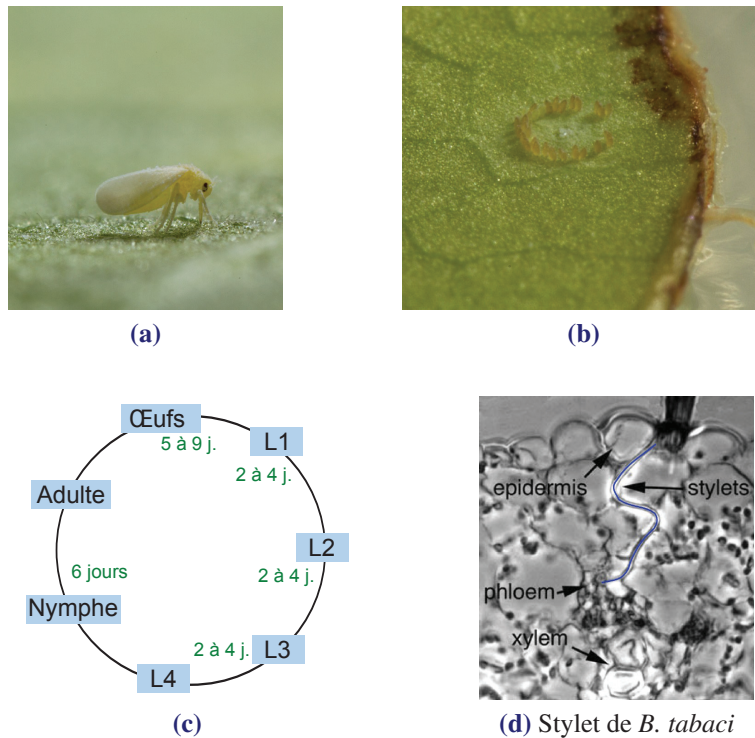
*Bemisia tabaci* (Gennadius, 1889), appelé aussi aleurode du tabac, de la patate douce ou du cotonnier, ou encore improprement mouche blanche, appartient à la famille des Aleyrodidae (Hémiptère : Sternorrhynche), large famille composée de plus de 1 500 espèces réparties en 161 genres (Martin & Mound 2007). Comme il n'y a pas assez de caractères discriminants chez l'adulte, la systématique des aleurodes repose sur l'observation des caractères de la puppe (voir figure 11), ce qui n'est pas toujours suffisant comme nous le verrons plus loin. *B. tabaci* est haplodiploïde (Blackman & Cahill 1998), c'est-à-dire que les femelles sont diploïdes, issues d'œufs fécondés alors que les mâles sont haploïdes, issus d'œufs non fécondés. Les adultes mesurent environ 1 mm.



**FIGURE 11** – Illustrations montrant le côté gauche des pupes chez différentes espèces de *Bemisia*. Les figures (b) et (c) illustrent le changement morphologique lié à l'environnement. (a) : *B. tabaci*, (b) : *B. berbericola*, (c) : *B. berbericola* avec développement de setas dû aux poils foliaires, (d) : *B. hancocki*

Le cycle de vie, d'une durée totale de 17 à 27 jours suivant la température, comprend quatre stades larvaires, puis le stade nymphe (ou puppe) donnant lieu à l'adulte (voir figure 12c). L'adulte vit sur les faces inférieures des feuilles, se déplaçant peu et principalement pendant les heures diurnes. Il se nourrit de sève phloémienne grâce à son stylet qu'il enfonce entre les cellules jusqu'à trouver les tubes criblés où elle circule (voir figure 12d). La sève phloémienne est élaborée, elle est donc riche en glucides tels que le saccharose, le sorbitol et le mannitol, mais pauvres en acides aminés et vitamines. Comme beaucoup d'autres insectes se nourrissant de phloème il peut se nourrir occasionnellement de xylème (Jiang 1999), très dilué et pauvre en nutriments, *a priori* lorsque l'insecte est déshydraté. La symbiose obligatoire entre *Portiera aleyrodidarum* et *B. tabaci* procure à ce dernier des acides aminés essentiels qu'il ne peut trouver dans son milieu.

*B. tabaci* a d'abord été étudié en tant que ravageur de cultures mondialement distribué (voir la carte page 44), en milieux anthropisés ou non dans les régions tropicales et subtropicales et seulement en serre sous les climats tempérés. Il touche autant des productions vivrières (tomate, patate douce, melon, concombre...) qu'ornementales (poinsettia, lantana...). On le retrouve énormément sur des cultures tels que le coton ou le tabac (Brown *et al.* 1995), plante sur laquelle



**FIGURE 12** – (a) : Femelle adulte, (b) : Œufs, pondus en cercle. Photos de Thibault Andrieux. (c) : Cycle de vie de *B. tabaci*, (d) : Microphotographie du stylet (en bleu) d'un adulte à travers le tissu d'une feuille de haricot (Stansly *et al.* 2010).

il a été décrit la première fois en Grèce en 1889 (Gennadius 1889). Ces quelques exemples de plantes sont très loin de décrire la diversité des hôtes possibles puisque *B. tabaci* a été décrit sur des centaines de plantes appartenant à plus de 70 familles. A titre d'exemples, nous avons répertorié les plantes hôtes des individus dont des séquences sont sur *GenBank*, ce qui fait un total de 141 espèces appartenant à 40 familles (voir l'annexe A page 221). On continue encore à répertorier de nouveaux hôtes, Simmons *et al.* (2008) ont ainsi identifié 49 nouvelles espèces ! Les dégâts que *B. tabaci* induit sont causés directement par succion de la sève phloémienne et indirectement, soit par excrétion de miellat<sup>1</sup> — qui facilite l'apparition de fumagine (maladie cryptogamique provoquée par une moisissure) réduisant la photosynthèse des feuilles et les asphyxiant, ce miellat diminue aussi la valeur marchande des produits —, soit par la transmission de virus phytopathogènes comme le TYLCV<sup>2</sup>, dont il est le vecteur exclusif (Oliveira *et al.* 2001). Il peut potentiellement transmettre environ 200 virus (Polston *et al.* 2014) appartenant aux genres (nombre de virus vectorisés par *B. tabaci*), *Begomovirus* (192), *Crinivirus* (12), *Ipomovirus* (4), *Carlavirus* (2), *Torradovirus* (2). C'est pourquoi *B. tabaci* est soumis par l'arrêté du 31 juillet 2000 à des mesures de lutte obligatoire et à des mesures de quarantaine dans l'importation de végétaux. La détection du TYLCV entraîne la destruction systématique des plants concernés (directive européenne). *B. tabaci* est classé parmi les 100 espèces invasives les plus

1. Le miellat est un liquide épais et visqueux excrété par l'anus. Issu du métabolisme de l'insecte, il est riche en sucres et acides aminés.

2. Tomato Yellow Leaf Curl Virus

problématiques en Europe d'après le projet DAISIE<sup>1</sup>. Les dégâts sont tels que jusqu'à la totalité d'une récolte peut être perdue. Par exemple au Brésil, de 60 à 100 % des récoltes de grenadille (*Passiflora edulis*) ont pu être détruites (Oliveira *et al.* 2001) ; au Mexique, à Sonora, 65 % des cultures de coton ont été perdues dans les années 1990 (Silva-Sanchez 1997) et beaucoup plus récemment, à Moulès (Bouches-du-Rhône), la totalité de cultures d'Aubergines a été détruite (Observation personnelle).

### Des races d'hôtes au complexe d'espèces

*B. tabaci* a été décrit d'abord sous le nom de *Aleyrodes tabaci*, l'aleurode du tabac (Gennadius 1889). En 1897 un individu collecté aux États-Unis sur patate douce est identifié comme *Bemisia inconspicua*, l'aleurode de la patate douce. En 1957, 19 taxons décrivant des aleurodes sont rassemblés sous le même nom : *B. tabaci*. C'est pour cette raison qu'aujourd'hui on le connaît sous différents noms : aleurode du tabac, du cotonnier ou encore de la patate douce.

Dans ces mêmes années 1950, la découverte de populations morphologiquement identiques mais qui exprimaient des phénotypes différents tels que la gamme de plantes hôtes ou la transmission de virus ont amené les auteurs à proposer l'existence de biotypes ou races d'hôtes chez *B. tabaci*, voir Brown *et al.* (1995) pour une revue. Par exemple, à Puerto-Rico, une race était plutôt retrouvé sur le genre *Jatropha* [Euphorbiaceae] et transmettait un seul virus tandis qu'une autre était polyphage (*Sida*<sup>2</sup>) et transmettait plusieurs virus (Bird & Brown 1998). Le même schéma était retrouvée en Côte d'Ivoire où un biotype était confiné au Manioc [Euphorbiaceae] et transmettait le virus ACMV<sup>3</sup> tandis qu'un autre ne transmettait pas ce virus et était polyphage mais absent sur Manioc (Burban *et al.* 1992). Les auteurs ont considéré que cette variabilité phénotypique était la conséquence de variabilité génétique sous-jacente. Les premières considérations sur la variabilité génétique chez *B. tabaci* ont donc été faites à partir de traits phénotypiques différenciant des populations sympatriques sans tenir compte des populations allopatriques.

Parallèlement, les marqueurs génétiques se sont développés. Les premiers utilisés furent les estérases qui sont des marqueurs nucléaires basés sur des profils de migration différentiels de ces enzymes. Par la suite on utilisera comme marqueurs nucléaires successivement les AFLP (*Amplification Fragment Length Polymorphism*) qui ont le désavantage de montrer beaucoup d'homoplasie, et les microsatellites, les plus utilisés jusqu'à aujourd'hui pour la génétique des populations. Les marqueurs mitochondriaux sont extrêmement utilisés : les marqueurs PCR RFLP (*Restriction Fragment Length Polymorphism*) qui consistent à distinguer des séquences sur la base de différences dans le profil de restriction par une enzyme, ou le séquençage qui est abon-

1. DAISIE : Delivering Alien Invasive Species In Europe. Projet fondé par la commission européenne pour répertorier les invasions biologiques sur le continent.

2. Ce nom provient de *Sida cordifolia* [Malvaceae] et n'a rien à voir avec le Syndrome d'immunodéficience acquise

3. *African cassava mosaic virus*, virus qui provoque des chloroses, notamment sur Manioc

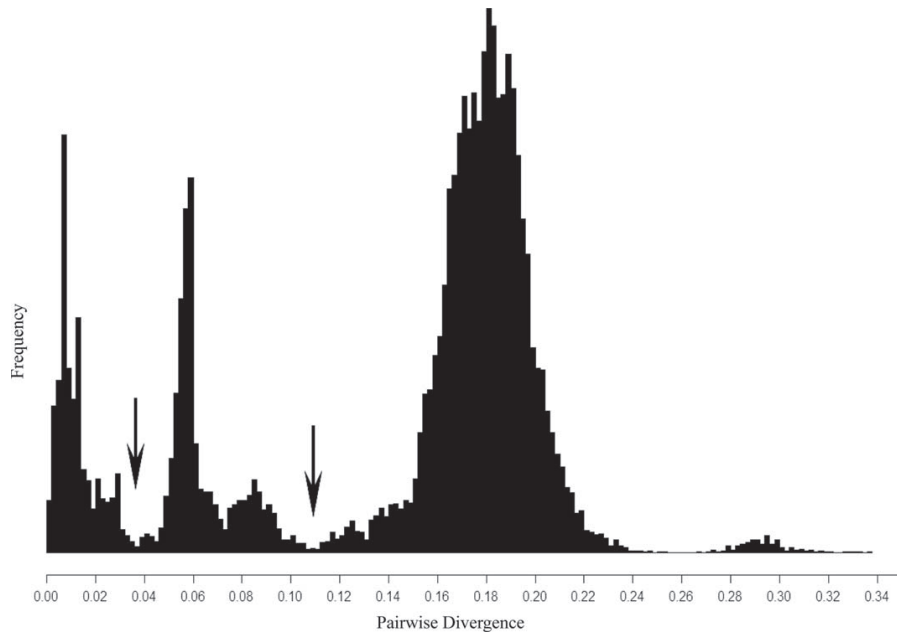
dament utilisé sur les gènes *ITS* (dont le signal sature et n'est plus très utilisé aujourd'hui) et *COI*, ce dernier étant à la base de la taxonomie actuelle du complexe d'espèces *B. tabaci*. Pour des revues retraçant l'historique d'utilisation de ces marqueurs, voir Brown *et al.* (1995) et De Barro *et al.* (2011)

La véritable course au biotype a commencé dans les années 1990 aux États-Unis avec différents épisodes invasifs très destructeurs en Floride puis en Arizona qui ont mis en évidence la présence de deux groupes génétiques dont l'hybridation n'était pas possible, avec des profils estérases différents et une utilisation différente des plantes hôtes (Costa & Brown 1991 ; Costa *et al.* 1993). Le groupe A était le groupe indigène (maintenant élevé au rang d'espèce potentielle sous le nom *B. tabaci* NEW WORLD), tandis que le B était l'invasif (maintenant *B. tabaci* Middle East Asia Minor 1, MEAM1). Le B sera élevé au rang d'espèce (*Bemisia argentifolii*) par Bellows *et al.* (1994), avant d'être à nouveau reconnu en tant que race de *B. tabaci* (De Barro *et al.* 2005) car il n'y avait pas de caractères morphologiques pour les différencier. La description avait été faite sur la présence d'un poil sur les pupes du biotype A, absent chez le biotype B. Ce critère s'étant avéré par la suite peu pertinent car la présence de poils sur les feuilles de la plante peu induire des changements phénotypiques de la puce (Stansly *et al.* 2010), voir par exemple la figure 11. De plus, ce critère était avant tout phénotypique et ne donnait aucune information sur d'éventuelles barrières reproductives. A l'inverse, rejeter cette espèce sur l'absence de critère morphologique est tout autant erroné puisqu'une barrière reproductive n'implique pas de différences morphologiques, ce qui est la définition même du complexe d'espèces cryptiques ! Cette absence de cadre clair a amené la communauté à décrire de très nombreux biotypes dont la réalité biologique sous-jacente était très peu connue, quasiment toutes les lettres de l'alphabet vont y passer pour désigner les nouveaux biotypes<sup>1</sup>, créant des synonymies et une grande confusion dans la littérature. Les critères utilisés étant la plante hôte, la capacité à disperser, l'induction d'argenture sur les feuilles, la transmission de virus, la résistance aux insecticides... Ces critères étaient couramment utilisés dans la littérature liée aux ravageurs de cultures (Saxena & Barrion 1987).

Ces données vont faire apparaître une grande variabilité dans le clade, variabilité compatible avec des groupes phylogéographiques, mis à part pour les deux groupes les plus invasifs : Q et B. En 2007 et 2010, deux études vont successivement identifier 12 puis 11 clades majeurs (superclade) à l'aide du gène mitochondrial de la cytochrome oxydase I (*mtCOI*, Boykin *et al.* 2007 ; Dinsdale *et al.* 2010). La puissance de la deuxième analyse, avec 202 haplotypes différents va permettre d'identifier 24 groupes mineurs et de quantifier les limites entre ces groupes (De Barro *et al.* 2011). En effet, la distribution des divergences nucléotidiques mitochondriales par paires d'haplotypes montre deux ruptures à 3,5 % et 11 % (pour un total de 650 paires de bases) qui vont permettre de délimiter des espèces potentielles sous l'hypothèse que ces ruptures sont dues à l'isolement reproducteur (voir figure 13).

1. Pour une liste (exhaustive ?) : A, AN, AnSL, ASL, B, B2, C, Cv, D, E, F, G, H, I, J, Jatropha, K, L, M, MS, N, NA, Nauru, Okra, P, PCG-1, PCG-2, PK1, Q, R, S, Sida, SY, T, ZHJ1, ZHJ2, ZHJ3

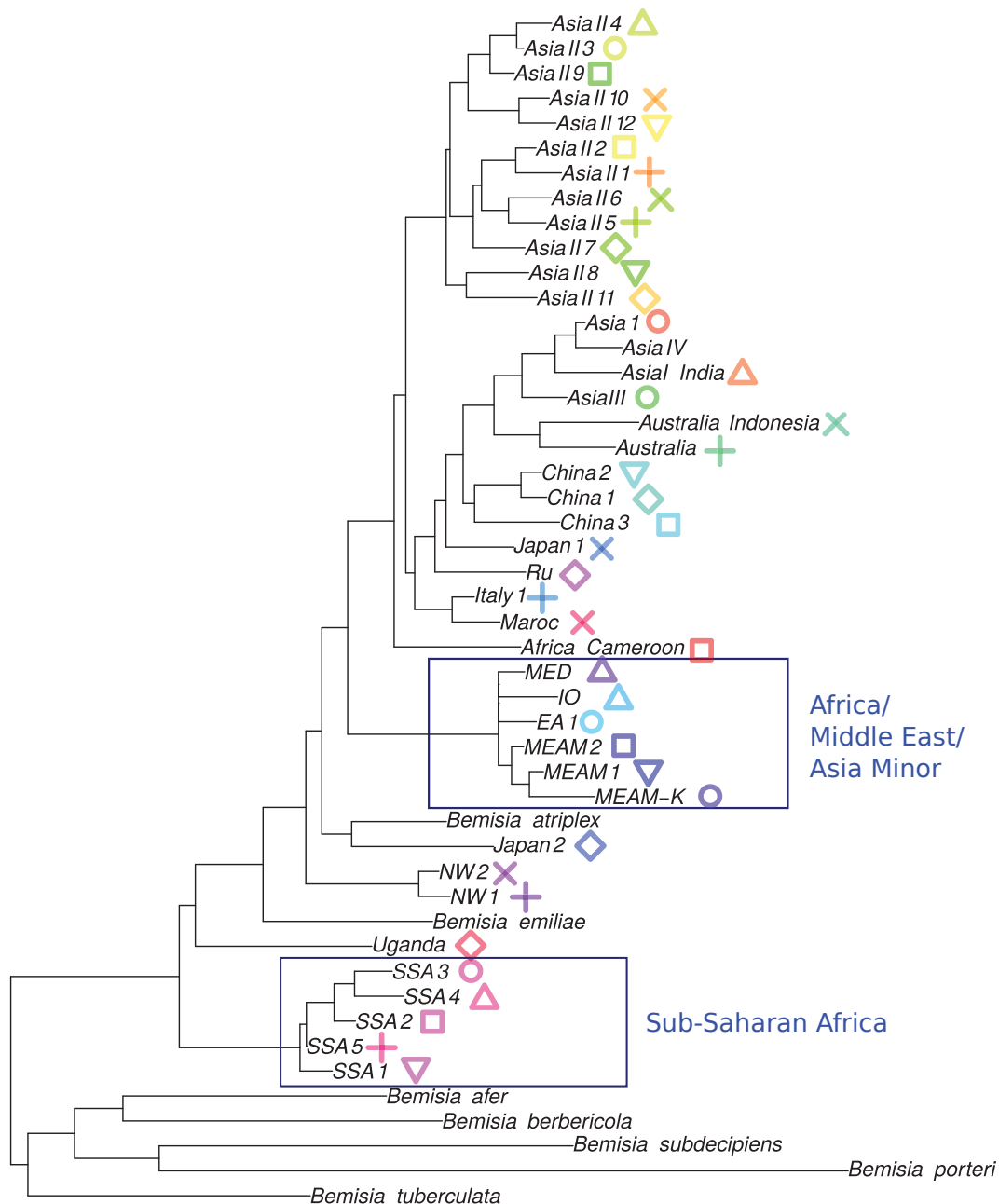




**FIGURE 13** – Distributions des distances mitochondriales de toutes les paires d'haplotypes connus de *B. tabaci* (Dinsdale *et al.* 2010). Les deux ruptures sont indiquées par des flèches.

La rupture à 11 % va permettre de définir des « super-clades » (voir la figure 14 page 44) et est probablement liée à des événements évolutifs anciens. La rupture à 3,5 % résulte de l'absence de flux de gènes à une échelle de temps plus courte, révélant la radiation récente du complexe *B. tabaci*, elle va permettre de définir les espèces potentielles du complexe *B. tabaci*. Cette valeur seuil de 3,5 % est compatible avec d'autres estimations de la différenciation entre espèces d'après ces auteurs, qui citent des travaux ponctuels (Barat *et al.* 2008 ; Burns *et al.* 2008) mais pas des méta-analyses. A l'issue de ce travail, il est proposé que le complexe d'espèces *B. tabaci* comprend 24 espèces potentielles identifiées d'après la valeur seuil de divergence mitochondriale à 3,5 % (De Barro *et al.* 2011) et corroborées en partie par des données de croisements expérimentaux mettant en évidence des barrières reproductives (Xu *et al.* 2010).

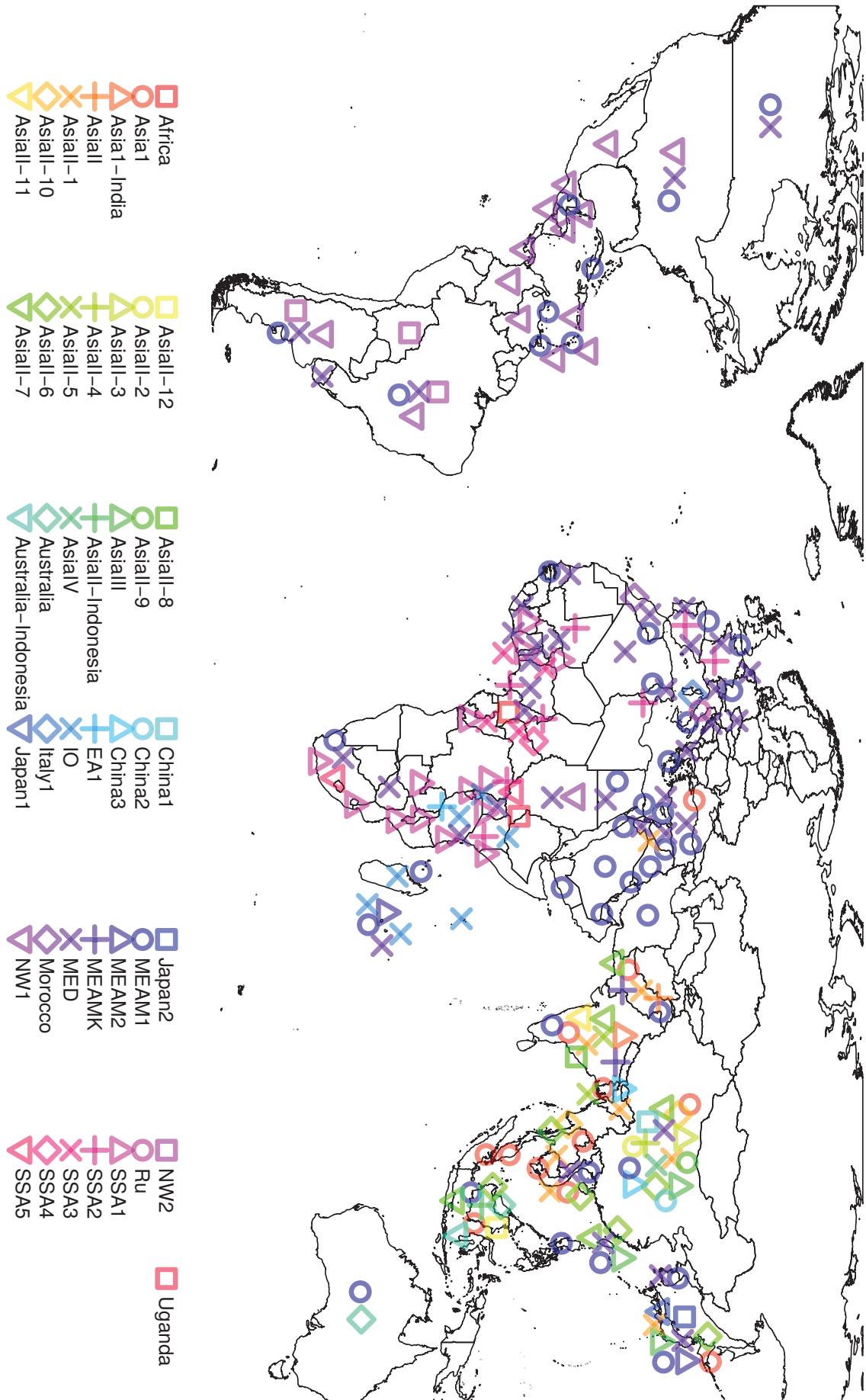
Il faut noter que les croisements contrôlés ne sont pas forcément faciles à mettre en œuvre car les groupes génétiques ne sont pas forcément retrouvés en sympatrie à cause de leur répartition géographique. De plus, lors des épisodes invasifs on assiste plutôt à des remplacements que des coexistences (Khasdan *et al.* 2005). En utilisant des analyses multivariées sur des dizaines de caractères morphologiques, diverses études ont essayé de différencier ces espèces potentielles avec plus ou moins de succès (Jayasekera *et al.* 2010 ; Thomas *et al.* 2014 ; Chaubey *et al.* 2015). En tout état de cause, ces analyses sont longues et il reste plus facile de « coller une étiquette » sur un clade grâce à une séquence, qu'à l'aide de la combinaison linéaire de plusieurs variables. Nous verrons que cette phylogénie basée sur ce seuil de 3,5 % est loin d'être parfaite et soulève de nombreuses questions. C'est pourquoi nous nous efforcerons de parler d'« espèces potentielles » pour les désigner. Néanmoins, il faut avoir à l'esprit que ce travail constitue une pre-



**FIGURE 14** – Phylogénie du complexe *B. tabaci* (*mtCOI*). Le complexe est polyphylétique puisque *B. emiliae* et *B. atriplex* sont incluses dedans, mais ce n'est pas un problème en soit puisque tous ces clades sont *a priori* des espèces différentes. La présence de ces clades émergeant au sein de *B. tabaci* a déjà été montré dans la littérature (Dinsdale *et al.* 2010 ; Firdaus *et al.* 2013 ; Lee *et al.* 2013). Nous mentionnons deux super-clades De Barro *et al.* (divergence supérieure à 11 % 2011) : Sub-Saharan Africa et Africa/Middle East/Asia Minor sur lequel nous nous focaliserons. Les symboles correspondent à la carte de droite.

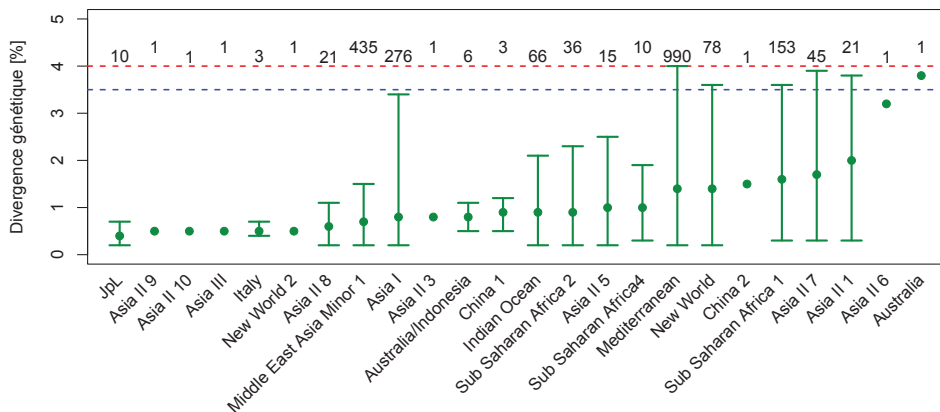
**FIGURE 15** – A droite, Distribution mondiale des espèces du complexe *B. tabaci*, chaque symbole indique que l'espèce en question a été détectée dans le pays correspondant. Attention au dégradé de couleurs qui ne reflète pas des liens phylogénétiques mais l'ordre alphabétique (voir phylogénie ci-dessus). On remarque la répartition cosmopolite des espèces potentielles MED et MEAM1 alors que d'autres ont une répartition beaucoup plus localisée comme NEW WORLD en Amérique, IO de la Réunion à l'Ouganda, ou encore l'espèce AFRICA, seulement au Cameroun. On remarque aussi une très importante diversité en Asie, 9 espèces en Inde et 14 en Chine, même s'il est vrai que ce sont de grands pays et que l'effort de recherche n'est probablement pas le même dans tous les pays.

2. Des races d'hôtes au complexe d'espèces



mière étape nécessaire pour clarifier la situation, s'accorder sur une nomenclature même si elle est provisoire, et surtout un modèle qui peut-être testé et réfuté le cas échéant. Depuis ce premier travail, de nouveaux haplotypes mitochondriaux divergeant de plus de 3,5 % de toutes les espèces potentielles précédemment identifiées ont été découverts, portant le nombre d'espèces potentielles de 24 à 41 : Hu *et al.* (2011) ont identifié 4 nouvelles espèces ; Parrella *et al.* (2012), 1 ; Alemandri *et al.* (2012), 1 ; Chowda-Reddy & Kirankumar (2012), 1 ; Lee *et al.* (2013), 1 ; Tahiri *et al.* (2013), 1 ; Legg *et al.* (2014), 1 et Firdaus *et al.* (2013), 7 nouvelles espèces ! Nous avons réalisé une phylogénie en prenant un haplotype de chacune de ces espèces potentielles (voir la figure 14, ainsi que la carte de distribution mondiale associée 15).

La variabilité mitochondriale à l'intérieur de ces espèces diffère énormément et tendrait en plus à sous estimer le nombre d'espèces total (Boykin *et al.* 2012 ; Lee *et al.* 2013). Par exemple la diversité intra chez les 6 espèces suivante dépasse en réalité les fameux 3,5 %, probablement car de nouveaux haplotypes émergeant au sein de ces groupes ont été décrits depuis : ASIA-II-1, ASIA-II-7, AUSTRALIA, MEDITERRANEAN, NEW WORLD, et SUB SAHARAN AFRICA 1. Par contre sept espèces ne sont représentées que par un seul haplotype : ASIA I-INDIA, ASIA-II-2, ASIA-II-4, CHINA3, MIDDLE EAST ASIA MINOR 2, SUB SAHARAN AFRICA 3, et UGANDA. Cette situation est illustrée sur la figure 16.



**FIGURE 16** – Divergences intraspécifiques observées sur les paires de séquences *mtCOI* (650 paires de bases), dans les espèces du complexe comprenant au moins deux haplotypes. Le point représente la moyenne, les deux extrémités le minimum et le maximum observé. Les nombres au dessus indiquent le nombre de paires comparées. Adapté de Lee *et al.* (2013).

Mais surtout, le seuil de 3,5 % qui est utilisé n'a pas forcément de réalité biologique : bien qu'on voit nettement la rupture dans la distribution des distances de toutes les paires d'haplotypes du complexe (voir figure 13), cela n'exclue pas la possibilité que dans certaines espèces potentielles des incompatibilités existent entre des groupes dont la distance est inférieure à ce seuil, et inversement des groupes peuvent s'hybrider alors que la distance mitochondriale est supérieure à ce seuil. Par exemple on peut se poser la question de la signification biologique de l'espèce potentielle Méditerranéenne (MED) : les distances intra sont parfois supérieures à 3,5 % (voir figure 16) : elle comprend plusieurs groupes d'haplotypes mitochondriaux, historiquement

identifiés comme biotypes, d'origines géographiques assez différentes.

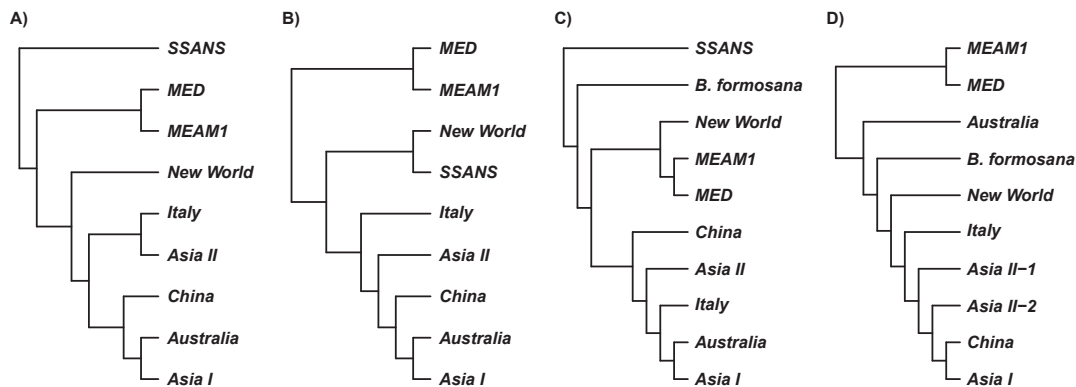
Pour illustrer la difficulté de se servir de la seule distance mitochondriale comme indice des barrières reproductives, prenons le cas de l'espèce MEAM 1 qui a envahi la plupart des régions du monde donnant lieu à des contacts secondaires avec d'autres espèces. Il a été montré que l'hybridation est possible avec l'espèce AUSTRALIA (Gunning *et al.* 1997) et dans un sens avec l'espèce INDIAN OCEAN (Delatte *et al.* 2005 ; Thierry *et al.* 2011), sur la base de marqueurs génétiques, respectivement des isozymes et des microsatellites, mais que l'isolement reproducteur était complet avec l'espèce NEW WORLD (Costa *et al.* 1993 ; Perring *et al.* 1993) ou l'espèce MED (Moya *et al.* 2001) (marqueurs RAPD). Si on regarde les distances mitochondriales, la plus proche de MEAM 1 est pourtant MED (c'est même la plus proche de tout le complexe) et la distance entre MEAM 1 et INDIAN OCEAN est de 7 % alors que l'hybridation reste possible. Pour rajouter à la complexité de la situation, de très rares hybrides ont été détectés entre MEAM 1 et MED (Tahiri *et al.* 2013 ; McKenzie *et al.* 2012) en populations naturelles, et des croisements expérimentaux ont permis d'obtenir des hybrides (rares, De Barro & Hart 2000 ; Sun *et al.* 2011), ce qui suggère encore une fois une certaine variabilité génétique dans ces espèces potentielles. A l'aide des génomes complets du symbiote primaire *Portiera* et de gènes mitochondriaux, Santos-Garcia *et al.* (2015) ont estimé la divergence entre les espèces potentielles MED et MEAM 1 à seulement quelques centaines de milliers d'années, (estimation compatible mais néanmoins inférieure à celle de Hsieh *et al.* (2014) qui l'estiment à 2,8 millions d'années) suggérant ainsi que le super-clade Africa/Middle East/Asia Minor (voir la phylogénie page 44) serait au tout début du processus de spéciation. Ceci expliquerait les résultats légèrement contradictoires précédant, à cause de barrières reproductives incomplètes.

L'utilisation de marqueurs mitochondriaux est problématique car ceux-ci sont soumis à plusieurs phénomènes biologiques compliquant leur interprétation. Tout d'abord, la mitochondrie reflète les processus de la lignée maternelle plutôt que ceux de la population dans son ensemble, ce qui peut être problématique si les deux sexes ne suivent pas les mêmes trajectoires. De plus, l'évolution des marqueurs mitochondriaux n'est pas neutre. Ils peuvent être soumis à sélection soit de manière directe (sur les gènes mitochondriaux eux-mêmes), soit à sélection indirecte par le biais d'autres éléments transmis maternellement.

L'introgression adaptative peut aussi amener à confusion. Par exemple, les espèces sœurs de drosophiles *D. yakuba* et *D. santomea* ont divergé il y a environ 400 000 ans mais peuvent s'hybrider à hauteur de 3 % dans la nature. Malgré cette divergence, plusieurs événements d'introgression mitochondriale ont été détectés, dont un qui a entièrement remplacé le génome d'une espèce par celui de l'autre il y a environ 14 000 ans. Les auteurs supposent qu'un des événements d'introgression est adaptatif et aurait permis de remplacer le génome de *D. santomea* qui aurait accumulé plus de mutations délétères que celui de *D. yakuba* dont la taille de population est plus élevée (Llopart *et al.* 2014). Rieseberg (2009) a proposé l'introgression comme moyen de « remplacer » des allèles délétères ou endommagés par ceux d'une espèce proche.

D'autres phénomènes peuvent aussi créer des biais lors de l'utilisation des seuls marqueurs

mitochondriaux : l'asymétrie liée au sexe, le tri incomplet des lignées, ou encore la sélection indirecte par la présence de symbiote en déséquilibre de liaison (Hurst & Jiggins 2005). Le transfert horizontal des bactéries capables d'envahir des populations d'hôtes peut ainsi entraîner l'invasion d'un haplotype et l'étude de celui-ci révèle plutôt l'histoire de l'infection que la phylogénie du groupe. Il y a de nombreux exemples de discordances dans la littérature entre les marqueurs nucléaires et les marqueurs mitochondriaux et notamment dans le cas de radiation récente comme chez les criquets du genre *Laupala* dans les îles Hawaïennes (Shaw 2002), pour une revue voir Toews & Brelsford (2012). Ce problème est visible chez *B. tabaci*, où la phylogénie de Hsieh *et al.* (2014) montre une incongruence, premièrement entre les phylogénies *mtCOI* de différents auteurs, mais en plus dans leur étude, entre une phylogénie construite sur le *mtCOI* et l'autre grâce à trois gènes nucléaires (voir figure 17).



**FIGURE 17** – Différentes phylogénies de *B. tabaci*, A : *mtCOI*, Boykin *et al.* (2007) ; B : *mtCOI*, Dinsdale *et al.* (2010) ; C : *mtCOI*, Hsieh *et al.* (2014) ; D : trois gènes nucléaires (*shaw*, *RNApyII*, *prp8*), Hsieh *et al.* (2014).

## Les associations symbiotiques chez *Bemisia tabaci*

### Le symbiote primaire

Chez les aleurodes, *Portiera aleyrodidarum* (Thao & Baumann 2004a) est l'endosymbiote obligatoire (ou primaire) car l'insecte ne peut vivre sans (et réciproquement). Il appartient à la famille des *Halomonadaceae* ( $\gamma$ -protéobactérie) et forme avec le symbiote primaire des psylles (*Carsonella ruddii*) et celui des *Coleorrhyncha* (*Evansia muelleri*), un clade phylogénétique exclusivement composé d'endosymbiotes d'hémiptères (Santos-Garcia *et al.* 2014). Ces bactéries sont situées dans les bactériocytes, des cellules spécialisées, formant une paire d'organes appelés bactériomes, de couleur orangée et facilement distinguable sur les pupes (voir figure 18).

*P. aleyrodidarum* a un génome très réduit (357 472 paires de bases) avec un taux en GC de 26 %, valeurs expliquées par le processus de réduction de son génome, qui a lieu chez tous les



symbiotes primaires (Moran *et al.* 2008), comme nous l'avons vu en introduction. Cette bactérie permet la synthèse des acides aminés essentiels suivants : thréonine, tryptophane et sérine (Sloan & Moran 2012 ; Santos-Garcia *et al.* 2012 ; Rao *et al.* 2015), en tout cas son génome contient les gènes de la voie de synthèse. Les gènes codant pour la synthèse de caroténoïdes sont aussi présents dans son génome permettant probablement de fournir ces molécules à *B. tabaci*. Les caroténoïdes sont très importants car ils sont sources de pigments et ont une action antioxydante. Les génomes de *Portiera* provenant des espèces potentielles MED et MEAM 1 sont très similaires, avec un contenu en gène identique et un taux de divergence de seulement 0,7 % (Moran *et al.* 2008 ; Santos-Garcia *et al.* 2012).



**FIGURE 18** – Pupa de *B. tabaci* avec les deux bactériomes visibles, en jaune plus foncé.

## Les symbiotes secondaires

### Généralités

Dans la partie précédente nous avons vu que *B. tabaci* est un complexe d'espèces potentielles, car définies d'après le gène mitochondrial *COI*, avec un certain manque de données biologiques telles que des croisements contrôlés ou des données génétiques nucléaires des populations naturelles. Nous avons vu que certaines de ces espèces potentielles montrent une grande variabilité *intra* avec de nombreux haplotypes, parfois fortement différenciés. Cette variabilité mitochondriale se retrouve dans les associations symbiotiques avec lesquelles ces haplotypes sont associés. En effet, en plus du symbiote primaire *Portiera* que nous avons évoqué précédemment, 7 autres bactéries secondaires (ou facultatives) sont retrouvées chez *B. tabaci*. Elles appartiennent à 5 ordres différents et sont transmises de façon maternelle (ou verticalement), mais aussi horizontalement (beaucoup plus rarement) pour certaines (*Arsenophonus*, *Rickettsia* et *Wolbachia*). Leur prévalence est importante puisque 95% des individus sont infectés par au moins un symbiote secondaire et 45 % par au moins deux symbiotes secondaires (Gueguen *et al.* 2010), on parle alors de multi-infections (sous entendu multiples infections **par des symbiotes secondaires**). Ces bactéries sont localisées dans les bactériocytes (avec le symbiote primaire *Portiera*), mais aussi dans d'autres tissus pour certaines, ce que nous allons détailler ensuite. Les bactéries impliquées sont les suivantes :

- *Arsenophonus* sp. (Enterobacterales, Zchori-Fein & Brown 2002)



- *Cardinium hertigii* (Bacteroidales, Weeks *et al.* 2003)
- *Hamiltonella defensa* (Enterobacteriales, Zchori-Fein & Brown 2002)
- *Rickettsia sp.* (Rickettsiales, Gottlieb *et al.* 2006)
- *Wolbachia sp.* (Rickettsiales, Jeyaprakash & Hoy 2000 ; Zchori-Fein & Brown 2002 ; Nirgianaki *et al.* 2003)
- *Hemipteriphilus asiaticus* (Rickettsiales, Bing *et al.* 2013a,b)
- *Fritschea bemisiae* (Chlamydiales, Thao *et al.* 2003 ; Everett *et al.* 2005)

Parmi ces bactéries, les deux dernières seront très peu évoquées pour les raisons qui suivent. *Fritschea* a été très peu détectée : seulement deux populations dans deux études, et sans données de fréquence. *Hemipteriphilus* n'a été décrite que très récemment et seulement sur deux espèces asiatiques (ASIA-II-7 et CHINA1), ce qui n'empêche pas qu'elle semble plutôt fréquente dans ces espèces (de 46 à 78 % des individus infectés dans six populations de l'espèce CHINA1 (Bing *et al.* 2013b).

Un des intérêts du modèle *B. tabaci* repose sur le fait que les multi-infections sont très courantes et surtout que des combinaisons de bactéries sont spécifiques de certains haplogroupes<sup>1</sup> mitochondriaux (Zchori-Fein & Brown 2002 ; Gueguen *et al.* 2010 ; Bing *et al.* 2013a ; Gnankiné *et al.* 2013a)<sup>2</sup>. Pour un haplogroupe mitochondrial donné, on ne pourra retrouver en association que certaines bactéries, par exemple, *Hamiltonella* ou *Wolbachia* associées à l'haplogroupe Q1 (qui appartient à l'espèce potentielle MED). Cette spécificité<sup>3</sup> est illustrée sur la figure 19b, qu'on appelle mondrian<sup>4</sup>. On peut observer cette spécificité d'association à deux échelles (au niveau spécifique ou inférieur) : par exemple l'espèce potentielle MEAM 1 peut héberger seulement *Hamiltonella* ou *Rickettsia*, mais les différents haplogroupes mitochondriaux de l'espèce potentielle MED hébergent des combinaisons de bactéries différentes (voir la figure 19b). Cette association de bactéries spécifique d'un haplogroupe mitochondrial sera appelée cytotype.

Le fait que certaines bactéries soient retrouvées avec certains haplogroupes n'implique pas que toutes ces bactéries soient retrouvées en même temps chez tous les individus, les prévalences sont variables entre populations. Si on reprend l'exemple de Q1, certaines populations ne présenteront pas *Wolbachia*, et d'autres n'auront pas *Cardinium* (Chiel *et al.* 2007 ; Chu *et al.* 2007 ; Gueguen *et al.* 2010) ; le cytotype est donc l'association entre un haplogroupe mitochondrial donné et des bactéries potentielles (celles-qui ont été identifiées jusqu'à présent). La figure 20 montre différents cas de figure de prévalence des bactéries du cytotype Q1.

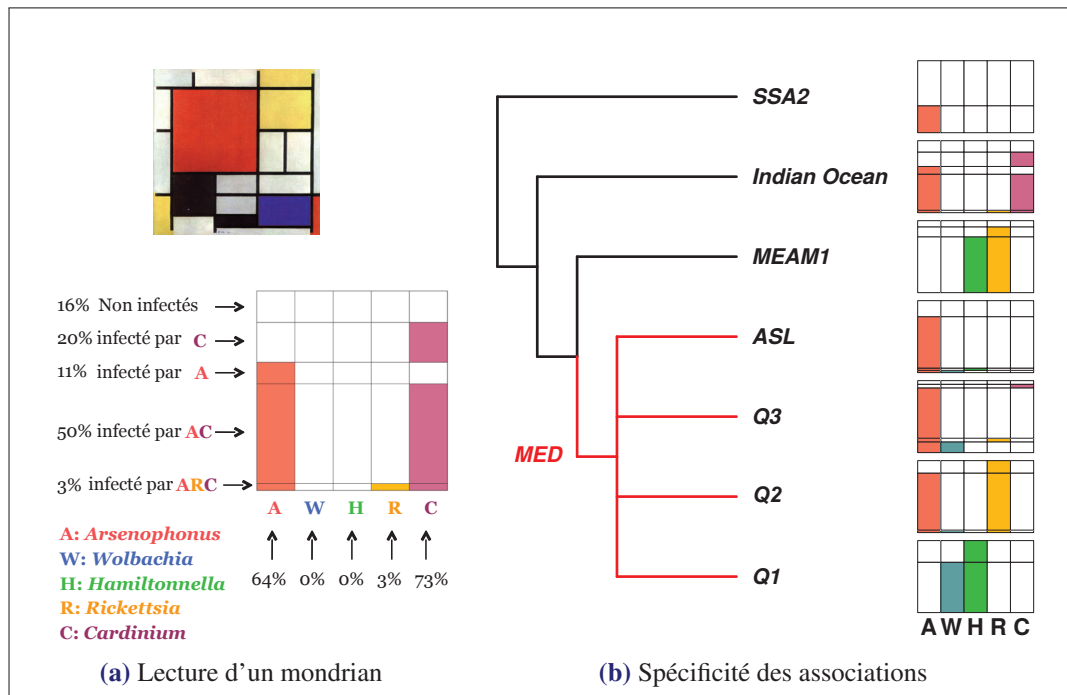
Ces cytotypes étant définis par l'association d'une communauté bactérienne avec un haplo-

1. Un haplogroupe est un ensemble d'haplotypes formant un groupe monophylétique.

2. Gnankiné O., Mouton L., Henri H., Terraz G., Houndeté T., Martin T., Vavre F., Fleury F. (2013). *Distribution of Bemisia tabaci biotypes (Homoptera : Aleyrodidae) and their associated symbiotic bacteria on host plants in western Africa. Insect Conservation and Diversity*, 6(3), 411–421.

3. Spécificité est utilisée ici dans le sens de caractère qui lui est propre et ne se réfère pas forcément au rang taxonomique de l'espèce.

4. Pieter Cornelis Mondriaan est un peintre néerlandais reconnu comme un des pionniers de l'abstraction. Ses peintures ont inspiré les représentations graphiques que nous utilisons pour décrire les prévalences des bactéries (voir figure 19a)



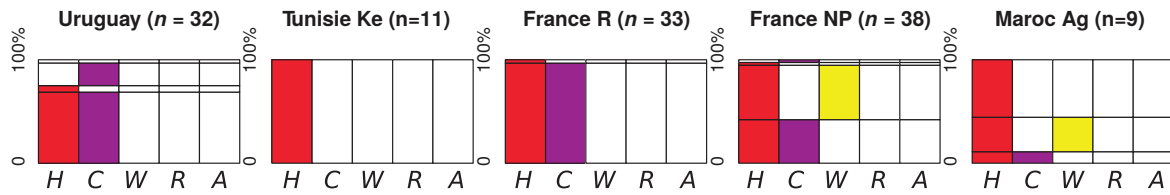
**FIGURE 19** – Illustration de la spécificité des associations. La figure (a) est un exemple type d’illustration utilisée (en bas), avec une peinture originale de Mondrian au dessus (*Composition en rouge, jaune, bleu et noir*, 1926), pour représenter ces multi-infections. En colonne sont représentées les bactéries, en ligne on peut lire les associations avec leur prévalence. H : *Hamiltonella*, C : *Cardinium*, W : *Wolbachia*, R : *Rickettsia*, A : *Arsenophonus* La figure (b) est un extrait de la phylogénie basée sur le *mtCOI* de *B. tabaci*, montrant seulement quatre espèces potentielles du complexe. L’espèce potentielle méditerranéenne (MED) est représentée en rouge par 4 haplogroupes. À droite des noms d’espèces sont représentées les fréquences de multi-infections.

groupe mitochondrial complique un peu la nomenclature. Par exemple, en parlant du cytotype Q1 on désigne l’haplogroupe mitochondrial et l’ensemble des bactéries potentielles associées, mais on peut vouloir désigner juste l’haplogroupe mitochondrial Q1 (le mytotype Q1), sans les bactéries ; on s’efforcera donc de le préciser dans le texte pour éviter les confusions.

Nous allons maintenant décrire les caractéristiques connues de ces différents symbiotes, (taxonomie, fréquence, effets phénotypiques...) qui sont, rappelons le, tous à transmission verticale maternelle, sans que cela n’exclue de possibles transferts horizontaux. Pour chaque bactérie, nous commencerons par décrire son incidence et ses effets phénotypiques chez les arthropodes, puis nous détaillerons ce qui concerne *B. tabaci*.

### Description des partenaires symbiotiques secondaires

*Arsenophonus* est un genre appartenant à la classe des  $\gamma$ -protéobactérie (Famille des Enterobacteriaceae) qui infecte plus d’une centaine d’hôtes dont de nombreux arthropodes (aleurodes, fourmis, tiques...) et même des plantes (Nováková *et al.* 2009). On sait que cette bactérie peut être transmise de façon horizontale (en plus de la transmission verticale), notamment chez les aleurodes (Thao & Baumann 2004b ; Mouton *et al.* 2012). Les phénotypes induits sont très



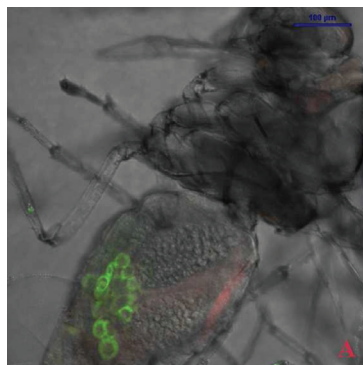
**FIGURE 20** – Illustration de la variabilité des associations chez le cytotype Q1 à travers cinq populations provenant de quatre pays. *H* : *Hamiltonella*, *C* : *Cardinium*, *W* : *Wolbachia*, *R* : *Rickettsia*, *A* : *Arsenophonus*. *Hamiltonella* est retrouvée dans toutes les populations, contrairement à *Cardinium* et *Wolbachia* qui sont retrouvées dans respectivement quatre et deux populations. *Rickettsia* et *Arsenophonus* ne sont jamais retrouvés en associations avec le Q1. La figure 19 détaille la lecture et l'interprétation de ce type de figure.

divers et vont de la manipulation de la reproduction, en l'occurrence le *male-killing* (Ghera *et al.* 1991) ou le mutualisme (Allen *et al.* 2007 ; Perotti *et al.* 2007).

Chez *B. tabaci*, les phénotypes associés sont mal connus : seulement deux articles existent dans la littérature et concernent l'espèce ASIA-II-1. Une étude suggère qu'*Arsenophonus* interagit avec le virus phytopathogène CLCuV et joue un rôle dans sa transmission (Rana *et al.* 2012). L'autre article semble montrer que l'absence de la bactérie (supprimée par traitement antibiotique) augmente la valeur sélective (fécondité, longévité, survie) de l'insecte (Raina *et al.* 2015). Sa localisation dans le corps de *B. tabaci* n'est pas très bien connue, une première étude l'avait localisée uniquement dans les bactériocytes, à tous les stades de développement (Gottlieb *et al.* 2008) mais des travaux plus récents l'ont localisée également dans l'intestin ainsi que dans les glandes salivaires (Rana *et al.* 2012), voir la figure 21.

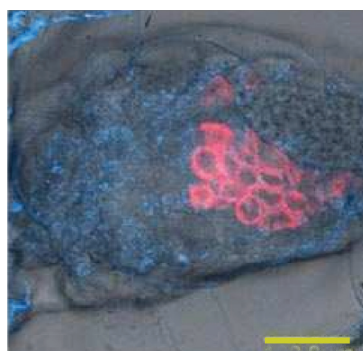
Il y a plusieurs souches d'*Arsenophonus* (au moins sept) présentes chez *B. tabaci* (Gueguen *et al.* 2010 ; Mouton *et al.* 2012) qui semblent transférer assez facilement dans le complexe d'espèce ou vers le genre *Trialeurodes* — autre aleurode à l'écologie similaire à *B. tabaci* — (Mouton *et al.* 2012), mais aussi vers des hyménoptères et des diptères. Cette capacité à transférer explique probablement pourquoi on la retrouve dans de nombreuses espèces potentielles du complexe d'espèces *B. tabaci* (10) : ASIA 1, ASIA-II-1, ASIA-II-3, ASIA-II-6, ASIA-II-7, IO, MED, SSA 1, SSA 3 et EA 1 (Ahmed *et al.* 2010, 2013 ; Bing *et al.* 2013a ; Fujiwara *et al.* 2015 ; Ghosh *et al.* 2015 ; Gnankiné *et al.* 2013a ; Gueguen *et al.* 2010 ; Jing *et al.* 2014 ; Raina *et al.* 2015 ; Singh *et al.* 2012 ; Tajebe *et al.* 2014 ; Thierry *et al.* 2011).

**Cardinium** est un genre appartenant à la classe des Bacteroidetes (Famille des Bacteroidaceae) relativement courant chez les arthropodes (aleurodes, cochenilles diaspines, cicadelles, araignées...) dont l'incidence chez les arthropodes se situerait autour de 6 % (Zchori-Fein & Perlman 2004). En ce qui concerne les phénotypes induits, on sait que *Cardinium* manipule la reproduction par parthénogénèse (Zchori-Fein *et al.* 2001) ou par incompatibilité cytoplasmique (Hunter *et al.* 2003) chez la guêpe parasitoïde *Encarsia pergandiella* ; mais aucun de ces effets n'ont été détectés chez *B. tabaci*.



**FIGURE 21** – Localisation d'*Arsenophonus* chez une femelle *B. tabaci* adulte. Photo prise au microscope électronique à transmission et coloré par FISH avec des sondes vertes pour *Arsenophonus* et rouge pour *Portiera*. *Arsenophonus* est localisé dans l'intestin, les glandes salivaires et les bactériocytes. Illustration tirée de Rana *et al.* (2012). Le trait mesure 100  $\mu\text{m}$ .

Trois souches ont été détectées chez *B. tabaci* (Gueguen *et al.* 2010). *Cardinium* est localisée dans les bactériocytes, et dans l'abdomen (voir figure 22), et peut-être dans les adipocytes (Gottlieb *et al.* 2008). Bien qu'il n'y ait à ce jour aucune étude sur de potentiels transferts horizontaux, on sait qu'elle est présente dans au moins 12 espèces potentielles du complexe : ASIA 1, ASIA-II-1, ASIA-II-3, ASIA-II-6, ASIA-II-7, IO, MEAM 1, MED, NEW WORLD 2, SSA 2, EA 1 et JAPAN 2 (Ahmed *et al.* 2013 ; Bing *et al.* 2013a ; Fujiwara *et al.* 2015 ; Gueguen *et al.* 2010 ; Park *et al.* 2012 ; Marubayashi *et al.* 2014 ; Pan *et al.* 2012 ; Singh *et al.* 2012 ; Tajebe *et al.* 2014 ; Thierry *et al.* 2011).



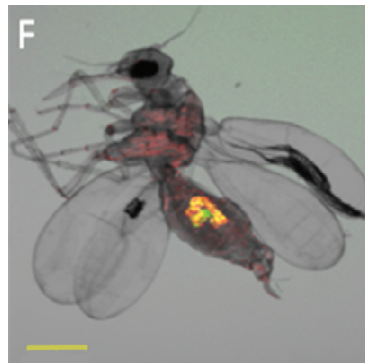
**FIGURE 22** – Localisation de *Cardinium* chez une femelle *B. tabaci* adulte. Photo prise au microscope électronique à transmission et colorée par FISH avec des sondes bleues pour *Cardinium* et rouge pour *Portiera*. Illustration tirée de Gottlieb *et al.* (2008). Le trait jaune mesure 100  $\mu\text{m}$ .

*Hamiltonella* est un genre appartenant à la classe des  $\gamma$ -protéobactéries (Famille des Enterobacteriaceae), trouvée seulement chez les aleurodes et les pucerons. Elle joue un rôle de protection chez le puceron contre les guêpes parasitoïdes (Oliver *et al.* 2003) ou contre le stress thermique, toujours chez le puceron (Russell & Moran 2006).

Chez *B. tabaci*, Gottlieb *et al.* (2010) ont montré que la transmission du virus TYLCV par l'insecte pourrait être facilitée par la présence de cette bactérie. De plus, Su *et al.* (2013b) montrent que son élimination par traitement antibiotique diminue la transmission de ce virus.

Son génome (Rao *et al.* 2012) indique qu'elle complémente les voies de synthèse de certains acides aminés essentiels comme la lysine, avec l'aide du symbiote primaire *Portiera*, amenant à une triple dépendance entre ces deux bactéries et l'hôte (Rao *et al.* 2015). Et en effet, une étude semble confirmer son rôle métabolique, notamment en période de stress nutritionnel (Su *et al.* 2014b). D'autres expériences en laboratoire semblent montrer son effet bénéfique sur la valeur sélective de *B. tabaci* : augmentation de la survie, de la fécondité, de la masse corporelle et diminution du temps de développement (Su *et al.* 2013a).

Chez *B. tabaci*, sa localisation tissulaire est restreinte aux bactériocytes (Gottlieb *et al.* 2008 ; Skaljic *et al.* 2010) ; voir la figure 23. Elle est présente chez six espèces potentielles du complexe, et notamment chez MEAM1 et MED chez qui elle est en très forte prévalence. On la retrouve aussi dans les espèces potentielles IO, NEW WORLD 2, SSA 1 et EA 1 (Bing *et al.* 2013a ; Chiel *et al.* 2007 ; Fujiwara *et al.* 2015 ; Gnankiné *et al.* 2013a ; Gueguen *et al.* 2010 ; Marubayashi *et al.* 2014 ; Pan *et al.* 2012 ; Tajebe *et al.* 2014).



**FIGURE 23** – Localisation d'*Hamiltonella* chez une femelle *B. tabaci* adulte. Photo prise au microscope électronique à transmission et coloré par FISH avec des sondes vertes pour *Hamiltonella*. Illustration tirée de Skaljic *et al.* (2010). Le trait jaune mesure 100  $\mu\text{m}$ .

*Rickettsia* est un clade très diversifié d' $\alpha$ -protéobactéries (famille des Rickettsiaceae), très connues pour les pathologies qu'elles provoquent chez les vertébrés (dont l'homme) comme les typhus provoqués par *Rickettsia prowazekii* ou *R. typhi*, la fièvre des Rocheuses induite par *R. rickettsii* ou encore la fièvre boutonneuse méditerranéenne induite par *R. conorii*.

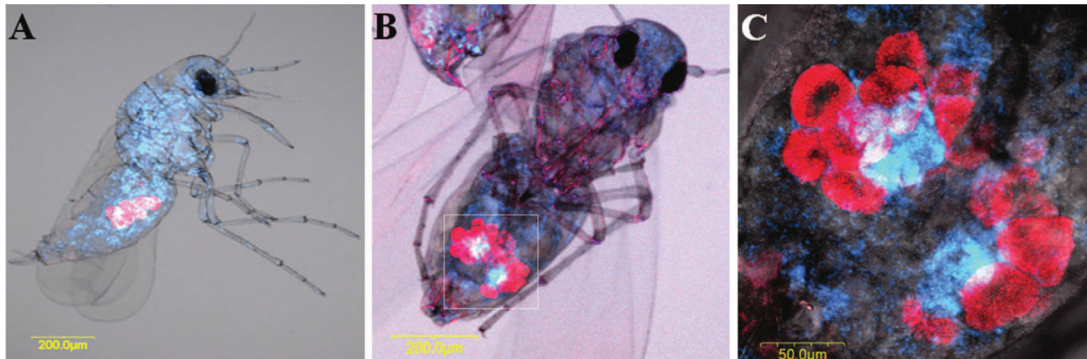
C'est surtout (pour nous) un endosymbiote très répandu chez les arthropodes (Perlman *et al.* 2006 ; Weinert *et al.* 2009). On sait qu'il peut manipuler la reproduction de ses hôtes, par exemple en induisant du *male-killing* chez la coccinelle à deux points *Adalia bipunctata* (Werren *et al.* 1994), ou la parthénogenèse chez la guêpe parasitoïde *Neochrysocharis formosa* (Hagimori *et al.* 2006). D'autres effets phénotypiques ont été répertoriés comme la tolérance au stress thermique chez le puceron du pois *Acyrtosiphon pisum* (Montllor *et al.* 2002).

La littérature commence à s'étoffer quant aux effets phénotypiques induits par *Rickettsia* chez *B. tabaci*. En 2008, Kontsedalov *et al.* (2008) ont montré que celle-ci pouvait augmenter la susceptibilité à plusieurs classes d'insecticides telles que les néonicotinoïdes, les pyridines ou



encore les acides tétroniques. Cette susceptibilité serait corrélée à la densité en *Rickettsia* dans l'hôte, plus élevée (Ghanim & Kontsedalov 2009). La densité en *Rickettsia* est liée à la température (Su *et al.* 2014a ; Brumin *et al.* 2011), ce qui augmenterait indirectement la thermotolérance (Brumin *et al.* 2011). La capacité de *B. tabaci* à transmettre le TYLCV pourrait être augmentée par la présence de *Rickettsia* (Kliot *et al.* 2014 ; Su *et al.* 2014a). Une très belle étude de Himler *et al.* (2011) a montré sa capacité à envahir des populations de l'espèce MEAM 1 en biaisant le sexe-ratio en faveur des femelles et en accroissant plusieurs paramètres liés à la valeur sélective de l'hôte (fécondité et survie) par rapport aux individus non infectés. Par contre, ce phénotype semble dépendant du génotype de l'hôte et n'est pas forcément retrouvé dans toutes les populations de l'espèce potentielle MEAM 1 (Cass *et al.* 2015a).

La localisation tissulaire de *Rickettsia* chez *B. tabaci* n'est pas encore tout à fait claire car il y aurait deux patrons possibles (Costa *et al.* 1995 ; Gottlieb *et al.* 2008 ; Brumin *et al.* 2012), soit une localisation restreinte aux bactériocytes, soit dans tout l'hémocoèle ainsi que les organes reproducteurs, digestifs, les glandes salivaires et le stylet. Néanmoins le deuxième patron de localisation semble plus courant (Caspi-Fluger *et al.* 2011 ; Su *et al.* 2014a). A ce jour, *Rickettsia* a été détectée chez neuf espèces potentielles du complexe : ASIA 1, ASIA-II-1, ASIA-II-3, IO, MEAM 1, MED, SSA 1, SSA 3 et EA 1 (Bing *et al.* 2013a ; Chiel *et al.* 2007 ; Fujiwara *et al.* 2015 ; Gnankiné *et al.* 2013a ; Gueguen *et al.* 2010 ; Marubayashi *et al.* 2014 ; Pan *et al.* 2012 ; Tajebe *et al.* 2014).



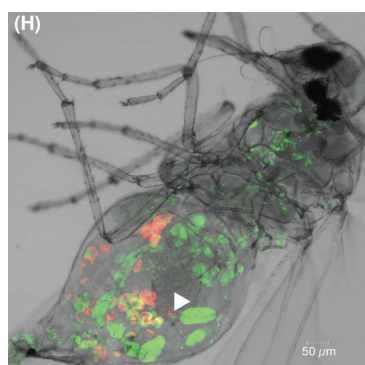
**FIGURE 24** – Localisation de *Rickettsia* chez des *B. tabaci* adultes. Photo prise au microscope électronique à transmission et coloré par FISH avec des sondes bleues pour *Rickettsia* et rouge pour *Portiera*. En A, *Rickettsia* occupe toute la cavité, en B, elle est trouvée à haute concentration près des bactériocytes. C : détails de l'image B. Illustration tirée de Brumin *et al.* (2012)

**Wolbachia** *B. tabaci* n'échappe pas à la « règle » et est infecté par *Wolbachia*, comme la moitié des arthropodes terrestres (Weinert *et al.* 2015). C'est une  $\alpha$ -protéobactérie de la famille des Rickettsiaceae.

Les effets phénotypiques induits sont peu connus chez *B. tabaci*, mais une étude suggère un rôle dans la valeur sélective et une protection contre les parasitoïdes (Xue *et al.* 2012). Plus récemment une équipe (Hu & Li 2015) a réussi à induire de l'incompatibilité cytoplasmique entre des femelles infectées naturellement par *Wolbachia* et des mâles transinfectés par une

souche de *Wolbachia* provenant de *Corcyra cephalonica*<sup>1</sup>. La *Wolbachia* de *B. tabaci* ne permet donc pas le sauvetage.

Sa localisation tissulaire est cohérente avec ces faits puisqu'elle peut-être retrouvée dans toute la cavité abdominale en plus des bactériocytes (Gottlieb *et al.* 2008 ; Bing *et al.* 2014) ou seulement dans les bactériocytes (Hu & Li 2015). Plusieurs groupes de *Wolbachia* ont été détectés chez *B. tabaci* (B et O, ce dernier étant propre à *B. tabaci*), parfois même les deux en coinfection (Bing *et al.* 2014). Ce dernier travail montre que le transfert horizontal est courant à l'intérieur du complexe, ce qui avait déjà été suggéré précédemment (Ahmed *et al.* 2013). *Wolbachia* a été détectée chez 13 espèces potentielles de *B. tabaci* pour le moment : ASIA I ASIA-II-1, ASIA-II-3, ASIA-II-6, ASIA-II-7, ASIA-II-9, AUSTRALIA, CHINA1, MEAM 1, MED, SSA 1, SSA 3, EA 1 et JAPAN 2 (Ahmed *et al.* 2010, 2013 ; Bing *et al.* 2013a, 2014 ; Fujiwara *et al.* 2015 ; Ghosh *et al.* 2015 ; Gnankiné *et al.* 2013a ; Gueguen *et al.* 2010 ; Jing *et al.* 2014 ; Pan *et al.* 2012 ; Raina *et al.* 2015 ; Singh *et al.* 2012 ; Skaljic *et al.* 2010 ; Tajebe *et al.* 2014 ; Tsagkarakou *et al.* 2012).



**FIGURE 25** – Localisation de *Wolbachia* chez une femelle *B. tabaci* adulte. Photo prise au microscope électronique à transmission et coloré par FISH avec des sondes vertes pour *Wolbachia* et rouge pour *Portiera*. Illustration tirée de Bing *et al.* (2014).

## Matériel et méthodes général

Pour finir cette partie introductive sur le modèle d'étude nous allons décrire rapidement l'élevage de l'insecte au laboratoire et la détermination des cytotypes. Ceux-ci se déterminent en deux temps : les bactéries par simple PCR spécifique, et l'haplogroupe mitochondrial par PCR-RFLP dont la mise au point a été publiée dans Henri *et al.* (2013), article qui est en annexe page 195<sup>2</sup>.

1. *Corcyra cephalonica* est un insecte de l'ordre des lépidoptères, de la famille des Pyralidae.

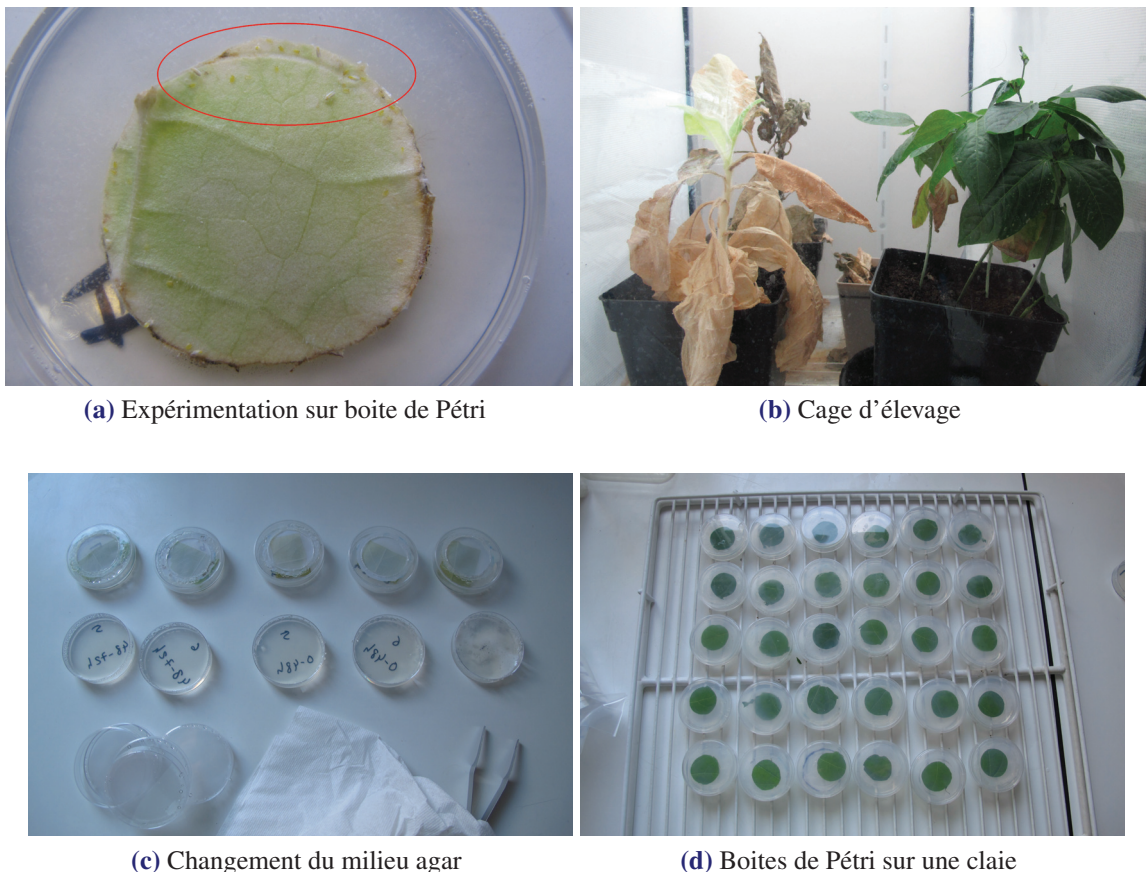
2. Henri, H., Terraz, G., Gnankiné, O., Fleury, F., & Mouton, L. (2013). *Molecular characterization of genetic diversity within the Africa/Middle East/Asia Minor and Sub-Saharan African groups of the Bemisia tabaci species complex*. *International Journal of Pest Management*, 59(4), 329–338.



## Élevage des insectes

L'élevage des insectes se fait en cages à populations (35 X 35 X 60 cm) sur des Hibiscus la plupart du temps mais également sur d'autres plantes (Coton, Solanum...) lors d'expérimentations qui le nécessite, ce qui sera alors précisé. L'élevage est maintenu à 25 °C avec un cycle de 16 h de jour pour 8 h de nuit.

Lorsque des expériences nécessitent d'être contrôlées plus finement qu'avec une simple cage à populations, notamment le nombre d'insectes, la température, ou alors qu'elles nécessitent une observation fine, impossible sur une plante entière, ces essais se font en boîte de Pétri. Un disque foliaire de 3,5 cm de diamètre est déposé dans une boîte de Pétri sur un milieu agar à 1 %. La boîte est fermée avec un couvercle aéré (avec du filet anti-insecte) pour éviter la condensation, et placée à l'envers (les insectes vivent sur la face abaxiale de la feuille) dans une étuve à la température et l'humidité contrôlées, tout comme le cycle jour-nuit.



**FIGURE 26** – Élevage de *B. tabaci* au laboratoire : cages et boîtes de Pétri. (a) : Disque foliaire sur gel d'agar, on distingue (difficilement) des adultes et des pupes (entourés en rouge), (b) : Plants de tabac (à gauche, légèrement défraîchi) et de poivron (à droite) dans une cage d'élevage de *B. tabaci*, (c) : Changement du milieu agar, (d) : Claie avec des boîtes de Pétri (retournées), avant d'être placée en chambre climatique.

## Détermination du cytotype

La détermination du cytotype implique l'identification de l'haplogroupe mitochondrial ainsi que les bactéries éventuellement présentes. Pour ceci on utilise généralement les femelles puisque ce sont elles qui transmettent les bactéries symbiotiques, elles sont plus grosses et généralement plus nombreuses dans les échantillons de populations naturelles.

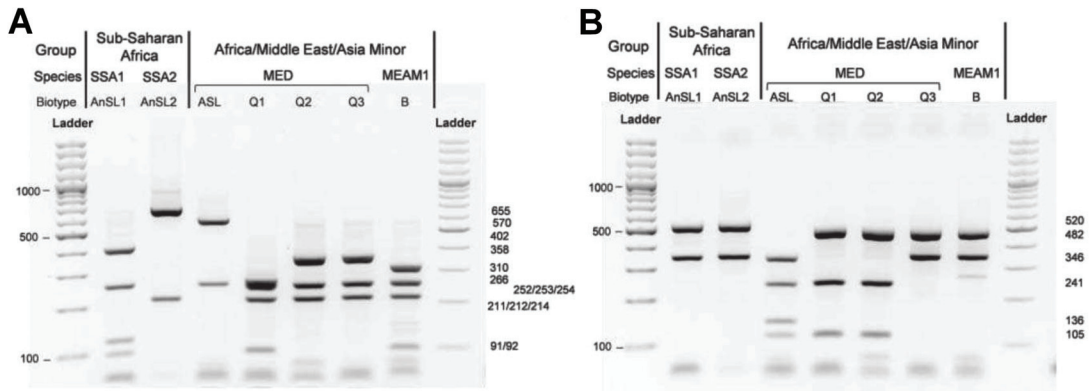
Les insectes sont isolés individuellement dans des micro-tubes pour l'extraction d'ADN. Celle-ci se fait dans 25  $\mu\text{L}$  de tampon Nonidet (50 mM KCl, 10 mM Tris-base pH8, 0,45 % Nonidet p-40, 0,45 % Tween 20) et 1  $\mu\text{L}$  de protéinase K (500  $\mu\text{g.ml}^{-1}$ ) après broyage mécanique des insectes à l'aide de billes en acier inoxydable (1,4 mm de diamètre, Carl Roth). Le lysat est incubé 1 h à 65 °C puis 15 mn à 100 °C pour inactivation de la protéinase K. 35  $\mu\text{L}$  d'eau ultra pure sont ajoutés aux échantillons, stockés à -20 °C avant utilisation.

L'haplogroupe mitochondrial est déterminé à l'aide de la technique de PCR-RFLP développée au sein du laboratoire (Henri *et al.* 2013)<sup>1</sup>. Une PCR est d'abord réalisée sur le gène mitochondrial *COI*. Le produit de PCR est digéré avec deux enzymes de restriction, dans deux réactions différentes. La combinaison des deux profils obtenus permet de discriminer les haplotypes du super-clade Africa/Middle East/Asia Minor et Sub-Saharan African. Après l'extraction, environ 800 paires de bases du gène mitochondrial *COI* sont amplifiées avec les amorces C1-J-2195 et L2-N-3014 (Simon *et al.* 1994). La réaction de PCR se fait dans un volume de 20  $\mu\text{L}$  contenant 1x de tampon, 200 nM pour chaque amorce, 200  $\mu\text{M}$  pour les dNTP, 0,5 UI de DreamTaq<sup>TM</sup> Green DNA polymerase (Fermentas), 2  $\mu\text{L}$  d'échantillon et q.s.p. eau. Le programme utilisé est : 95 °C 2 mn suivi par 35 cycles (94 °C 30 s, 50 °C 1 mn, 72 °C 1 mn) et enfin 10 mn d'élongation à 72 °C. Les produits de PCR sont digérés pour moitié avec XapI (10 U) et l'autre avec BfmI (10 U) à 37 °C pendant 3 heures. Les produits sont mis à migrer 1 h sur gel d'agarose (2 %, 100 V). La figure 27 montre les gels obtenus permettant de discriminer les haplotypes (Henri *et al.* 2013). Les quelques profils ambigus sont séquencés par la méthode de Sanger (Après une nouvelle PCR !).

Les symbiotes secondaires sont détectés par PCR à l'aide d'amorces spécifiques. Pour détecter certaines bactéries, plusieurs PCR avec différentes paires d'amorces peuvent être nécessaires pour confirmer un résultat. En tout état de chose, on ne peut jamais prouver l'absence d'une bactérie puisque l'absence de signal peut provenir de la sensibilité trop faible de la technique, notamment lorsque les bactéries sont en faible densité mais aussi d'autres artefacts techniques inhérent à toute technique expérimentale.

Les réactions de PCR se font séparément pour chaque bactérie dans un volume de 9  $\mu\text{L}$  contenant, 1x de tampon, 200 nM pour chaque amorce, 200  $\mu\text{M}$  pour les dNTP, 0,5 UI de DreamTaq<sup>TM</sup> Green DNA polymerase (Fermentas), 2  $\mu\text{L}$  d'échantillon et q.s.p. eau. Le cycle utilisé est : 94 °C 2 mn ; 94 °C 30 s, température d'hybridation spécifique des amorces (Voir tableau 1) pendant

1. Henri, H., Terraz, G., Gnankiné, O., Fleury, F., & Mouton, L. (2013). *Molecular characterization of genetic diversity within the Africa/Middle East/Asia Minor and Sub-Saharan African groups of the Bemisia tabaci species complex*. *International Journal of Pest Management*, 59(4), 329–338.



**FIGURE 27** – Profils obtenus sur gel après PCR RFLP pour discriminer les haplotypes mitochondriaux. A : XapI, B : BfmI. Figure issue de Henri *et al.* (2013)

30 s, 72 °C 1 mn (35 cycles) ; et 10 mn d'élongation à 72 °C. Le produit est mis à migrer 20 mn sur gel d'agarose 1 % à 100 V, la présence de la bande à la bonne taille indiquant la présence de la bactérie, l'inverse n'étant pas forcément vrai.

Symbiote	Gène	Nom	Séquence	Référence	Taille	Tm
<i>Rickettsia</i>	16S	Rb-F	5'-GCTCAGAACGAACGCTATC-3'	Gottlieb <i>et al.</i> (2006)	900	58
		Rb-R	5'-GAAGGAAAGCATCTCTGC-3'			
<i>Arsenophonus</i>	23S	Ars23S1	5'-CGTTTGATGAATTCATAGTCAAA-3'	Thao & Baumann (2004b)	650	60
		Ars23S2	5'-GGTCCTCCAGTTAGTGTTACCCAAC-3'			
	16S	ArsF	5'-GGGTTGTAAAGTACTTTCAGTCGT-3'	Duron <i>et al.</i> (2008)	800	52
		ArsR2	5'-GTAGCCCTRCTCGTAAGGGCC-3'			
<i>Cardinium</i>	16S	CFBF	5'-GCGGTGTAAAATGAGCGTG-3'	Weeks <i>et al.</i> (2003)	500	56
		CFBR	5'-ACCTMTTCTTAAGTCAAGCCT-3'			
	16S	CLO-f1	5'-GGAACCTTACCTGGGCTAGAATGTATT-3'	Gotoh <i>et al.</i> (2007)	466	54
		CLO-r1	5'-GCCACTGTCTTCAAGCTCTACCAAC-3'			
<i>Hamiltonella</i>	16S	H14F	5'-GCATCGAGTGAGCACAGTTT-3'	Mouton, com. pers.	243	58
		H257R	5'-TATCCTCTCAGACCCGCTAGA-3'			
<i>Wolbachia</i>	wsp*	81F	5'-TGGTCCAATAAGTGATGAAGAAAC-3'	Zhou <i>et al.</i> (1998)	550	53
		691R	5'-AAAAATTAACGCTACTCCA-3'			
	FtsZ*	F2 R2	5'-TTGCAGAGCTTGGACTTGAA-3' 5'-CATATCTCCGCCACCAGTAA-3'		400	55

**TABLE 1** – Caractéristique des amorces utilisées pour la détection des symbiotes. La longueur des fragments obtenus ainsi que la température d'hybridation sont indiqués. Les gènes 16S et 23S sont des gènes ribosomiaux. \*wsp : *wolbachia* surface protein



# Chapitre 1

## Description et dynamique spatio-temporelle des cytotypes

1	Problématique	63
2	Historique des populations de <i>B. tabaci</i>	64
2.1	En France et dans le monde, deux invasions successives par deux espèces différentes	64
3	Description des cytotypes à l'échelle du bassin méditerranéen	67
3.1	Méthode	67
3.2	Résultats	67
4	Dynamique des cytotypes en France	70
4.1	Méthode	70
4.2	Répartition des haplogroupes	71
4.2.1	Répartition spatio-temporelle	71
4.2.2	Effets de la plante hôte sur la structuration	72
4.3	Description du compartiment symbiotique	74
5	Discussion	75
5.1	Répartition géographique des haplogroupes : compétition, remplacement, coexistence ?	75
5.2	Stabilité des associations : les cytotypes	78
5.2.1	Les associations de bactéries ne sont pas recombinaisonnées	78
5.2.2	Quels mécanismes expliquent cette stabilité ?	78



## Problématique

Pour appréhender les phénomènes d'admixture, il est tout d'abord nécessaire de décrire finement la répartition des différents groupes génétiques en présence, tant dans l'aire d'étude (la France) que dans les aires potentielles d'où sont originaires les populations. Pour décrire les groupes génétiques nous allons utiliser les haplogroupes mitochondriaux puisque la taxonomie de *B. tabaci* repose principalement sur le gène mitochondrial *COI*. Nous décrirons en parallèle les bactéries symbiotiques, sources de variabilité génétique et phénotypique, qui sont associées à ces haplogroupes.

Des travaux précédents effectués au sein du laboratoire (Gueguen 2009) ont montré l'arrivée du cytotype Q2 (espèce MED) en France en 2007 où l'on retrouvait jusqu'alors principalement le cytotype Q1 (espèce MED aussi). Cette arrivée du cytotype Q2 ne s'est pas accompagnée de la disparition du cytotype Q1, les deux semblent coexister. Cette situation originale de sympatrie rend ce système particulièrement propice à l'étude de la différenciation génétique et des interactions entre communautés symbiotiques. A l'échelle de la France, on pourra suivre l'évolution des interactions (compétition, coexistence ou déplacement) entre les cytotypes Q1 et Q2 que l'on retrouve en populations mixtes, grâce à un large échantillonnage dans de nombreuses serres, réalisé sur plusieurs années. D'après la littérature (Moya *et al.* 2001 ; Khasdan *et al.* 2005 ; Horowitz *et al.* 2005), on s'attend à ce qu'un groupe déplace l'autre et que les situations de coexistence soient rares. La description des fréquences relatives de ces deux groupes génétiques permettra de déterminer si l'un des deux cytotypes va déplacer l'autre ou s'ils vont coexister. L'analyse du compartiment symbiotique déterminera l'évolution des infections lors de la confrontation des deux groupes génétiques et la possibilité d'échanges génétiques sous la forme de bactéries et potentiellement la sélection de certaines combinaisons cytoplasmiques. À une échelle plus fine, celle de la région Rhône-Alpes et particulièrement en Savoie, lieu de production intensive des poinsettias [Euphorbiaceae], culture fortement touchée par *B. tabaci*. Nous pourrions envisager de comprendre la dynamique spatio-temporelle des ravageurs dans des exploitations agricoles.

À une échelle plus large, nous faisons l'hypothèse que les populations trouvées en France sont originaires du bassin méditerranéen, aire d'origine de l'espèce potentielle MED d'après la littérature, ce que nous allons décrire plus loin. Nous allons donc décrire les différents clades que l'on retrouve autour du bassin méditerranéen grâce à un échantillonnage dans de nombreux pays. Nous pourrions ainsi obtenir des informations sur les routes d'introductions potentielles et les vecteurs de *B. tabaci* (comme le commerce de plantes ou un déplacement sans l'intervention de l'homme). C'est cette échelle qui donne des informations sur les groupes potentiels qui peuvent se trouver en contact.

Avant de passer aux travaux de thèse proprement dit nous allons décrire l'historique des populations de *B. tabaci* en France et dans le bassin méditerranéen, en faisant le parallèle avec les colonisations successives de nombreuses régions du monde, respectivement par les espèces



potentielles MEAM 1 et MED qui ont chacune colonisé plus de 40 pays (Voir la carte 15 page 44 et De Barro *et al.* 2011), invasions à l'origine de la situation retrouvée en France.

## Historique des populations de *B. tabaci*

### En France et dans le monde, deux invasions successives par deux espèces différentes

Dans ce chapitre nous allons parler des espèces potentielles MED, MEAM 1 et SSA 2 (principalement), leur position phylogénétique est visible sur la figure 14 page 44.

En Europe de l'ouest *B. tabaci* a été identifié pour la première fois en Espagne dans les années 1940 (Gomez-Menor 1943) et en France en 1989 en serre (Giustina *et al.* 1989), mais on ne sait pas *a posteriori* à quel clade ces populations appartenaient. Des populations induisant de l'argenture<sup>1</sup> ont été observées ensuite en Espagne (de façon similaire à la situation aux États-Unis), ce qui suggère l'espèce potentielle MEAM 1<sup>2</sup> connue pour déclencher ces symptômes. La présence de MEAM 1 sur la côte méditerranéenne française est confirmée en 1995 (Guirao *et al.* 1997). C'est durant cette décennie que l'espèce potentielle MEAM 1 se répand sur toute la planète et devient un problème économique majeur, notamment aux États-Unis (Brown *et al.* 1995) et en Chine (Chu *et al.* 2007). Cette extension mondiale a souvent conduit au remplacement des clades autochtones, comme aux États-Unis (Brown *et al.* 1995) et dans les Caraïbes (Bird & Brown 1998). En Australie le suivi de populations de 1995 à 2005 montre clairement le remplacement de l'espèce potentielle locale AUSTRALIA par l'espèce potentielle MEAM 1 (voir figure 1.1).

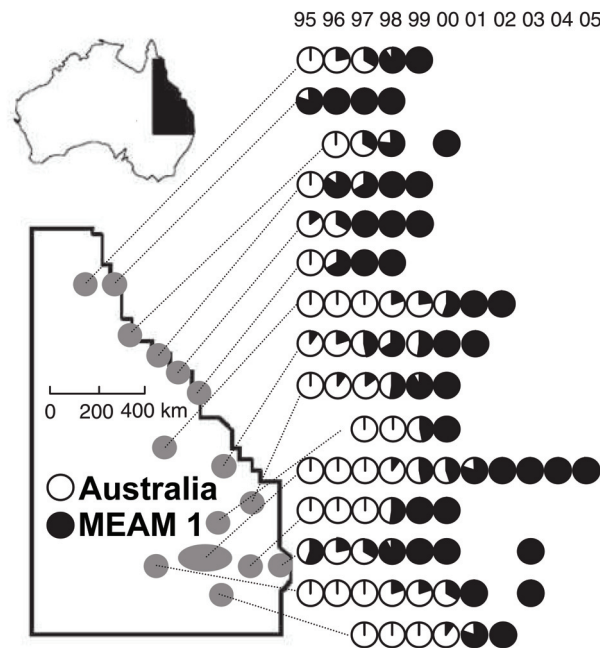
*B. tabaci* a été détecté ensuite de façon sporadique en France jusqu'au début des années 2000. Les populations ont alors fortement augmenté et causé de nombreux dégâts, notamment par la vectorisation de virus phytopathogènes, mais sans les symptômes de l'argenture, indice de la présence d'un groupe autre que MEAM 1. Cette explosion est liée à l'arrivée de l'espèce potentielle MED<sup>3</sup> en Europe de l'ouest, simultanément dans plusieurs pays : France, Espagne et Italie. Les premières études faisaient l'hypothèse que l'espèce potentielle MED était originaire de la péninsule ibérique car les premières détections de cette espèce potentielle (jusqu'en 1997) se limitaient à cette région, alors que MEAM 1 avait été détectée en Israël, France, Espagne, Italie (voir la carte du bassin méditerranéen sur la figure 1.2). Parallèlement à la diffusion de l'espèce potentielle MED, on note une très forte diminution des populations de MEAM 1 dans l'ouest méditerranéen, sa détection étant très sporadique (en 2004, en France, on le trouve seulement à Nice dans un jardin botanique, Dalmon *et al.* 2008).

---

1. Maladie physiologique provoquée par un développement anormal des cellules du limbe, qui donne aux feuilles une coloration gris-vert à argentée.

2. À l'époque on parlait de biotype B

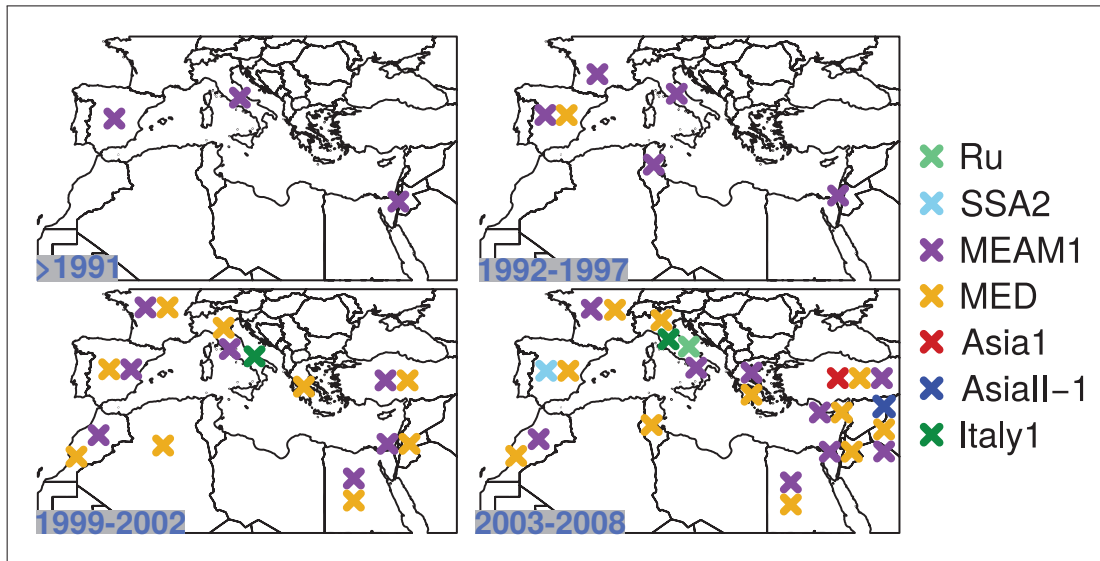
3. À cette époque on parlait du biotype Q



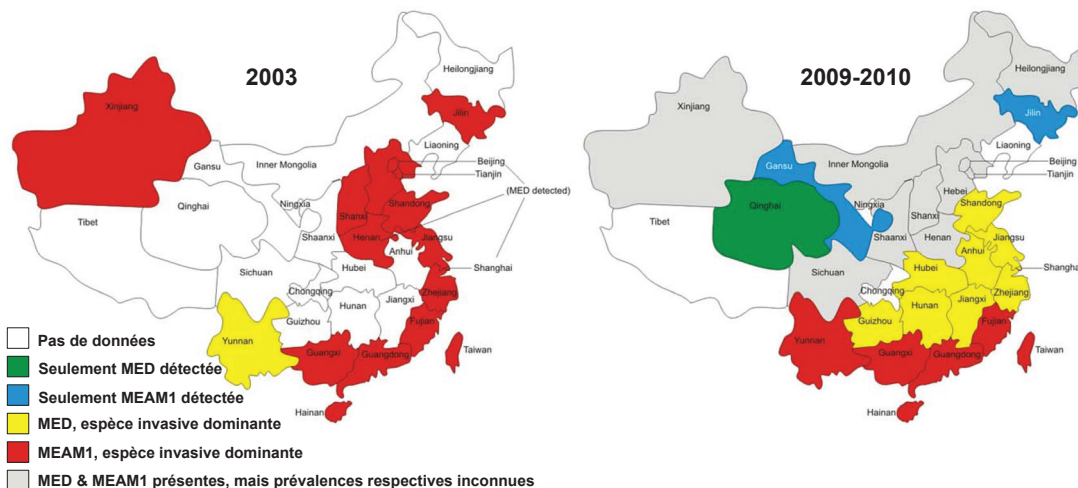
**FIGURE 1.1** – Remplacement de l'espèce AUSTRALIA par l'espèce MEAM1 en Australie. Les diagrammes représentent les proportions respectives des deux clades. Chaque ligne de diagramme représente une population suivie au cours du temps. Les diagrammes d'une même ligne verticale proviennent d'une même année d'échantillonnage, de 1995 à 2005. L'illustration provient de Liu *et al.* (2007) dont les données sont issues de Chu *et al.* (2006) ; Hsieh *et al.* (2006) ; Chu *et al.* (2007) ; Hu *et al.* (2011) et Genbank.

Une nouvelle fois, ce phénomène d'expansion démographique et géographique n'est pas restreint au bassin méditerranéen, l'espèce potentielle MED envahissant à peu près toutes les régions du monde (voir la carte 15 de détection mondiale page 44 en introduction). Ce phénomène de remplacement se fait bien souvent au détriment de l'espèce potentielle MEAM1 mais aussi des clades autochtones. La situation en Chine illustre tout à fait ce déplacement, puisqu'au début des années 2000, MEAM1 était l'espèce potentielle dominante, alors que moins de 10 ans plus tard l'espèce potentielle MED était majoritaire (voir figure 1.3). L'hypothèse dominante expliquant ce déplacement est la plus grande tolérance de l'espèce potentielle MED à de nombreux insecticides, et particulièrement les néonicotinoïdes (Hsieh *et al.* 2007 ; Dennehy *et al.* 2010 ; Luo *et al.* 2010 ; Wang *et al.* 2010), comparée à l'espèce potentielle MEAM1, plus susceptible (Horowitz *et al.* 2005 ; Ye *et al.* 2014). L'espèce potentielle MED pourrait aussi être plus tolérante aux épisodes de fortes températures (au delà de 40 °C, Xiao *et al.* 2016).

En introduction nous avons vu qu'il y a beaucoup de diversité mitochondriale dans l'espèce potentielle MED (On dénombre 5 haplogroupes : Q1, Q2, Q3, ASL, ainsi qu'un plus rare trouvé seulement en Croatie, Dinsdale *et al.* 2010). Ces haplogroupes ont des répartitions géographiques différentes : Q3 est retrouvé seulement au Burkina-Faso, ASL dans quelques pays comme la Côte d'Ivoire, le Ghana, le Burkina-Faso... Plus proche de nous, Q2 est retrouvé seulement dans l'est méditerranéen (Israël, Chypre, Liban Chiel *et al.* 2007 ; Hsieh *et al.* 2007 ; Chu *et al.* 2008) où il est le seul représentant de l'espèce potentielle MED, ainsi que quelques



**FIGURE 1.2** – Dynamique des espèces potentielles de *B. tabaci* à l'échelle du bassin méditerranéen, de 1992 à 2008. Les données proviennent de Bedford *et al.* (1994) ; Chermiti *et al.* (1997) ; Guirao *et al.* (1997) ; Demichelis *et al.* (2000) ; Elbert & Nauen (2000) ; Horowitz *et al.* (2003) ; Simón *et al.* (2003) ; Maruthi *et al.* (2004) ; Pascual & Callejas (2004) ; Brown & Idris (2005) ; Delatte *et al.* (2005) ; Horowitz *et al.* (2005) ; Khasdan *et al.* (2005) ; Hsieh *et al.* (2006) ; Rua *et al.* (2006) ; Tahiri *et al.* (2006) ; Ueda & Brown (2006) ; Fujiie *et al.* (2007) ; Simón *et al.* (2007) ; Tsagkarakou *et al.* (2007) ; Dalmon *et al.* (2008) ; Erdogan *et al.* (2008) ; Vassiliou *et al.* (2008) ; Papayiannis *et al.* (2009) ; Dinsdale *et al.* (2010) ; Gueguen *et al.* (2010) ; Skaljac *et al.* (2010) ; Vassiliou *et al.* (2011) ; Ahmed *et al.* (2012) ; Parrella *et al.* (2012) ; Tsagkarakou *et al.* (2012) ; Tahiri *et al.* (2013) ; Gauthier *et al.* (2014) ; Karut & Tok (2014) ; Karut *et al.* (2014), ainsi que d'autres non publiées (*Genbank*) : Papayiannis, Brown, Brown *et al.* 2004, Tahiri *et al.* , Gauthier *et al.*, Ahmed *et al.*, Brown, Shatters *et al.*. Les données complètes par année sont disponibles en annexe page 223.



**FIGURE 1.3** – Épisodes invasifs successifs par les clades MEAM1 et MED dans les différentes provinces chinoises. En 2003, MEAM1 est l'espèce potentielle invasive dominante tandis que MED est très peu détectée. En 2009, ces deux espèces potentielles sont détectées quasiment partout, et dans beaucoup de provinces on retrouve seulement l'espèce potentielle MED. L'illustration est extraite de Hu *et al.* (2011)

autres pays voisins, où on le retrouve en plus de Q1 (Turquie, Egypte et Syrie Rua *et al.* 2006 ; Dinsdale *et al.* 2010). L'haplogroupe Q1 est retrouvé dans une large partie ouest du bassin

### 3. Description des cytotypes à l'échelle du bassin méditerranéen

Méditerranéen : Maroc, Tunisie, Portugal, Espagne, France, Grèce (Tsagkarakou *et al.* 2007 ; Dalmon *et al.* 2008 ; Gueguen *et al.* 2010 ; Gorsane *et al.* 2011). D'ailleurs, l'analyse récente des individus utilisé par Gennadius en Grèce en 1889 pour la toute première description de *B. tabaci* révèle que ces individus appartiendraient à l'haplogroupe Q1 (Tay *et al.* 2012). Toutes les données disponibles montrent que seul cet haplogroupe est retrouvé en France et en Espagne jusqu'en 2006 (références précédentes ainsi que Simon *et al.* 1999 ; Reynaud 2000 ; Rua *et al.* 2006).

À l'échelle mondiale l'invasion par l'espèce MED est due à ce même haplogroupe Q1 qui est invasif, on le retrouve au Japon, Pays-Bas, États-Unis, Chine, Corée du Sud, Uruguay, Costa-Rica, République Tchèque, Taïwan, Guatemala, Argentine (Bethke *et al.* 2009 ; Grille *et al.* 2011 ; Guevara-Coto *et al.* 2011 ; Firdaus *et al.* 2013). Quant à l'haplogroupe Q2, en dehors des pays méditerranéens cités précédemment, il a été détecté seulement en Amérique du Nord (McKenzie *et al.* 2012). En France, jusqu'en 2007, on observait seulement l'haplogroupe Q1, l'année de la première détection de Q2.

## Description des cytotypes à l'échelle du bassin méditerranéen

### Méthode

Des individus *Bemisia tabaci* provenant de plusieurs pays du bassin méditerranéen autres que la France (que nous détaillerons ensuite) ont été échantillonnés entre 2007 et 2013 (voir tableau 1.1), pour un total de 855 individus, provenant de 22 localités différentes. Leur haplogroupe a été déterminé par PCR-RFLP, comme vu en introduction du modèle d'étude page 58<sup>1</sup>, sur le gène mitochondrial *COI* et seule une (petite) partie des individus a été séquencée. Les cinq bactéries symbiotiques les plus courantes chez *B. tabaci* (*Arsenophonus*, *Cardinium*, *Hamiltonella*, *Rickettsia* et *Wolbachia*) ont été aussi recherchées par simples PCR spécifiques chez 522 de ces 855 individus (méthode décrite à la fin de la partie présentant le modèle biologique page 58). D'autres bactéries peuvent être présentes car on ne peut les détecter avec la méthode de PCR spécifique qui nécessite un *a priori*. La bactérie *Portiera* n'a pas été recherchée car obligatoire et donc considérée comme présente. Les autres individus (333) ont été génotypés seulement sur le *mtCOI* puisque leur ADN devait être utilisé pour des analyses *RADSeq* (voir chapitre 4), leurs données symbiotiques ne sont donc pas connues à ce jour.

### Résultats

On observe une assez forte diversité à l'échelle du bassin puisque 3 espèces potentielles sont retrouvées : MED, MEAM1 et SSA2 (voir figure 1.4 et la phylogénie de la figure 1.5 page 70).

---

1. Henri, H., Terraz, G., Gnankiné, O., Fleury, F., & Mouton, L. (2013). *Molecular characterization of genetic diversity within the Africa/Middle East/Asia Minor and Sub-Saharan African groups of the Bemisia tabaci species complex*. *International Journal of Pest Management*, 59(4), 329–338.

ID	N	Ville	Pays	Année	ID	N	Ville	Pays	Année
Alx	10	Alexandrie	Egypte	2012	Hof	29	Hof-Carmel	Israël	2011
Cab	134	Cabrils	Espagne	2007	Krm	24	Kerem Maharal	Israël	2012
Vdc1	22	Viladecans	Espagne	2008	Nas	36	Nachshon	Israël	2013
Bln	3	Blanes	Espagne	2010	Tdg	25	Torre del Greco	Italie	2012
Cab2	59	Cabrils	Espagne	2010	Cas	16	Casablanca	Maroc	1999
Vdc2	117	Viladecans	Espagne	2010	Unk1	16	Unknown	Maroc	2006
Cdm	58	Cabrera de mar	Espagne	2011	Aga	11	Agadir	Maroc	2012
Elp	32	El Perello	Espagne	2012	Blf	5	Belfaa	Maroc	2012
Mcd	37	Montcada	Espagne	2012	Trx1	10	Troinex	Suisse	2011
Irp	16	Ierapetra	Grèce	2007	Trx3	23	Troinex	Suisse	2013
Irp2	10	Ierapetra	Grèce	2010	Msr	16	Monastir	Tunisie	2006
Kas	10	Kasteli-Chania	Grèce	2010	Fty	24	Fethiye	Turquie	2013
The	10	Thessaloniki	Grèce	2010	Mug	26	Mugla	Turquie	2013
Hrk	10	Heraklion	Grèce	2012	Sam	19	Samandag	Turquie	2013
Apz	26	Ampelouzozos	Grèce	2012	Sam2	21	Samandag	Turquie	2013

**TABLE 1.1** – Échantillons provenant du pourtour méditerranéen, à l’exception de la France. N : nombre d’individus génotypés.

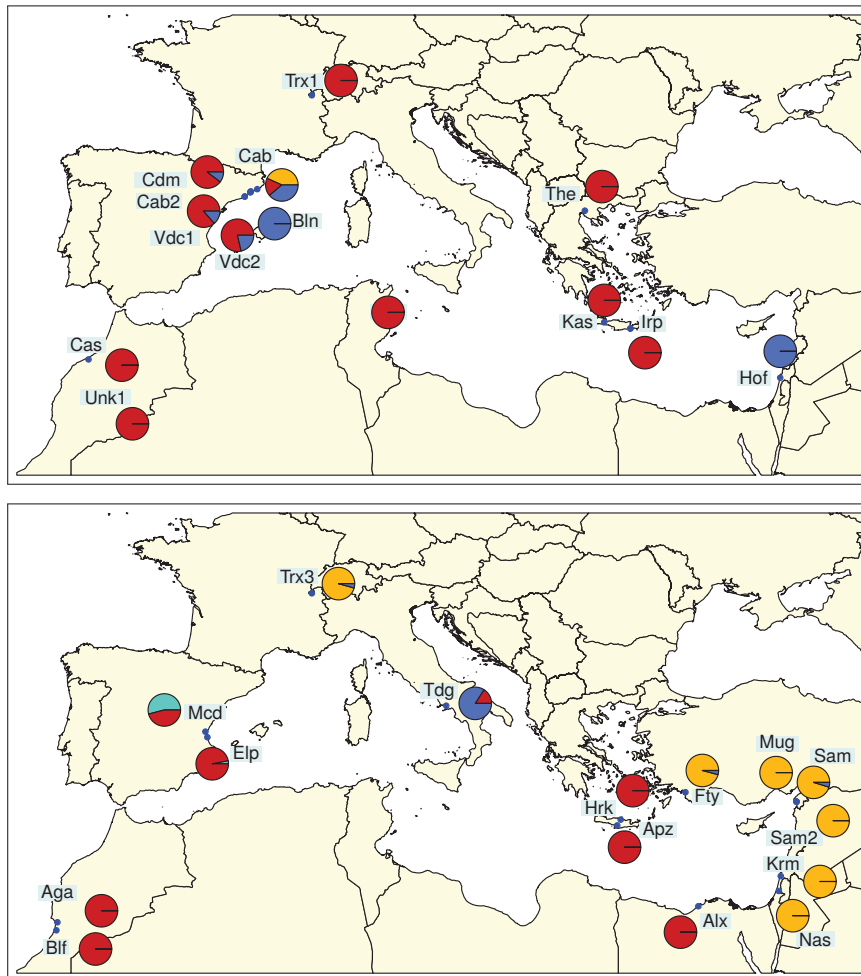
Dans plusieurs pays on trouve seulement l’haplogroupe Q1 de l’espèce potentielle MED : Grèce, Égypte, Tunisie, Maroc. L’haplogroupe Q2 est présent dans tous les autres pays non cités précédemment (Italie, Israël et Espagne). Ces données ne révèlent pas de surprises et sont conformes à celles déjà présentes dans la littérature (voir les nombreuses références de la figure 1.2 page 66). A l’est (Turquie et Israël), on retrouve l’haplogroupe Q2 de l’espèce MED ainsi que l’espèce MEAM1, avec une très forte prévalence pour la deuxième dans les dernières années (la disparition de Q2 dans ces pays sera discutée par la suite). L’espèce potentielle MEAM1 est aussi retrouvée en Suisse en 2013 dans des serres où elle n’était pas retrouvée auparavant.

Après avoir décrit les haplogroupes seulement à l’aide de PCR-RFLP, certains ont été séquencés. Ici, les données de tous les pays y compris la France ont été regroupées. A l’aide de ces séquences, un arbre phylogénétique a été construit (voir figure 1.5) pour décrire la variabilité ainsi que les liens d’apparentement (seulement sur le plan mitochondrial !) des populations à l’échelle du bassin. Il y a de nombreux éléments qui indiquent des voies d’introductions liées à l’homme plutôt que par voie naturelle. L’haplotype MEAM1 de Suisse est plus proche de ceux d’Israël qu’il ne l’est de celui de Turquie, pourtant un pays frontalier d’Israël. On retrouve des profils similaires dans l’haplogroupe Q1 et l’haplogroupe Q2 : la phylogénie ne suit pas la géographie même si certains groupes apparaissent comme cohérents. Par exemple les 3 haplotypes Q1 grecs sont monophylétiques, tout comme ceux de Tunisie. Par contre, les haplotypes Q1 provenant d’Espagne, du Maroc et de France ne forment pas des groupes monophylétiques. Un élément intéressant est l’apparente plus grande variabilité dans le groupe Q1 que dans le groupe Q2. Ce fait apparaît cohérent car les données indiquent (celles-ci ainsi que la littérature) que ce groupe était restreint aux pays de l’est méditerranéen.

Après avoir regardé les haplogroupes présents dans ces pays, les cinq bactéries symbiotiques

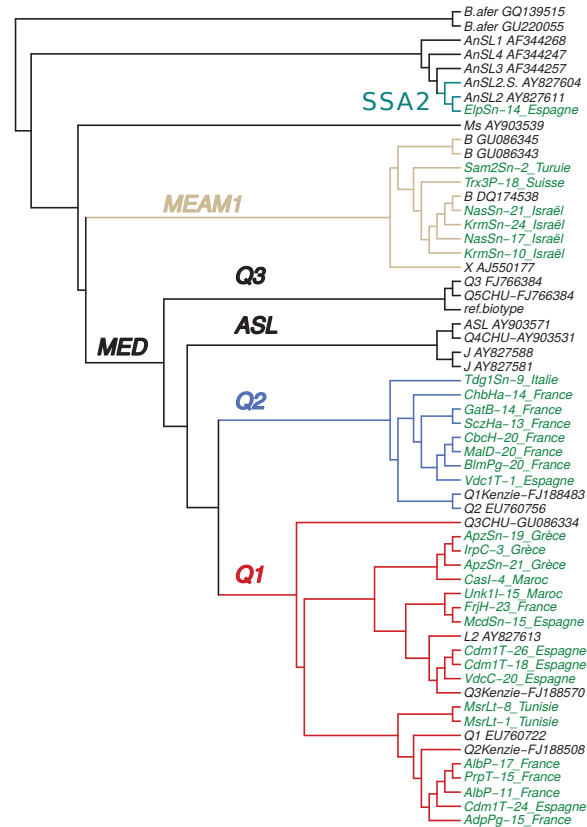


### 3. Description des cytotypes à l'échelle du bassin méditerranéen



**FIGURE 1.4** – Distribution des espèces potentielles du complexe *B. tabaci* en Méditerranée. Pour l'espèce MED, les haplogroupes Q1 et Q2 sont détaillés. En haut : données récoltées jusqu'en 2011, en bas : à partir de 2012. Pour les populations françaises, voir la figure 1.7. Jaune : MEAM 1, turquoise : SSA 2, rouge : Q1 (MED), bleu : Q2 (MED).

les plus courantes chez *B. tabaci* ont été recherchées. Les résultats sont présentés en figure 1.6. On retrouve la spécificité des cortèges symbiotiques évoqués auparavant, mais on remarque aussi l'absence complète de certaines bactéries dans certains pays. Ainsi on ne retrouve pas *Wolbachia* associée à l'haplogroupe Q1 en Egypte, en Suisse et en Espagne, *Cardinium* n'est retrouvée ni en Grèce, ni en Suisse. Par contre de façon analogue aux données de la littérature, *Hamiltonella* est à très forte prévalence dans l'haplogroupe Q1 (94 % des 222 individus), et dans l'espèce potentielle MEAM 1 (92 % des 115 individus). Associée à cette dernière on retrouve aussi *Rickettsia* avec une prévalence plus faible (38 %). Dans l'haplogroupe Q2 on trouve *Rickettsia* à des prévalences très élevées (fixée dans 7 populations, 67 et 88 % pour les deux autres), ainsi qu'*Arsenophonus*, et *Wolbachia* à des prévalences assez variées selon les populations mais plus faibles que la prévalence de *Rickettsia*. *Wolbachia* n'est jamais retrouvée dans les populations italiennes et israéliennes.



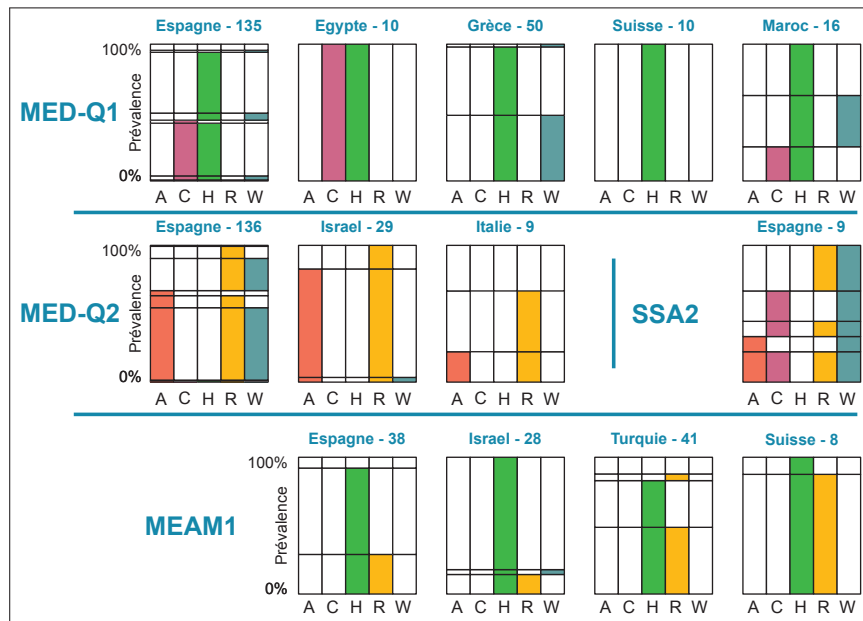
**FIGURE 1.5** – Cladogramme des séquences mitochondriales *mtCOI* des individus génotypés autour du bassin méditerranéen. Les individus en vert sont ceux de l'étude, les autres sont des individus de référence pris dans les bases de données. Les 3 espèces potentielles SSA 2, MEAM 1 et MED sont indiquées, ainsi que les quatre principaux haplogroupes MED (Q1, Q2, Q3 et ASL), les deux derniers étant retrouvés seulement en Afrique jusqu'à présent).

## Dynamique des cytotypes en France

### Méthode

Le suivi des populations a commencé au laboratoire en 2006, depuis, 1 255 individus provenant de 46 populations ont été génotypés. Une population est ici une exploitation agricole à un temps donné, ainsi une même exploitation échantillonnée à plusieurs reprises sera décomptée comme autant de populations (voir tableau 1.2). De plus, certains échantillonnages se font sur plusieurs plantes (au sens spécifique), dans ce cas l'information est précisée mais ce n'est pas décompté comme autant de populations. Pour tous ces individus, l'haplogroupe a été déterminé par PCR-RFLP et une (petite) partie de ceux-ci ont été séquencés pour vérification. Les mêmes bactéries symbiotiques que dans la partie précédente ont été recherchées, toujours par PCR, chez 838 de ces 1 255 individus. Les autres (417) ont été génotypés (*mtCOI*) dans le but d'être analysé par *RADSeq* (voir chapitre 4 page 127) et n'ont donc pas été utilisés pour la recherche de bactéries symbiotiques. Les données antérieures à 2008 sont des données acquises lors de travaux précédents (Dalmon *et al.* 2008 ; Gueguen *et al.* 2010), mais sont utilisées ici pour pouvoir discuter





**FIGURE 1.6** – Prévalences des types d’infection associées aux haplogroupes Q1 et Q2 de l’espèce potentielle MED, de MEAM1 et de SSA2 dans les pays méditerranéens à l’exception de la France. Pour chaque haplogroupe mitochondrial, les données sont divisées par pays. Le nombre d’individus testés est indiqué à côté du nom du pays. A : *Arsenophonus*, C : *Cardinium*, H : *Hamiltonella*, R : *Rickettsia*, W : *Wolbachia*. Voir la figure 19 page 51 en introduction qui explique comment lire ce type de figure.

de l’évolution de la situation.

## Répartition des haplogroupes

### Répartition spatio-temporelle

En 2007 un haplotype mitochondrial Q2 est détecté dans quatre endroits différents en France, sur les six populations analysées (Castelmaurou, Launaguet, Lavour et Lyon ; voir la figure 1.7). Cette situation est similaire à celle trouvée dans d’autres pays méditerranéen comme l’Italie et l’Espagne (Gauthier *et al.* 2014 ; Parrella *et al.* 2013 ; Terraz *et al.* 2014)<sup>1</sup>. Étant donné que la situation antérieure était assez bien connue dans ces pays, on peut supposer très fortement que ceci est dû à l’introduction récente de Q2 dans ces pays. Nous avons vu que les données sont trop incomplètes dans les autres pays du pourtour méditerranéen pour généraliser cette expansion géographique de Q2. Cette apparition soudaine ne se fait pas au détriment de l’haplogroupe Q1, car celui-ci est toujours retrouvé dans ces mêmes populations, parfois sur la même plante hôte (La même plante hôte signifie ici le même pied, au sens individuel donc). Une population mixte signifie donc que des individus sont trouvés sur les mêmes plantes. Ces situations de mixité sont retrouvées chaque année depuis, sans pour autant exclure totalement des situations où seulement l’un des deux haplogroupes est retrouvé. Le nombre de situations de mixité par

1. Terraz G., Gueguen G., Arnó J., Fleury F., Mouton L. (2014) Nuclear and cytoplasmic differentiation among Mediterranean populations of *Bemisia tabaci* : testing the biological relevance of cytotypes. *Pest management science*, 70, 1503–1513.

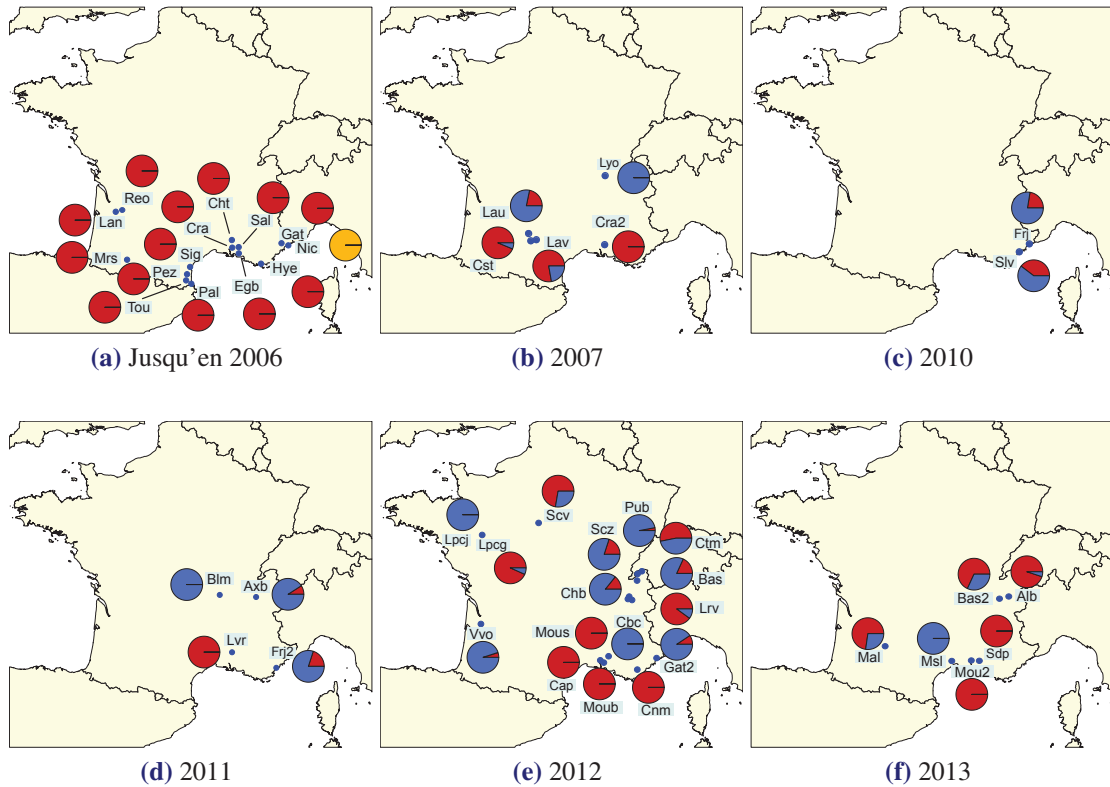
ID	N	Ville	Plante	Année	ID	N	Ville	Plante	Année
Cht	30	Chataurenard	T	2003	Frj2	10	Fréjus	H	2011
Egb	23	Etang-de-Berre	T	2003	Lvr	26	La Vernède	C, M	2011
Hye	19	Hyerès	G	2003	Bas	7	Bassens	P	2012
Lan	15	Langon	Dv	2003	Cap	1	Caphan	A	2012
Mrs	15	Marseillan	Dv	2003	Cbc	36	Cheval Blanc	H, Lt	2012
Pal	19	Palau	T	2003	Chb	30	Chambéry	Ha	2012
Pez	21	Pezilla	T	2003	Cnm	10	Cannet des Maures	R	2012
Reo	14	La reole	H	2003	Ctm	32	Contamine-sur-Arve	H, L	2012
Sal	15	Salon-de-Provence	T	2003	Gat2	32	Gattières	B	2012
Tou	24	Toulouges	C	2003	Lpcg	30	Les Ponts-de-Cé	H	2012
Gat	23	Gattieres	B	2004	Lpcj	24	Les Ponts-de-Cé	D	2012
Cra	34	St-Martin	Pv	2006	Lpcx	25	Les Ponts-de-Cé	H	2012
Nic	10	Nice	H	2006	Lrv	5	La Ravoire	P	2012
Sig	15	Sigean	T	2006	Moub	34	Moulès	A	2012
Cra2	63	St-Martin	A, Pv, T	2007	Mous	10	Moulès	A	2012
Cst	16	Castelmaurou	Mg	2007	Pub	34	Publier	H	2012
Lau	23	Launaguet	H	2007	Scv	47	St-Cyr-En-Val	H, M	2012
Lav	27	Lavaur	Dv	2007	Scz	44	Sciez	H, Ha, Lt	2012
Lyo	19	Lyon	Dv	2007	Vvo	20	Villenave D Ormon	P	2012
Prp	16	Perpignan	T	2008	Alb	40	Albertville	P	2013
Srp	63	St-Rémy de Provence	T	2009	Bas2	44	Bassens	P	2013
Frj	116	Fréjus	H, G, P, Rn, T	2010	Mal	44	Malause	D	2013
Slv	58	St-Laurent du Var	A, H	2010	Mou2	10	Moulès	A	2013
Adp	43	Althen-des-Paluds	P	2011	Msl	10	Montferrier-sur-lez	Ch	2013
Axb	22	Aix-les-Bains	P	2011	Sdp	6	Salon de Provence	T	2013
Blm	36	Boisset-lès-Montrond	P	2011					

**TABLE 1.2** – Récapitulatif des populations échantillonnées en France. Les données antérieures à 2008 sont des données acquises lors de travaux précédents. N : nombre d’individus génotypés. Plantes : A, Aubergine ; B, Basilic ; C, Concombre ; Ch, Chou ; D, Dipladénia ; Dv, Divers ; G, Gerbera ; H, Hibiscus ; Ha, Haricot ; L, Lamium ; Lt, Lantana ; M, Melon ; Mg, Magnolia ; P, Poinsettia ; Pv, Poivron ; R, Rosier ; Rn, Renoncule ; T, Tomate.

rapport au nombre total de populations, n’est pas différent de 50 %, quelle que soit l’année considérée (tests exacts de Fisher, toutes les valeurs de  $p$  largement supérieures à 5 %). Il n’y a donc pas plus de populations mixtes que de populations pures. L’espèce potentielle MEAM 1 a quant à elle été retrouvée seulement une fois ces dix dernières années en France (à Nice en 2004), contrairement à l’est du bassin méditerranéen où elle est souvent trouvée (Israël et Turquie).

### Effets de la plante hôte sur la structuration

*B. tabaci* est extrêmement polyphage d’après la littérature (voir aussi l’annexe A page 221) mais avec l’avancée des méthodes de génotypage et la révélation maintenant acceptée qu’il est en réalité un complexe d’espèces, il semblerait que certaines préférences de plantes hôtes existent. Par exemple en Afrique de l’ouest, Q3 n’est retrouvé que sur *Lantana camara* alors qu’il s’hybride avec Q1 qui a lui une large gamme de plantes hôtes (Communication personnelle). Ces préférences sont peut-être provoquées par la présence de différentes bactéries jouant un rôle dans le métabolisme (Luan *et al.* 2015 ; Rao *et al.* 2015).



**FIGURE 1.7** – Distribution des haplogroupes mitochondriaux en France, échantillonnés sur plusieurs années : de 2006 à 2013. Chaque diagramme représente une serre. Jaune : MEAM 1, rouge : Q1 (espèce MED), bleu : Q2 (espèce MED).

Nous avons tenté de tester la possibilité de préférence de plante hôte entre Q1 et Q2 avec notre échantillonnage, qui n'a pas été pensé au départ pour répondre à cette question. Nous avons 16 plantes hôtes différentes représentant de très nombreuses familles botaniques (*Brassicaceae*, *Cucurbitaceae*, *Euphorbiaceae*, *Fabaceae*, *Lamiaceae*, *Malvaceae*, *Rosaceae*, *Solanaceae*...) sur lesquelles on retrouve les deux haplogroupes, ce qui semble illustrer l'absence de spécialisation. Nous avons tout d'abord testé globalement la préférence potentielle de plante en décomptant les individus retrouvés sur chaque espèce (en retenant seulement les plantes échantillonnées à au moins deux reprises) et à partir de 2007 car Q2 était absent auparavant. Les données des différentes années ont été regroupées. Les données testées sont dans le tableau 1.3a. Le test ( $\chi^2 = 130,52$ ;  $p < 0,001$ ) est significatif. Ainsi il semblerait qu'on retrouve plus de Q1 sur Aubergine, Poinsettia et sur Tomate, alors que sur Hibiscus on retrouve plus de Q2. La faiblesse de ce test est de considérer les populations présentant potentiellement un seul haplogroupe, on ne peut exclure alors que l'autre haplogroupe n'a jamais été introduit dans cette exploitation et que la présence d'un seul groupe n'est pas forcément le résultat d'une compétition. Un deuxième test conserve seulement les population mixtes, où l'on retrouve les deux haplogroupes, on conserve donc deux lignes du tableau précédent. Une nouvelle fois le test est significatif ( $\chi^2 = 24,95$ ;  $p < 0,001$ ), les Q1 étant préférentiellement sur Poinsettia tandis que les Q2 sont plutôt sur Hibiscus. Il faut rester très prudent avec ces résultats car ce ne sont pas

les résultats de tests contrôlés en laboratoire, mais les résultats d'un échantillonnage prévu pour d'autres raisons !

	Q1	Q2		Q1	Q2
Aubergine	109	18	Hibiscus	104	201
Hibiscus	104	201	Poinsettia	142	114
Poinsettia	142	114			
Tomate	57	8			

(a) Test global

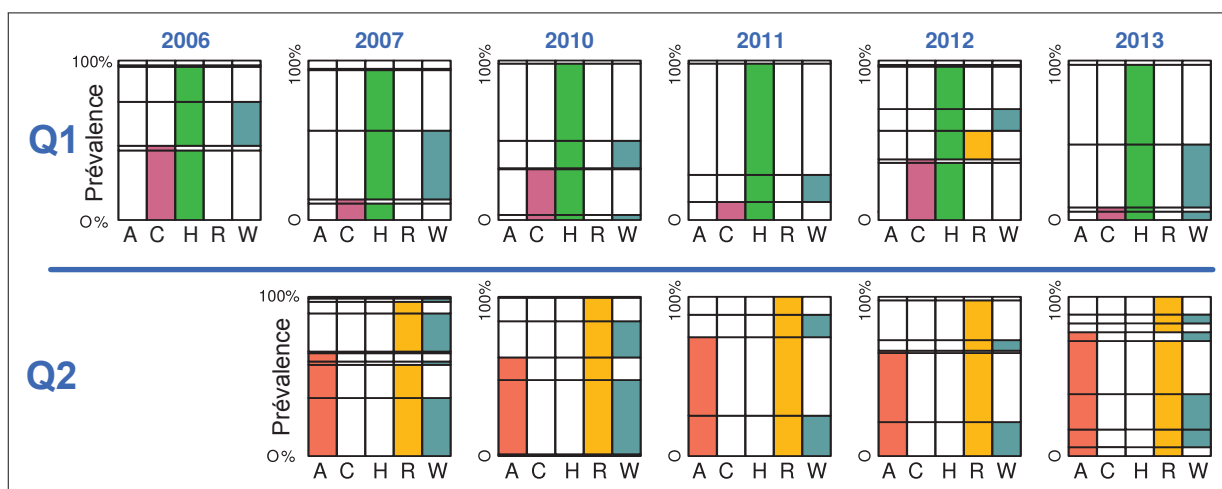
(b) Populations mixtes

**TABLE 1.3** – Nombre d'individus collectés par haplogroupe et par plante pour la détection d'une potentielle préférence de plante hôte. Ne sont conservées que les populations strictement postérieures à 2006 et dont la plante a été échantillonnée au moins à deux reprises. (a) : Test global, (b) : Seulement les populations mixtes.

## Description du compartiment symbiotique

La recherche systématique des symbiotes dans les individus montre une extrême stabilité des associations. L'haplogroupe Q1 est associé seulement à *Hamiltonella*, qui est quasiment fixée (93 % des individus sont infectés) ainsi qu'à *Wolbachia* et *Cardinium*, à des fréquences intermédiaires (respectivement 28 et 38 %). Les multi-infections, c'est-à-dire la présence d'au moins deux symbiotes secondaires, sont très fréquentes puisque 62 % des individus sont concernés. Les bactéries *Cardinium* et *Wolbachia* sont rarement trouvées dans une même population (seulement trois sur 27), et dans ces cas là on les retrouve moins fréquemment qu'attendu (produit des fréquences des symbiotes) dans un même individu ( $\chi^2 = 5,8$  ;  $p = 0,016$ ). L'haplogroupe Q2 est quant à lui retrouvé en association avec les bactéries *Rickettsia*, presque fixée avec un taux d'infection de 95 %, et *Arsenophonus* (62 %). La bactérie *Wolbachia* est retrouvée chez 69 % des individus mais c'est *a priori* une souche différente que celle retrouvée avec Q1 (Gueguen *et al.* 2010) mais nous ne l'avons pas testé. Les résultats sont consignés graphiquement sur la figure 1.8. Les multi-infections sont aussi très fréquentes puisque 48 % des individus ont deux symbiotes et 20 % en ont trois.

Une seule combinaison symbiotique atypique a été retrouvée, à deux endroits différents (deux serres ayant le même fournisseur de plantes) en 2012. Cette association est composée d'un haplotype mitochondrial Q1 associé aux bactéries *Rickettsia* et *Hamiltonella*. Ce profil d'infection étant atypique et très rare par rapport aux nombre d'individus et populations échantillonnés (et aussi aux données de la littérature, notamment la méta-analyse de Zchori-Fein *et al.* 2014), elle fera l'objet du chapitre suivant et ne sera pas plus longuement discuté ici.



**FIGURE 1.8** – Prévalences des infections associées aux haplogroupes Q1 (en haut) et Q2 (en bas) de l'espèce MED, pour les populations françaises, regroupées par année d'échantillonnage. A : *Arsenophonus*, C : *Cardinium*, H : *Hamiltonella*, R : *Rickettsia*, W : *Wolbachia*.

## Discussion

### Répartition géographique des haplogroupes : compétition, remplacement, coexistence ?

Les données montrent que Q1 était le seul haplogroupe présent en France jusqu'en 2006 (si on exclu MEAM1 dont la présence est anecdotique malgré sa domination antérieure dans ces régions comme nous l'avons indiqué en introduction de ce chapitre). L'apparition de Q2 en France en 2007 est simultanée à son apparition en Espagne et Italie (Gauthier *et al.* 2014 ; Parrella *et al.* 2013). Cet haplogroupe provient des pays de l'est méditerranéen où il était cantonné (Rua *et al.* 2006 ; Chu *et al.* 2008 ; Gueguen *et al.* 2010). Pour le moment, à l'échelle mondiale, c'est toujours l'haplogroupe Q1 qui est retrouvé (mis à part aux États-Unis), le Q2 n'est donc pas encore en phase d'invasion dans toutes les régions du monde. L'explosion démographique de Q1 (espèce MED) face à MEAM1, auparavant largement réparti en Europe, est probablement due à sa plus forte résistance aux insecticides comme il avait été proposé et comme certaines données génomiques semblent l'indiquer (sur-expression de certains gènes de détoxification comme les cytochromes *P450*, Wang *et al.* 2011). Dans un contexte d'utilisation intensive d'insecticides de toutes sortes pour contrôler les aleurodes, il apparaît évident qu'un génotype disposant de meilleures capacités de résistance soit invasif. Par exemple en Grèce, les principales méthodes de lutte contre *B. tabaci* ont été pendant plusieurs années l'utilisation des pyréthrianoïdes et des organophosphorés, ce qui a conduit rapidement à l'émergence de résistances (Roditakis *et al.* 2005). En Chine, les mêmes causes ont produit les mêmes effets (Wang *et al.* 2010), en Afrique de l'ouest l'utilisation massive des pyréthrianoïdes pendant plus de 30 ans sur coton a favorisé l'émergence de résistance, tout comme l'utilisation de l'acétamipride mais cette fois en quelques années seulement (Houndété *et al.* 2010). Ainsi l'apparition de résistances aux

insecticides n'est pas un phénomène rare chez *B. tabaci*.

Cette différence de tolérance aux insecticides entre les espèces potentielles MED et MEAM1 pourraient expliquer les données récentes de distribution de MEAM1. L'échantillonnage de 2013 montre la présence de MEAM1 à Troinex en Suisse (auparavant on ne trouvait que du Q1), ainsi qu'en Israël où l'on retrouve assez souvent l'espèce potentielle MEAM1 en populations naturelles (communication personnelle). Cette « ré-apparition » de l'espèce potentielle MEAM1, pourrait peut-être s'expliquer par une diminution de l'usage des insecticides, ce qui est par exemple le cas en Savoie et à Troinex en Suisse où toutes les exploitations échantillonnées utilisent principalement des auxiliaires de culture. Par contre nous n'avons pas d'informations sur les traitements utilisés pour la protection des cultures en Israël.

Cette différence de résistance aux traitements insecticides est peut être aussi présente entre les Q1 et Q2, ce qui expliquerait la coexistence entre les deux. Expliquons. Parrella *et al.* (2013) avancent l'hypothèse d'une différence dans l'utilisation des plantes hôtes, les *Solanaceae* seraient plus propices à Q2, qui serait aussi présent en plus grande proportion (par rapport à Q1) sur les plantes traitées que celles non traitées comme les adventices (87 % contre 40 %). Ainsi, les plantes exemptes de traitements serviraient de réservoir pour Q1 qui se maintiendrait localement où il serait plus compétitif. La réglementation sur les insecticides ainsi que leur utilisation évoluant constamment, une différence de résistance à ceux-ci (due à un bagage génétique différent, tels que les enzymes de détoxification) pourrait expliquer leur répartition et leur dynamique. Ces mécanismes de résistance, associés à des mutations ponctuelles ou des différences d'expression sont assez bien documentés chez *B. tabaci* (Rauch & Nauen 2004 ; Alon *et al.* 2008 ; Karunker *et al.* 2008 ; Nauen *et al.* 2013 ; Ye *et al.* 2014). Ces mêmes auteurs proposent que Q2 serait plus tolérant aux fortes températures car originaire de l'est du bassin méditerranéen où les chaleurs sont plus fortes, ce qui serait un atout compétitif pour lui.

Lors de la description des haplogroupes en France, nous avons trouvé des indices d'une différence de répartition selon la plante hôte, mais on pourrait peut-être aussi l'expliquer par des différences de traitements insecticides. Les cultures de Poinsettia sont globalement gérées en Production Biologique Intégrée avec une utilisation intensive des auxiliaires parasitoïdes, ce qui n'est pas du tout le cas dans les cultures d'Aubergine ou de Tomate par exemple, qui sont complètement ravagées soit dit en passant. Ainsi l'émergence de nouveaux modes de production pourrait favoriser des groupes génétiques au détriment des groupes dont la résistance aux produits phytosanitaires est plus importante mais plus coûteuse, en terme de valeur adaptative, lorsque la pression de sélection est absente. La littérature n'est pas d'accord sur ce point, il n'y aurait pas de coût associé à la résistance à l'acétamipride (Basit *et al.* 2012), ni au pyriproxifène (Crowder *et al.* 2009) mais pas pour d'autres auteurs (Horowitz *et al.* 2002). La présence d'allèles connus de résistances aux insecticides sera cherchée dans ces populations dans le chapitre 2.

La présence de plusieurs haplogroupes dans des populations mixtes a été hypothétisée comme plutôt rare et transitoire (Moya *et al.* 2001 ; Khasdan *et al.* 2005 ; Horowitz *et al.* 2005), abou-



tissant au déplacement de l'un par un autre (Kaplan & Denno 2007). Mais ce n'est pas un fait systématiquement observé : l'espèce potentielle MEAM 1 (à l'époque biotype B) a déplacé l'espèce potentielle NEW WORLD (à l'époque biotype A) aux États-Unis dans les années 1990 (Brown *et al.* 1995), mais cette même espèce potentielle n'a pas déplacé le clade autochtone (non nommé) en Espagne sur la même période (Guirao *et al.* 1997). Si il y a déplacement d'un groupe par l'autre, on devrait observer de nombreuses situations de populations pures, le groupe dominant excluant le moins adapté, l'identité de celui-ci variant avec les conditions environnementales locales. Si les conditions de cultures sont homogènes d'une population à l'autre (plante, traitements phytosanitaires, températures..) l'haplogroupe dominant devrait être le même, celui qui a la meilleure valeur sélective dans ces conditions. Dans le cas d'environnements variables d'une populations à l'autre, comme dans notre cas, on devrait obtenir une mosaïque de situations variées : mixité ou populations pures de l'un ou l'autre des cytotypes. Bien sûr, si la ressource n'est pas limitante, les haplogroupes ne devraient pas rentrer en compétition, est-ce le cas dans nos échantillonnage ? Dans les populations que nous avons échantillonnées, certaines semblent sous forte pression démographique, d'autres non. D'après nos observations personnelles, les cultures de poinsettias en Savoie ne sont pas soumises à une très forte pression de la part de *B. tabaci* car la lutte biologique est assez efficace et il peut même être difficile de trouver une dizaine d'individus dans une serre de plusieurs milliers de mètres carré. Dans ces conditions, la compétition pour les ressources paraît limitée. A l'inverse, dans des cultures maraichères (tomate ou aubergine principalement) dans le sud de la France, la densité d'individus est énorme, et certaines cultures entièrement détruites. On a donc des conditions variables en terme de pression démographique qui *in fine* contribuent au maintien de situations différentes en terme de cytotypes retrouvés

Dans notre échantillonnage, la plupart des populations sont mixtes, les haplogroupes ne s'excluent donc pas. L'absence de patron d'exclusion signifie donc que les conditions rencontrées dans les serres ne sont donc pas les mêmes partout, et qu'il y a une forte interaction entre environnement et fond génétique, créant de multiples situations différentes qui paraissent difficilement prédictibles. Les conditions à l'intérieur d'une serre peuvent être changeante (rotation des cultures, calendrier de traitement phytosanitaire, température) et faire varier au cours du temps les pressions de sélection, maintenant une situations où les groupes coexistent à des fréquences variables. L'apport répété de populations dans les serres par les échanges commerciaux peut aussi favoriser cette pluralité, qui ne serait pas durable et « artificiellement » entretenue. Seul un suivi répété dans le temps (hebdomadaire, mensuel ?) de ces populations pourraient répondre à ce type de question.



## Stabilité des associations : les cytotypes

### Les associations de bactéries ne sont pas recombinaées

En France et en Espagne, la diffusion de l'haplogroupe Q2 ne s'est pas accompagnée du déplacement de l'haplogroupe Q1, nous avons donc affaire à une situation particulière où la présence de plusieurs haplogroupes en sympatrie est détectée. Cette particularité va nous permettre de montrer que les haplogroupes et les bactéries forment des associations stables que sont les cytotypes, qui ont dès lors une signification biologique. De 2007 à 2013, 21 situations de mixité sont détectées, dans presque autant de localités (voir figure 1.7 et 1.4). Bien que cette cohabitation aurait pu être l'occasion de transfert horizontaux de bactéries d'un haplogroupe vers un autre, car les insectes sont retrouvés souvent sur la même plante, les données montrent une parfaite stabilité des cytotypes sur des centaines d'individus de 2003 à 2011. Tandis que la communauté symbiotique de Q1 peut être composée (tous les symbiotes ne sont pas forcément retrouvés dans un individu) de *Cardinium* et *Hamiltonella*, celle de Q2 est composée d'*Arsenophonus* et *Rickettsia*. Notons que *Wolbachia* peut être trouvée dans les deux. Le fait que l'on a jamais trouvé de Q2 en Grèce — le premier *B. tabaci* décrit en 1887 était du Q1 (Tay *et al.* 2012) et depuis Q2 n'a jamais été détecté —, et réciproquement jamais de Q1 en Israël nous amène à considérer ces deux pays comme références pour ces deux haplogroupes. On a donc bien adéquation entre les associations retrouvées dans ces pays de référence et celles en France et en Espagne. Ces données d'infection sont corroborées par celles de la littérature, en tout cas dans la région méditerranéenne (Chiel *et al.* 2009 ; Gueguen *et al.* 2010). Certaines bactéries ne se retrouvent pas associées à certains haplogroupes, il n'y a donc pas homogénéisation des cortèges symbiotiques.

Nous ne constatons pas de nouvelles associations, pourtant on sait que la plupart des symbiotes secondaires de *B. tabaci* peuvent transférer horizontalement d'un individu à l'autre (attention ce phénomène est rare, les symbiotes sont à transmission verticale la plupart du temps !) : *Arsenophonus*, *Cardinium*, *Rickettsia* et *Wolbachia* (Ahmed *et al.* 2013 ; Mouton *et al.* 2012) et notamment par la plante (Caspi-Fluger *et al.* 2012). Justement, nos données semblent indiquer que les occasions de transfert horizontaux sont très nombreuses car beaucoup de populations sont mixtes, avec cohabitation des cytotypes parfois sur la même plante. Nous devrions donc retrouver de très nombreuses associations différentes, mais ce n'est pas le cas ! Ces éléments nous permettent d'affirmer que le cytotype a une réalité biologique et qu'il n'est pas un assemblage aléatoire de bactéries. Le problème est donc d'expliquer cette stabilité.

### Quels mécanismes expliquent cette stabilité ?

L'absence de transferts horizontaux et la parfaite association de certaines bactéries avec les haplogroupes Q1 ou Q2 pourraient n'être qu'une « apparence ».

On peut faire l'hypothèse que les transferts sont nombreux, mais que les combinaisons créées ne

sont pas stables dans le temps pour plusieurs raisons, d'abord au niveau intra-hôte : 1) Les bactéries sont en compétition et certaines sont éliminées de l'individu 2) L'hôte n'est pas permissif aux bactéries 3) La bactérie ne s'adapte pas à son nouvel hôte. Ensuite, même si ces critères sont satisfaits, il se peut que l'individu porteur d'une telle association ait une valeur sélective fortement diminuée et être éliminé au niveau populationnel du fait de la compétition inter-hôtes. Les données que nous avons sont donc toujours des images à un instant  $t$  d'une population ; nous n'avons jamais de suivis temporels, très rarement des populations échantillonnées à plusieurs moments, et dans ce cas échantillonnées à plusieurs années d'intervalles. Pour illustrer cette hypothèse, imaginons l'association Q2-*Cardinium* qui apparaît dans une population composée de Q1-*Cardinium-Hamiltonella* et Q2-*Arsenophonus-Rickettsia*. L'association hypothétique Q2-*Cardinium* pourrait très bien être créée très régulièrement par transfert horizontal mais éliminée, nous ne pouvons pas l'exclure avec nos seules données si l'élimination est rapide. Mais dans le cas d'une élimination lente, on devrait en trouver parmi nos nombreux échantillons. Par contre l'existence de cette association pourrait être limitée par la capacité des entités la composant (les symbiotes) à cohabiter dans un même organisme ou une mauvaise adaptation de l'association hôte bactérie : le « cytotype est instable », au sens où l'ensemble des bactéries ne sont pas conservées à long terme.

Néanmoins si cette association fonctionne, c'est-à-dire que les bactéries cohabitent durablement dans un individu, celui-ci pourrait être moins compétitif que les autres et être éliminé de la population par la compétition inter-individu (faible fécondité ou survie). On ne le détecte donc pas. Ce que nous observons alors dans les populations naturelles est le résultat de la compétition intra-individu entre les bactéries, puis de la compétition inter-individus (les deux phénomènes de compétition peuvent bien sûr être simultanés).

Le chapitre 3 sera l'occasion de traiter plus en détail les interactions entre bactéries d'un même individu.



# Chapitre 2

## Analyse du compartiment nucléaire des cytotypes

1	Problématique	83
2	Matériels et méthodes	83
2.1	Allèles de résistance aux insecticides	84
2.2	Microsatellites	84
2.2.1	Génotypage	84
2.2.2	Analyse des données	85
3	Le cas des cytotypes ASL et Q1 en Afrique	86
3.1	Présentation de la situation	86
3.2	Échantillonnage	87
3.3	Résultats	88
3.3.1	Analyse microsatellite	88
3.3.2	Détection des allèles de résistance aux insecticides	89
3.4	Conclusion sur ce premier exemple	90
4	Le cas des cytotypes Q1 et Q2 en Europe	92
4.1	Présentation de la situation	92
4.2	Croisements contrôlés	92
4.2.1	Méthode	92
4.2.2	Résultat	93
4.3	Analyse du compartiment nucléaire	93
4.3.1	Échantillonnage	93
4.3.2	Allèles de résistance	94
4.3.3	Analyse microsatellite	94
4.4	Comparaison de populations mixtes et pures dans l'aire sympatrique	97
4.5	Conclusion sur le modèle Q1-Q2 en Europe de l'ouest	100

## Chapitre 2. Analyse du compartiment nucléaire des cytotypes

5	Discussion	101
5.1	Les bactéries n'induisent pas d'incompatibilités	101
5.2	Taxonomie de l'espèce potentielle MED	101
5.3	Limites des microsatellites, vers de nouveaux marqueurs	104

## Problématique

Aujourd'hui, les espèces potentielles formant le complexe d'espèces *B. tabaci* sont définies par la divergence observée sur le gène mitochondrial *COI*, qui doit être supérieure à 3,5 % pour pouvoir définir une nouvelle espèce potentielle. Dans l'introduction nous avons vu qu'il y a actuellement 41 espèces potentielles. Potentielles car malgré l'accumulation de données génétiques et comportementales sur les barrières reproductives (Xu *et al.* 2010 ; Liu *et al.* 2012), il y a toujours un manque d'informations pour de nombreux groupes ainsi que des questions soulevées par ce critère de divergence.

Un problème soulevé par ces espèces potentielles est le critère de divergence qui a été érigé grâce aux données de tout le complexe, ainsi la très forte divergence entre certaines paires d'espèces potentielles peut très bien masquer certains phénomènes ayant lieu dans des espèces potentielles en particulier. Ainsi des barrières reproductives peuvent exister dans certains groupes pour des valeurs de divergence inférieures à 3,5 % et *a contrario* on peut imaginer des groupes interféconds malgré une divergence plus élevée.

Dans ce chapitre, nous allons nous intéresser aux potentielles barrières reproductives à l'intérieur du complexe *B. tabaci* en nous focalisant sur l'espèce potentielle MED et plus particulièrement sur deux situations retrouvées en Afrique et en Europe, mettant respectivement en jeu les cytotypes ASL et Q1 (nous évoquerons aussi Q3), puis dans un deuxième temps Q1 et Q2.

Dans les deux cas nous utiliserons des marqueurs nucléaires pour explorer les flux génétiques entre les deux groupes : des microsatellites comme marqueurs neutres de la diversité, ainsi que des gènes décrits dans la littérature comme conférant des résistances aux insecticides et donc sous sélection. Dans le cas de Q1 et Q2 nous testerons l'interfécondité au laboratoire à l'aide de croisements contrôlés.

Les marqueurs microsatellites vont nous permettre de mesurer la diversité et la structuration génétique de ces cytotypes. En effet, les populations mixtes rencontrées nous donnent l'opportunité de tester l'hybridation entre ces cytotypes ou inversement la présence de barrières reproductives, en population naturelle. Quant aux marqueurs de résistance aux insecticides ils vont nous apporter une information supplémentaire sur les phénotypes sélectionnés en condition d'admixture. En effet, les populations d'aleurodes sont soumises à une forte pression insecticide, nous avons donc aussi étudié des gènes connus comme étant impliqués dans la résistance aux organophosphorés et aux pyréthrinoïdes (Alon *et al.* 2006).

## Matériels et méthodes

Les haplogroupes des individus échantillonnés ont été identifiés par PCR-RFLP et séquencés lorsque nécessaire, comme décrit dans le chapitre précédent.

## Allèles de résistance aux insecticides

Les insecticides de la famille des pyréthriinoïdes et des organophosphorés agissent tous les deux sur le système nerveux. Les premiers perturbent le fonctionnement du canal sodium et ainsi la transmission de l'influx nerveux (Morin *et al.* 2002), le deuxième inactive l'acétylcholine estérase et provoque l'accumulation d'acétylcholine dans la synapse, bloquant l'influx nerveux (Alon *et al.* 2008). Des mécanismes de résistance à ces deux substances, soit par insensibilité de la cible, soit par augmentation des mécanismes de détoxification, ont été identifiés chez *B. tabaci*. (Byrne & Gorman 2000 ; Alon *et al.* 2008). Une résistance aux organophosphorés a été caractérisée sur le gène de l'acétylcholine estérase (*ace1*), la mutation F331W (allèle R) (Alon *et al.* 2008) ainsi que deux mutations ponctuelles sur le gène du canal sodium (L925I et T929V, respectivement allèles r1 et r2) provoquant une résistance aux pyréthriinoïdes de l'espèce potentielle MED, dans de nombreux pays (Italie, Espagne, Allemagne, Israël, Grèce, Japon ; Alon *et al.* 2006 ; Roditakis *et al.* 2006). Nous avons donc recherché ces trois mutations ponctuelles dans les populations d'aleurodes.

Les mutations L925I et F331W ont été identifiées par une PCR-RFLP, tandis que la mutation T929V l'a été par PCR spécifique (Tsagkarakou *et al.* 2009). Les profils ambigus ont été séquencés.

Les fréquences alléliques ont été analysées avec des tests exacts de Fisher ou des tests de  $\chi^2$  lorsque cela était possible, test utilisé aussi pour tester l'équilibre de Hardy-Weinberg. Toutes ces analyses ont été effectuées avec le logiciel R (R Core Team 2015).

## Microsatellites

### Génotypage

Sept locus microsatellites ont été utilisés comme marqueurs nucléaires neutres pour étudier la diversité génétique et la différenciation des populations et cytotypes de *B. tabaci*. Seules les femelles ont été utilisées car *B. tabaci* est un insecte haplo-diploïde, les mâles sont donc haploïdes et ne permettent pas d'estimer l'hétérozygotie. Les caractéristiques des amorces utilisées sont dans le tableau 2.1.

Les PCR ont été réalisées séparément pour chaque paire d'amorces dans un volume final de 10  $\mu$ L contenant 1x de tampon, 0,375  $\mu$ L de  $MgCl_2$ , 200  $\mu$ M de dNTP, 200 nM de chaque amorce et 0,5 UI de Taq polymérase Euroblue<sup>TM</sup> (Eurobio). Le programme d'amplification utilisé est : 2 mn à 94°C ; 30 s à 94°C, 30 s à la température d'hybridation (Voir tableau 2.1) 1 mn à 72°C (35 fois) ; 10 mn d'élongation à 72°C. Une amorce de chaque paire est fluorescente (marqueurs FAM et HEX) pour permettre l'analyse ultérieure, chez la société Biofidal à Lyon (séquenceur ABIPrism), avec le marqueur de taille GS600LIZ<sup>®</sup>. Les tailles des allèles ont été déterminées avec les logiciels Peak Scanner v1.0 et GeneMapper<sup>®</sup> v4.0.



Locus	Amorce	Séquence	Motif	Référence	T (°C)
BtIS2.3	BtIs2.3F	5'-CAGAACGACAGGTTCGAG-3'	(GA) <sub>n</sub>	(Gauthier <i>et al.</i> 2008)	50
	BtIs2.3R	5'-CAAAATTAATGGTATTGACTC-3'			
BtIS1.13	BtIs1.13F	5'-CTAAGACCGATTCCCTCC-3'	(CA) <sub>n</sub>	(Gauthier <i>et al.</i> 2008)	52
	BtIs1.13R	5'-GAATACTACACCTTCAATTACC-3'			
Locus 11	11a	5'-CCAGAAAAGTGGACTTAAGA-3'	(GT) <sub>n</sub>	(Delatte <i>et al.</i> 2006)	57
	11b	5'-GATCTGGGTGTTTTCTTCTA-3'			
BT83	BT83F	5'-GATGCCACAGGTTGTCTGG-3'	(TC) <sub>n</sub> G(CA) <sub>n'</sub>	(Tsagkarakou & Roditakis 2003)	57
	BT83R	5'-GCTTGCCAGGCACTTCTAG-3'			
BtIS1.1	BtIs1.1F	5'-CCCATAGAACACGCTCC-3'	(CA) <sub>n</sub>	(Gauthier <i>et al.</i> 2008)	57
	BtIs1.1R	5'-CATTGGAAGCCTCGAATAC-3'			
BtIS1.2	BtIs1.2F	5'-CTTACCTTCCATTACC-3'	(CA) <sub>n</sub> N <sub>n'</sub> (CA) <sub>n''</sub>	(Gauthier <i>et al.</i> 2008)	57
	BtIs1.2R	5'-ATCCCAGATCTTATGTTG-3'			
Locus 145	145a	5'-CCTACCCATGAGAGCGGTAA-3'	(AC) <sub>n</sub>	(Dalmon <i>et al.</i> 2008)	57
	145b	5'-TCAACAAACGCGTTCCTCAC-3'			

**TABLE 2.1** – Caractéristiques des amorces utilisées pour amplifier les locus microsatellites.

### Analyse des données

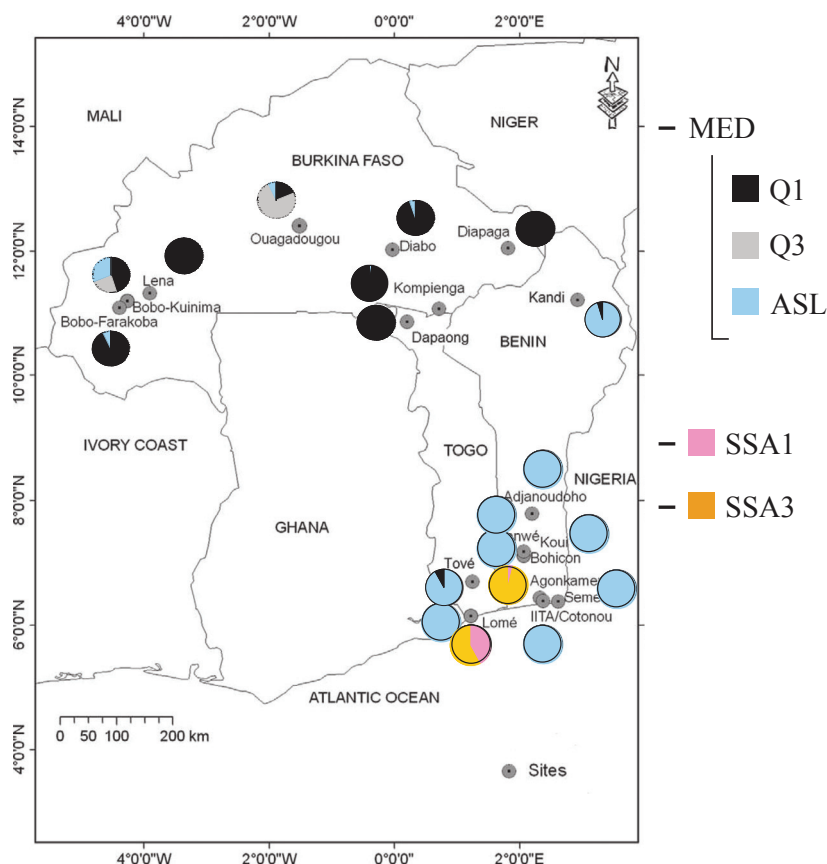
Le nombre moyen d'allèles par locus, l'hétérozygotie observée  $H_o$  et attendue  $H_e$  ont été calculés avec GENETIX (Belkhir *et al.* 1996). Les estimateurs de l'indice de fixation  $F_{IS}$  (Weir & Cockerham 1984) ont été calculés avec GENEPOP 4.0 (Raymond & Rousset 1995) ainsi que les tests exacts de l'équilibre de Hardy-Weinberg, le déséquilibre de liaison entre les paires de locus et les estimateurs de différenciation  $F$ . Nous avons utilisé le logiciel STRUCTURE (Pritchard *et al.* 2000) pour établir le nombre de groupes génétiques à l'équilibre. Ce logiciel bayésien utilise la fréquence des allèles à chaque locus. Les paramètres suivants ont été utilisés : 500 000 chaînes dont 50 000 écartées ; avec possibilité d'admixture et corrélation des fréquence alléliques. Nous avons calculé la vraisemblance pour un nombre de groupes allant de 1 au nombre de populations, en faisant 10 répétitions. Le nombre de groupes a été établi avec la vraisemblance ainsi que la variation d'ordre 2 ( $\Delta$ ) de la vraisemblance d'après la méthode proposée par Evanno *et al.* (2005).

En plus, pour les populations européennes, une analyse en composante principale a été réalisée sur ces données avec la bibliothèque *adeigenet* (Jombart 2008) de l'environnement R (R Core Team 2015). L'avantage d'une telle méthode est qu'elle est exploratoire dans le sens où elle ne nécessite pas de modèle génétique sous-jacent et n'est pas contrainte par le déséquilibre de liaison ou l'absence d'équilibre d'Hardy-Weinberg. En plus, comme la méthode ne fait que résumer les données en réduisant leurs dimensions, on peut mettre en évidence des gradients contrairement à la plupart des approches qui décrivent des populations discrètes (Jombart *et al.* 2009). Un dernier avantage est la rapidité des calculs (pour nos jeux de données microsatellites, il y a au moins trois ordres de grandeurs de différences en temps de calcul). Pour finir, une analyse hiérarchique de la variance moléculaire a été réalisée avec la bibliothèque *hierfstat* (Goudet 2005) de l'environnement R, afin de tester l'effet de plusieurs facteurs sur la structuration génétique : symbiotes, plante hôte et structure géographique.

## Le cas des cytotypes ASL et Q1 en Afrique

### Présentation de la situation

Des campagnes d'échantillonnage réalisées en Afrique de l'ouest dans trois pays (Bénin, Togo et Burkina-Faso) de 2007 à 2009 ont montré la présence de trois espèces potentielles du complexe *B. tabaci* : MED, SSA 1 et SSA 3 (voir la figure 2.1 et Gnankiné *et al.* 2013a)<sup>1</sup>. Ces deux dernières espèces potentielles ont une distribution très restreinte, au bord de l'océan. Les données de la littérature montrent que SSA 1 est présente dans à peu près tous les pays d'Afrique au sud du Sahel et que SSA 3 est présente dans les pays en bordure du golfe de Guinée.



**FIGURE 2.1** – Distribution des espèces potentielles du complexe *B. tabaci* en Afrique de l'ouest (MED, SSA 1 et SSA 3), échantillonnées de 2007 à 2009. Pour l'espèce MED, ses différents cytotypes sont précisés : Q1, Q3 et ASL.

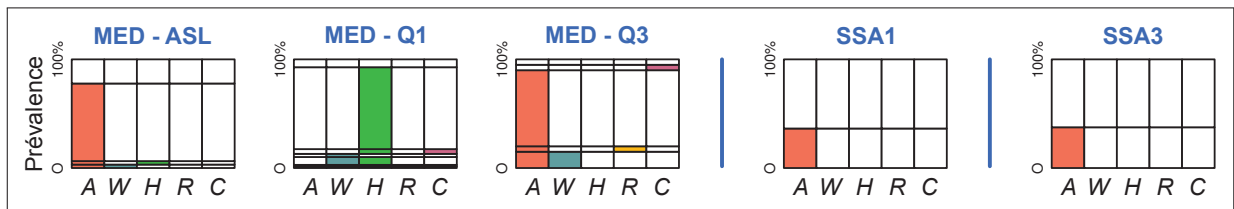
En tout état de cause, dans cette partie nous allons nous restreindre à l'étude de l'espèce potentielle MED représentée par plusieurs cytotypes en sympatrie (Q1, Q3 et ASL), ce qui fait l'intérêt de cette situation. Nous avons déjà longuement évoqué Q1 dans le chapitre précédent car invasif au niveau mondial tandis que nous avons seulement évoqué les autres en introduction. Le

1. Gnankiné O., Mouton L., Henri H., Terraz G., Houndeté T., Martin T., Vavre F., Fleury F. (2013). *Distribution of Bemisia tabaci biotypes (Homoptera : Aleyrodidae) and their associated symbiotic bacteria on host plants in western Africa. Insect Conservation and Diversity*, 6(3), 411–421.

cytotype ASL est surtout présent dans le sud du Togo et du Bénin ainsi qu’au Burkina-Faso où il est à plus faible fréquence (Il est néanmoins présent dans d’autres pays d’après les données *Genbank* : Cameroun, Zimbabwe, Ouganda, Nigeria, Côte d’Ivoire et Ghana). Dans nos échantillons, on le retrouve souvent dans les mêmes populations que Q1. Le Q3 est quant à lui extrêmement rare (seulement deux localités au Burkina-Faso, seul pays où il a été détecté d’après les données *Genbank*).

Les cortèges symbiotiques de ces cytotypes sont les suivants : Q1 en Afrique est principalement infecté par *Hamiltonella* et à une moindre fréquence par *Cardinium* et *Wolbachia* ; ASL est infecté par *Hamiltonella* et *Arsenophonus* alors que Q3 est seulement infecté par *Rickettsia* et *Arsenophonus*. Sur la figure 2.2 sont montrés des exemples de prévalence des différents symbiotes pour ces cytotypes (ainsi que pour les espèces SSA 1 et SSA 3).

Les cytotypes Q1 et ASL sont en sympatrie au Burkina-Faso et se trouvent parfois dans des populations mixtes, c’est-à-dire sur la même plante hôte. Cette configuration particulière nous a permis d’explorer le degré de divergence génétique nucléaire entre ASL et Q1, sachant que la distance mitochondriale entre les deux est inférieure à 3,1 % sur le gène *COI* (Henri *et al.* 2013)<sup>1</sup>. Quant au cytotype Q3, il n’est pas retrouvé en population mixte avec les deux autres cytotypes Q1 et ASL car on le retrouve seulement sur *Lantana camara* (*Verbenaceae*) où aucun autre cytotype n’a été détecté. Par contre on trouve des hybrides entre Q3 et ASL en situation naturelle.



**FIGURE 2.2** – Prévalence des différents symbiotes des espèces potentielles retrouvées en Afrique de l’ouest, pour l’espèce MED sont détaillés les trois cytotypes Q1, Q3 et ASL. Les données complètes par populations sont dans l’article A page 183<sup>2</sup>.

## Échantillonnage

Des échantillons ont été prélevés en différents endroits du Burkina-Faso (Afrique de l’ouest) au cours de l’année 2007 pour le génotypage microsatellite (voir tableau 2.2) et en 2009 pour le génotypage des gènes de résistance aux insecticides (voir tableau 2.3). Les individus, adultes, ont été prélevés sur différentes plantes maraîchères : tomate, aubergine, courge, concombre ; ainsi que sur coton et tabac. La figure 2.3 montre une vue générale du Burkina-Faso avec les différentes localités où les échantillonnages ont eu lieu.

1. Henri, H., Terraz, G., Gnankiné, O., Fleury, F., & Mouton, L. (2013). *Molecular characterization of genetic diversity within the Africa/Middle East/Asia Minor and Sub-Saharan African groups of the Bemisia tabaci species complex. International Journal of Pest Management*, 59(4), 329–338.



FIGURE 2.3 – Carte des lieux d’échantillonnage au Burkina-Faso.

Lieu	Plante hôte	Cytotype	N	$F_{IS}$
Kompienga	Coton	Q1	15	0,224
Diapaga	Coton	Q1	14	0,223
Lena	Coton	Q1	15	0,190
Bobo/Farako-Bâ	Coton	Q1	14	0,276
	Aubergine	Q1	8	0,119
		ASL	1	-
Bobo/Kuinima	Tomate	Q1	9	0,041
		ASL	34	0,039
	Courge	Q1	11	0,068
		ASL	14	0,064

TABLE 2.2 – Échantillons de *B. tabaci* du Burkina-Faso utilisés pour le génotypage microsatellite. *N* : nombre d’individus.

## Résultats

### Analyse microsatellite

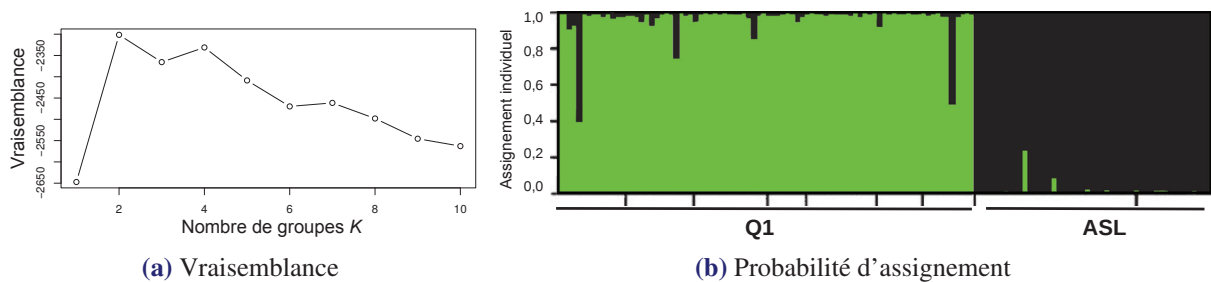
Les résultats de cette analyse du compartiment nucléaire des cytotypes Q1 et ASL ont été publiés dans Mouton *et al.* (2015)<sup>1</sup>, en annexe A 205.

Un total de 135 femelles de cytotypes Q1 (n = 86) ou ASL (n = 49), provenant de cinq sites, ont été génotypées sur les sept locus microsatellites. Les individus ASL proviennent d’un seul endroit mais de trois plantes hôtes différentes et ont donc été considérés comme provenant de trois populations différentes. La diversité allélique est très variable, allant de 4 à 31 allèles par locus. Le nombre moyen d’allèles par locus pour ASL est de 5, ce qui est plus faible que la moyenne de 11 trouvée pour Q1. Les locus sont indépendants car aucun déséquilibre de liaison n’a été détecté.

1. Mouton L., Gnankiné O., Henri H., Terraz G., Ketoh G., Martin T., Fleury, F., Vavre, F. (2015). *Detection of genetically isolated entities within the Mediterranean species of Bemisia tabaci : new insights into the systematics of this worldwide pest. Pest Management Science, (February). doi :10.1002/ps.3834*

Un très fort  $F_{IS}$  a été détecté (voir tableau 2.2), même en regroupant les individus par cytotype (globalement 0,19 pour Q1 et 0,22 pour ASL ;  $p < 0,001$ ) ; cette valeur est expliquée par le très fort déficit en hétérozygotes. Les valeurs de  $F_{IS}$  par population s'échelonnent de 0,04 à 0,27 pour Q1 et de 0,04 à 0,06 pour ASL (attention il y a seulement deux valeurs).

Une analyse de la structuration génétique a été réalisée avec STRUCTURE afin de déterminer le nombre  $K$  de groupes génétiques à l'équilibre d'Hardy-Weinberg ; en faisant varier  $K$  de 1 à 10. La plus forte vraisemblance a été obtenue pour  $K = 2$  (voir figure 2.4a), résultat soutenu par le  $\Delta K$  de Evanno *et al.* (2005), résultat non présenté ici. Les deux groupes correspondent aux cytotypes Q1 et ASL qui sont donc aussi différenciés au niveau nucléaire (voir figure 2.4b). La différenciation globale entre ces deux groupes est plutôt élevée avec une valeur de  $F_{ST}$  valant 0,183 ( $p < 0,01$ ). En analysant le jeu de données globalement, on peut ne pas déceler une structuration génétique à l'intérieur de certains groupes (Rosenberg *et al.* 2002), l'analyse a donc été réalisée aussi sur les deux cytotypes séparément. Celle-ci montre l'absence de différenciation parmi les populations de chaque cytotype ( $K = 1$  pour chaque), ce qui est confirmé par les faibles valeurs de  $F_{ST}$ , non significatives, pour ces deux groupes (0,009 et 0,003 respectivement pour Q1 et ASL).



**FIGURE 2.4** – Structure génétique de Q1 et ASL au Burkina-Faso obtenue avec l'ensemble des données microsatellites analysées avec le logiciel STRUCTURE. (a) : Vraisemblance des résultats en fonction du nombre de groupes  $K$ , (b) : Probabilités individuelles d'assignement aux deux groupes génétiques (ce nombre 2 de groupes génétiques est la valeur la plus vraisemblable d'après la figure (a)). Un groupe est en vert, l'autre en noir. Chaque barre verticale représente un individu, la proportion relative de chaque couleur pour un individu représente les probabilités relatives d'assignement pour chaque groupe. Les individus sont triés par cytotype (ASL et Q1).

#### Détection des allèles de résistance aux insecticides

Les mutations dans le gène *ace1* (F331W) et dans le gène du canal sodium (L925I et T929V) ont été recherchées chez 166 individus du Burkina-Faso, soit de cytotype ASL soit Q1, à l'aide de PCR spécifiques ou de PCR-RFLP. Les résultats sont présentés dans la table 2.3. Aucun déséquilibre de liaison n'a été détecté entre les deux gènes (test exact de Fisher,  $p > 0,05$ ), nous pouvons donc les considérer comme indépendants. Pour les deux gènes, la fréquence des mutations est très variable d'une population à l'autre et d'un cytotype à l'autre. Pour le gène *ace1*, la fréquence de la mutation F331W est de 0,98 pour le cytotype Q1 et 0,59 pour ASL. Pour

## Chapitre 2. Analyse du compartiment nucléaire des cytotypes

le canal sodium, seule l'une des deux mutations a été détectée (L925I) qui est pratiquement fixée chez Q1 (99 %) mais extrêmement rare chez ASL avec une fréquence de 2 %. Ces fréquences sont assez homogènes entre les populations d'un même cytotype (test exact de Fisher,  $p > 0,05$ ). Les fréquences des génotypes ne diffèrent pas significativement de celles attendues à l'équilibre d'Hardy-Weinberg pour les deux gènes et pour les deux cytotypes.

Ville	Plante	Cytotype	<i>Canal sodium</i>							<i>ace1</i>								
			Mâles			Femelles				Mâles			Femelles		f(R)			
			n	r1	s	n	r1r1	r1s	ss	f(r1)	n	R	S	n		RR	RS	SS
Kompienga	Coton	Q1	6	6	0	10	8	2	0	0,92	6	5	1	10	8	2	0	0,88
Bobo/Koko	Tomate	ASL	0	0	0	5	0	0	5	0	0	0	0	5	0	2	3	0,20
		Q1	1	1	0	10	10	0	0	1	1	1	0	10	10	0	0	1
Léguéma	Tomate	ASL	2	0	2	12	0	0	12	0	2	1	1	12	4	2	6	0,42
		Q1	3	3	0	12	12	0	0	1	3	3	0	12	12	0	0	1
Léguéma	Courge	ASL	0	0	0	5	0	0	5	0	0	0	0	5	1	1	3	0,3
		Q1	0	0	0	6	6	0	0	1	0	0	0	6	5	1	0	0,92
Toukoro	Tomate	Q1	2	2	0	13	13	0	0	1	2	2	0	13	13	0	0	1
Bobo/Kuinima	Concombre	Q1	0	0	0	29	29	0	0	1	0	0	0	29	29	0	0	1
Bittou	Coton	Q1	4	4	0	19	19	0	0	1	4	4	0	19	18	1	0	0,97
Soumouso	Tomate	ASL	4	1	3	6	0	0	6	0,06	4	4	0	6	1	3	2	0,56
		Q1	2	2	0	15	15	0	0	1	2	2	0	15	15	0	0	1

**TABLE 2.3** – Fréquences des mutations sur les gènes *ace1* et du canal sodium dans les populations d'Afrique de l'ouest, classées par plante hôte et localité. Pour chaque sexe et gène, les génotypes sont indiqués, puis la fréquence de l'allèle de résistance pour les femelles (les mâles étant haploïdes, la fréquence du génotype résistant est égale à la fréquence de l'allèle).  $n$  : nombre d'individus,  $r1$  et  $s$  : respectivement allèles de résistance et sensible du gène *Canal sodium*, R et S : respectivement allèles de résistance et sensible du gène *ace1*.

### Conclusion sur ce premier exemple

Les cytotypes Q1 et ASL sont classés dans l'espèce potentielle MED sur la base de leur différenciation (3,1 %) mitochondriale sur le gène mitochondrial *COI* (Dinsdale *et al.* 2010 ; De Barro *et al.* 2011). A l'aide de marqueurs microsatellites nous avons montré une faible différenciation entre populations du même cytotype, ce qui suggère que des échanges génétiques ont lieu, même à cette échelle : les populations sont distantes de 23 à 655 kilomètres et ne sont pas toujours prélevées sur la même plante. Par contre les données montrent que les individus Q1 et ASL sont isolés génétiquement car on ne détecte pas d'hybrides malgré les populations mixtes : les deux groupes identifiés par STRUCTURE correspondent strictement aux deux cytotypes (voir figure 2.4) identifiés — rappelons le — par un marqueur mitochondrial. Le  $F_{ST}$  mesuré entre ces deux groupes est de plus assez élevé (0,18), confirmant l'absence d'hybridation malgré le contact étroit. La différenciation pourrait être due à la formation du désert du Sahara qui aurait séparé les deux groupes : ASL en Afrique de l'ouest, Q1 en méditerranée (Rua *et al.* 2006). La situation actuelle serait alors la conséquence de la colonisation l'Afrique de l'ouest par Q1, comme d'ailleurs un peu partout dans le monde.

Les deux mutations recherchées conférant des résistances aux insecticides sont en très forte fréquence dans le cytotype Q1, celle conférant la résistance aux organophosphorés (sur le gène



*ace1*) étant presque fixée (99 %). A l'inverse, les deux mutations sont beaucoup moins présentes dans le cytotype ASL, la fréquence de l'allèle de résistance sur le gène *ace1* est de 0,59 et la mutation sur le gène du *Canal sodium* est quasiment absente (1 seul mâle la porte). L'émergence des résistances apparaît évidente car les traitements insecticides sont utilisés à forte doses depuis de nombreuses années dans cette région, autant les organophosphorés (Gnankiné *et al.* 2007) que les pyréthrinoïdes (Houndété *et al.* 2010). Par contre ce qui est notable est la différence observée entre les deux cytotypes. En effet les deux sont en sympatrie et même dans des populations mixtes, ils sont donc soumis tous les deux aux mêmes pressions insecticides. Une première possibilité pour expliquer cette différence de résistance est que les individus ASL utilisent d'autres mécanismes comme la sur-expression de gènes de détoxification (Karunker *et al.* 2008), ou l'utilisation d'autres molécules pour détoxifier comme par exemple les glutathion S-transférase (Rauch & Nauen 2004). Une deuxième possibilité est la présence de plantes refuges pour ASL. Dans le modèle source-puits, les parcelles de coton traitées seraient le puits où les populations sont maintenues par l'immigration depuis une plante hôte non identifiée.

Avec ces résultats nous pouvons faire l'hypothèse que l'espèce potentielle MED est composée en réalité d'au moins deux espèces distinctes. L'isolement reproducteur n'est pas explicable par des barrières physiques puisqu'ils se trouvent parfois sur la même plante.

Il est peu probable que l'incompatibilité cytoplasmique en soit la cause puisque les bactéries que ces deux cytotypes — ou peut être espèce maintenant — portent avec un taux d'infection suffisamment élevé (respectivement *Hamiltonella* dans Q1 et *Arsenophonus* dans ASL), bien qu'étant différentes, n'induisent pas d'incompatibilité quel que soit l'hôte qui les héberge. La prévalence des autres bactéries n'est pas assez élevée (moins de 10 %) pour expliquer l'isolement reproducteur qu'elles provoqueraient. Il est donc plus probable que l'absence d'échanges génétiques entre Q1 et ASL soit due à une absence de copulation — quel qu'en soit le mécanisme comportemental — ou des incompatibilités post-copulatoires provoquées par des conflits nucléo-cytoplasmiques par exemple. Il y a beaucoup d'études comportementales dans la littérature s'attachant à déterminer si les mécanismes d'isolement reproducteur entre les différentes espèces du complexe *B. tabaci* sont de type pré- ou post-copulatoires. Il semblerait que pour la plupart des paires d'espèces testées, les mâles et les femelles procèdent à des parades nuptiales, parfois longues, mais qui n'aboutissent pas à la copulation (Liu *et al.* 2012). Les observations entre les espèces potentielles MED et MEAM1 montrent que mâles et femelles effectuent les premiers stades de la parade nuptiale (battements antennaires), mais qu'ils restent « bloqués » à cette étape (Elbaz *et al.* 2010). Des observations similaires ont été faites entre les espèces potentielles AUSTRALIA et ASIA-II-3 (Liu *et al.* 2007). En plus de ces barrières pré-copulatoires, il existe aussi des barrières post-copulatoires comme l'ont montré Sun *et al.* (2011) entre MED et MEAM1, De Barro & Hart (2000) entre MEAM1 et AUSTRALIA ainsi que Byrne *et al.* (1995) entre NEW WORLD et MEAM1.



## Le cas des cytotypes Q1 et Q2 en Europe

### Présentation de la situation

Les principaux résultats de cette partie ont été publiés dans Terraz *et al.* (2014)<sup>1</sup>, article en annexe page 171.

Dans le chapitre précédent nous avons décrit les haplogroupes mitochondriaux Q1 et Q2 ainsi que leur cortège bactérien spécifique, d'où l'emploi du terme cytotype pour les désigner. Tous deux sont originaires du bassin méditerranéen, mais Q2 est originaire des pays le plus à l'est du bassin. Comme Q1 n'a jamais été retrouvé en Israël, nous avons considéré le Q2 en Israël comme étant le Q2 dit de référence. Inversement, en Grèce seul Q1 a été trouvé, nous l'avons donc appelé Q1 de référence. En France, mais la situation est similaire en Espagne et en Italie, Q1 est le principal cytotype présent à partir de 2003 (Gueguen *et al.* 2010 ; Gauthier *et al.* 2014). En 2007, la situation évolue avec l'apparition du cytotype Q2 simultanément en France, Espagne et Italie. Son arrivée ne provoque pas la disparition de Q1, mais engendre plutôt une mosaïque de situations avec de très nombreux cas de mixité, ce qui est encore vérifiable en 2013, dernière année d'échantillonnage. On a donc une situation très particulière où deux cytotypes aux cortèges symbiotiques très différents et précédemment isolés géographiquement se retrouvent en contact direct, dans de très nombreux endroits, ce qui va nous permettre de tester la présence d'hybrides en populations naturelles. Pour cela nous avons procédé en deux étapes, en mettant tout d'abord en place un protocole en laboratoire pour tester l'hybridation en conditions contrôlées. En plus de ces tests comportementaux, nous avons comparé les compartiments nucléaires des Q1 et Q2 dits de référence, puis nous avons comparé avec ceux présents en sympatrie (en l'occurrence 4 points d'échantillonnage en Espagne et en France) à l'aide des mêmes sept locus microsatellites que dans la partie précédente. Nous avons également déterminé la fréquence des allèles de résistances (ceux utilisés dans la partie précédente) pour voir comment évolue la fréquence de ces allèles dans un contexte d'admixture et de sélection.

### Croisements contrôlés

#### Méthode

Pour les croisements, nous avons utilisé deux lignées provenant de Sigeon pour le Q1 (Languedoc Roussillon) et de Lyon (Rhône-Alpes) pour le Q2. La lignée Q1 est infectée par *Hamiltonella* et *Cardinium* alors que la lignée Q2 est infectée par *Rickettsia* et *Arsenophonus*. Ces deux lignées ont été maintenues plusieurs années au laboratoire (12 à 15 générations par an) avant les expérimentations. Elles sont maintenues sur *Hibiscus moscheutos* var. Galaxy, à 25 °C avec un cycle jour/nuit de 16/8 heures.

---

1. Terraz G., Gueguen G., Arnó J., Fleury F., Mouton L. (2014) ; *Nuclear and cytoplasmic differentiation among Mediterranean populations of Bemisia tabaci : testing the biological relevance of cytotypes*. **Pest management science**, 70, 1503–1513.

Des pupes (dernier stade larvaire) des cytotypes Q1 et Q2 sont isolées, de façon à obtenir des émergents vierges pour le lendemain (les femelles émergentes ne sont pas réceptives pendant 16 à 24 h). Ces adultes sont alors triés selon le sexe. Les quatre croisements possibles ont été réalisés (Q1♀ \* Q1♂, Q1♀ \* Q2♂, Q2♀ \* Q1♂ et Q2♀ \* Q2♂), avec 10 répétitions pour chaque. Cinq femelles et cinq mâles sont placés sur un disque foliaire d'Hibiscus de 30 mm de diamètre, lui-même sur un milieu agar à 1 % dans une boîte de Pétri de 50 mm de diamètre, pour une durée de 6 jours. Les oeufs pondus sont comptés 3 jours après avoir enlevé les femelles, puis les adultes émergents. Pour tester la fertilité des F1, les descendants mâles et femelles d'une même boîte ont été conservés ensemble pour permettre la fécondation des femelles et ainsi vérifier la production de descendants mâles et femelles dans le F2.

### Résultat

Des descendants viables mâles et femelles ont été obtenus pour tous les croisements réalisés, suggérant qu'il n'y a ni incompatibilité nucléaire ni incompatibilité cytoplasmique, en tout cas pas assez fortes pour être détectées ici. Le nombre moyen d'oeufs pondus et le pourcentage d'émergence est indiqué dans le tableau 2.4. Le croisement Q1♀ \* Q1♂ a produit plus d'oeufs que les trois autres (Tukey HSD ;  $p < 0,01$ ). Le nombre de descendants était légèrement plus petit pour le croisement Q1♀ \* Q2♂ que dans les autres ( $p < 0,01$ ). La fertilité des F1 a été vérifiée en laissant les mâles et les femelles ensemble pour accouplement : la présence de mâles et femelles dans la descendance de ces F1 indique que celles-ci sont fertiles.

	Q1♀ * Q1♂	Q1♀ * Q2♂	Q2♀ * Q1♂	Q2♀ * Q2♂
Oeufs	127* ± 43	75 ± 13	40 ± 13	65 ± 9
Adultes émergents	32 ± 15	48* ± 9	32 ± 12	39 ± 8
Adultes émergents (%)	24* ± 14	65 ± 13	79 ± 5	62 ± 16

**TABLE 2.4** – Résultats des croisements expérimentaux. Nombre moyen d'oeufs par répétition ( $\pm \hat{\sigma}$ ), nombre moyen d'adultes émergents ( $\pm \hat{\sigma}$ ) et pourcentage moyen d'adultes émergents par rapport au nombre d'oeufs pondus ( $\pm \hat{\sigma}$ ). Les astérisques désignent les résultats significativement différents des autres (Tukey HSD,  $p < 0,01$ ). Le nombre de répétitions est de dix par croisement.

## Analyse du compartiment nucléaire

### Échantillonnage

Un total de 301 femelles provenant de 8 localités ont été collectées en 2010. Parmi celles-ci, 59 sont des cytotypes de référence (Grèce et Israël) et 242 proviennent de situations de mixité en France et en Espagne où ils sont en sympatrie (voir tableau 2.5). Toutes ont été collectées en serre sur différentes plantes hôtes (maraichères ou ornementales) précisées dans le tableau 2.5.

## Chapitre 2. Analyse du compartiment nucléaire des cytotypes

	Pays	Ville	Plante	Q1					Q2					Q1 + Q2				
				N	Ho	He	$F_{IS}$	$A^-$	N	Ho	He	$F_{IS}$	$A^-$	N	Ho	He	$F_{IS}$	$A^-$
Sympatrique	Espagne	Cabrils	Melon	5	0.41	0.42	0.14	2.3	26	0.38	0.49	0.25	4.4	31	0.39	0.51	0.24	4.6
			Tomate	7	0.33	0.43	0.31	3.3	33	0.35	0.49	0.30	4.3	40	0.35	0.50	0.30	4.6
		Viladecans	Tomate	21	0.45	0.58	0.25	4.6	10	0.43	0.57	0.29	3.3	31	0.44	0.60	0.26	4.9
	Concombre		40	0.41	0.59	0.33	4.1	7	0.44	0.55	0.27	3.3	47	0.41	0.60	0.30	4.3	
	France	Fréjus	Hibiscus	14	0.42	0.52	0.23	3.4	34	0.37	0.52	0.31	3.6	48	0.38	0.53	0.28	4.9
St L. du Var			Aubergine	30	0.39	0.52	0.30	3.9	15	0.40	0.55	0.33	4.1	45	0.39	0.55	0.32	4.9
Référence	Grèce	Thessaloniki	Concombre	10	0.28	0.56	0.52	3.4	0	-	-	-	-	-	-	-	-	-
		Kasteli-Chania	Aubergine	10	0.38	0.56	0.34	3.4	0	-	-	-	-	-	-	-	-	-
	Israël	Ierapetra	Concombre	10	0.43	0.54	0.21	3.4	0	-	-	-	-	-	-	-	-	-
		Hof Carmel	Divers	0	-	-	-	-	29	0.4	0.52	0.37	4.3	-	-	-	-	-

**TABLE 2.5** – Diversité génétique des données microsatellite des populations de référence ainsi que des populations mixtes dans l'aire de sympatrie. N : nombre d'individus ; Ho : hétérozygotie observée ; He : hétérozygotie attendue non biaisée ;  $F_{IS}$  : indice de fixation de Weir & Cockerham (1984) ;  $A^-$  : nombre moyen d'allèles par locus

### Allèles de résistance

Tout d'abord nous avons considéré seulement les cytotypes de référence en Israël (Q2) et en Grèce (Q1). Les résultats sont consignés dans le tableau 2.6. L'allèle de résistance F331W sur le gène *ace1* et les deux allèles sur le gène du canal sodium (L925I et T929V) ont été détectés dans les deux cytotypes de référence le Q1 et le Q2 (voir le tableau 2.6). F331W est fixé dans les deux cytotypes ; L925I est présent à hauteur de 0,70 et 0,74 respectivement chez Q1 et Q2, ce qui n'est pas significativement différent (test exact de Fisher,  $p > 0,05$ ) ; la fréquence de l'allèle T929V diffère selon le cytotype : 0,02 chez les Q1 contre 0,29 chez les Q2 (test exact de Fisher,  $p < 0,05$ ).

Maintenant, considérons les populations sympatriques où l'on retrouve les populations mixtes Q1-Q2 en France et en Espagne, et regardons l'évolution de ces fréquences. Dans ces populations, (111 femelles génotypées), les trois mutations ont été détectées, dans les deux cytotypes Q1 et Q2 et dans toutes les populations avec des fréquences très variables. Tous les individus sont homozygotes sur le gène *ace1*, l'allèle de résistance *R* est fixé. La mutation L925I (allèle r1) est très fréquente (0,56 à 1) et la fréquence de la mutation T929V (allèle r2) est plutôt basse (0 à 0,30). Ces fréquences ne diffèrent pas entre les deux cytotypes, dans les quatre populations (test exact de Fisher,  $p > 0,05$ ). Aucun écart à l'équilibre de Hardy-Weinberg n'a été détecté dans trois populations (quelle que soit le cytotype) pour les deux mutations du gène du *Canal sodium* (Fréjus, Saint-Laurent du Var et Viladecans ;  $p > 0,05$ ). Seule la population de Cabrils (Espagne) n'est pas à l'équilibre pour la mutation T929V ( $p < 0,05$ ).

### Analyse microsatellite

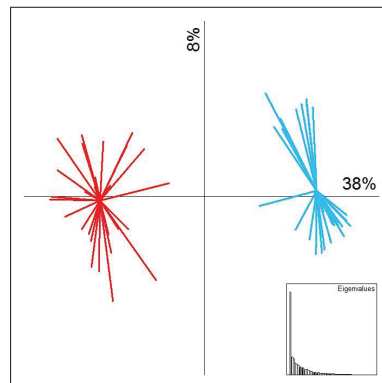
A nouveau, nous allons tout d'abord analyser les données des cytotypes de référence (Israël et Grèce), puis nous comparerons avec ceux en sympatrie. L'analyse microsatellite des 59 individus génotypés (30 provenant de trois populations en Grèce et 29 provenant d'une population Israélienne) révèle une richesse allélique légèrement plus grande dans le cytotype Q1 que dans

#### 4. Le cas des cytotypes Q1 et Q2 en Europe

	Pays	Villes	Cytotype	<i>Canal sodium</i>										<i>ace1</i>					
				Locus1 (L925I) genotypes				Locus 2 (T929V) genotypes						Genotypes					
				n	r1r1	r1s1	s1s1	$f_{r1}$	n	r2r2	r2s2	s2s2	$f_{r2}$	n	RR	RS	SS	$f_R$	
Réf.	Grèce	Diverses (3)	Q1	29	16	11	2	0,74	29	0	1	28	0,02	29	29	0	0	1	
	Israël	Hof-Carmel	Q2	23	12	8	3	0,70	27	3	11	13	0,29	27	27	0	0	1	
Sympatriques	Espagne	Cabrils	Q1	7	3	4	0	0,71	7	0	3	4	0,21	7	7	0	0	1	
			Q2	25	6	16	3	0,56	25	0	15	10	0,30	25	25	0	0	1	
		Viladecans	Q1	28	12	14	2	0,68	28	0	6	22	0,11	28	28	0	0	1	
			Q2	9	6	2	1	0,78	9	0	3	6	0,17	9	9	0	0	1	
	France	St-L. du Var	Q1	12	11	1	0	0,96	12	0	0	12	0	12	12	0	0	1	
			Q2	7	6	1	0	0,93	7	0	0	7	0	7	7	0	0	1	
			Frejus	Q1	8	8	0	0	1	8	0	0	8	0	8	8	0	0	1
				Q2	15	12	3	0	0,90	15	0	2	13	0,07	15	15	0	0	1

**TABLE 2.6** – Fréquences alléliques des gènes de résistance aux insecticides. Les populations sont triées par type : cytotypes de référence (Réf.) et zone sympatrique. Pour chaque sexe et gène, les génotypes sont indiqués, puis la fréquence de l'allèle de résistance pour les femelles (les mâles étant haploïdes, la fréquence du génotype résistant est égale à la fréquence de l'allèle).  $n$  : nombre d'individus,  $r1$ ,  $r2$  et  $s$  : respectivement les deux allèles de résistance ainsi que l'allèle sensible du gène *Canal sodium*, R et S : respectivement allèles de résistance et sensible du gène *ace1*.

le Q2 (4,3 contre 3,4 ; voir tableau 2.5). L'analyse en composante principale montre des cytotypes très structurés entre pays, la part de variance expliquée par les deux premiers axes étant respectivement 38 et 8 %. L'analyse en composante principale confirme les résultats obtenus par le logiciel STRUCTURE (résultats non présentés ici), il y a deux groupes génétiques distincts, sans hybrides, et très différenciés : l'un est formé par les populations de Grèce (Q1) l'autre par les populations d'Israël (Q2) La valeur de  $F_{ST}$  entre les deux groupes atteint 0,27.

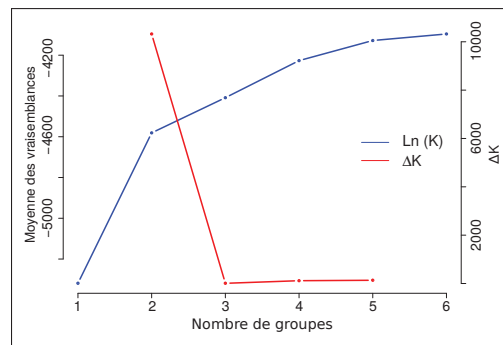


**FIGURE 2.5** – Analyse en composante principale sur les données microsatellites des cytotypes Q1 et Q2 dits de référence. Les Q1 (Grèce) sont en rouge, les Q2 (Israël) sont en bleu. Les pourcentages indiquent la part de variabilité expliquée par chaque axe.

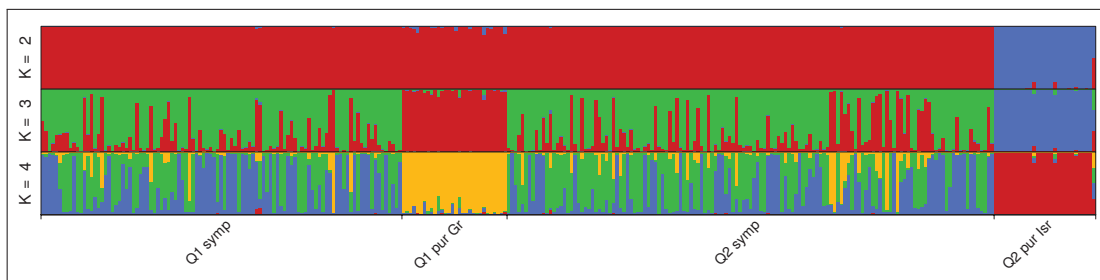
Ensuite nous avons ajouté les populations de l'aire sympatrique. La richesse allélique est plutôt homogène entre les populations mixtes (de 4,3 à 4,9), mais la différence n'est pas significative avec celles des populations de référence.

Toutes les populations de la région sympatrique montrent un écart à l'équilibre de Hardy-Weinberg ( $p < 0,001$ ). Un déficit systématique en hétérozygotes a été observé, les valeurs de  $F_{IS}$  allant de 0,243 à 0,357. Ce déficit n'est pas dû à un locus en particulier, l'hétérozygotie observée étant inférieure à l'attendu dans chacun d'entre eux. Dans les quatre localités (Cabrils,

Viladecans, Saint-Laurent du Var et Fréjus), le  $F_{IS}$  n'était pas supérieur quand tous les individus étaient considérés dans leur ensemble plutôt qu'en séparant les cytotypes, suggérant que les hautes valeurs de  $F_{IS}$  ne sont pas expliquées par une structuration génétique entre les cytotypes. Ce résultat a été confirmé avec l'analyse par le programme STRUCTURE révélant trois groupes génétiques distincts dans l'ensemble du jeu de données (populations de références et populations sympatriques). Le nombre  $K$  de groupes génétiques testés va de 1 à 6, les probabilités d'assignement sont représentées sur la figure 2.7 pour  $K = 2$  à  $K = 4$ . Il n'existe pas de méthode parfaite pour trouver le nombre optimal de groupes génétiques malgré l'utilisation de la vraisemblance ou du  $\Delta K$  de Evanno *et al.* (2005), il vaut mieux regarder les résultats pour plusieurs  $K$  et retenir celui qui paraît avoir le plus de signification biologique (Rosenberg *et al.* 2002), et le cas échéant discuter les autres. C'est pour cette raison que nous présentons ici les résultats pour trois  $K$  différents.



**FIGURE 2.6** – Vraisemblance en fonction du nombre de groupes génétiques, calculée par le logiciel STRUCTURE (en bleu), ainsi que le  $\Delta K$  (en rouge) calculé par la méthode de Evanno *et al.* (2005).



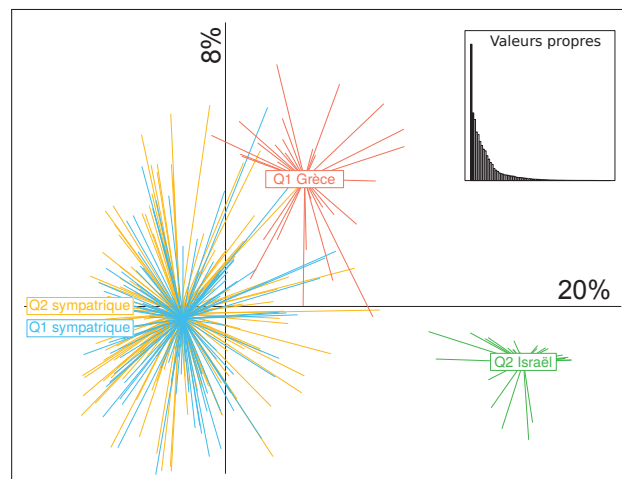
**FIGURE 2.7** – Probabilité d'assignement des individus génotypés par microsatellites, de  $K = 2$  (en haut) à  $K = 4$  (en bas). Chaque barre verticale représente un individu, la couleur indiquant le groupe auquel il est rattaché. Les individus sont groupés par cytotype (Q1 ou Q2) mais aussi par type : *symp* veut dire sympatrique et regroupe les individus espagnols et français où les populations sont mixtes ; *pur* signifie que les aires en Grèce et en Israël sont occupées seulement par l'un des deux cytotypes.

Le premier groupe est composé des individus d'Israël (Q2 référence). Les individus grecs (Q1 référence) forment un second groupe. Les individus sympatriques, qu'ils soient Q1 ou Q2, sont assignés avec des probabilités plus ou moins importantes soit au groupe grec, soit à un troisième groupe. Ces probabilités sont représentées sur la figure 2.7. Les cytotypes en sympatrie montrent ainsi une composition génétique similaire qui résulte probablement de l'hybridation de deux

fonds génétiques. L'un de ces fonds génétiques est apparenté aux populations grecques, l'autre n'est retrouvé que dans les populations de l'aire sympatrique.

L'analyse en composante principale (ACP) sur les fréquences alléliques individuelles confirme ces résultats (voir figure 2.8). Le premier axe sépare clairement trois groupes : les cytotypes en sympatrie, les Q1 de référence (Grèce) et les Q2 de référence (Israël). Quant au second axe, il discrimine faiblement les populations grecques des populations en sympatrie. Les mesures de  $F_{ST}$  par paire de ces trois groupes sont très élevées ( $0,09 < F_{ST}(\text{sympatriques-Grèce}) < 0,14$  ;  $0,29 < F_{ST}(\text{sympatriques-Israël}) < 0,33$ ). Le point important ici est que les cytotypes Q1 et Q2 sont homogènes sur le plan nucléaire lorsqu'ils sont en sympatrie.

L'effet de la plante hôte et du compartiment symbiotique sur la structuration génétique a été testé dans les quatre populations de l'aire sympatrique. L'AMOVA n'est pas significative ( $p > 0,1$ ) pour ces deux facteurs. Par contre la différenciation géographique est faible mais significative entre les paires de populations inter-pays (France *versus* Espagne) : les valeurs de  $F_{ST}$  sont comprises entre 0,01 et 0,02 ( $p < 0,05$ ).



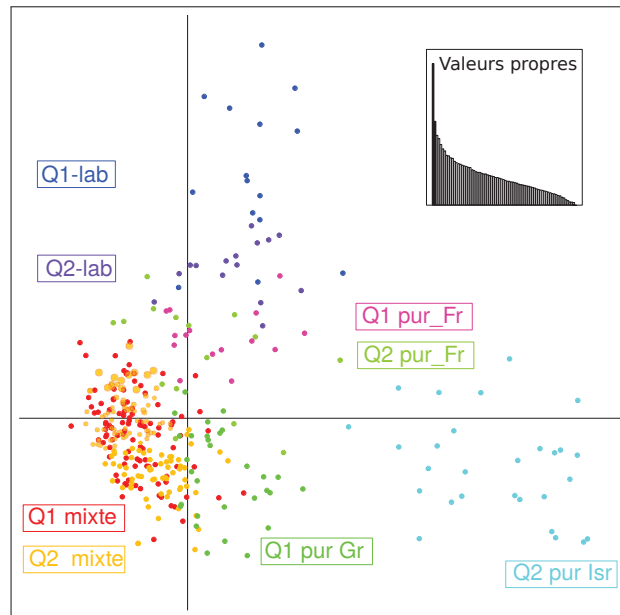
**FIGURE 2.8** – Analyse en composante principale sur les données microsatellites des cytotypes Q1 et Q2 en sympatrie, ainsi que des cytotypes dit de référence.

### Comparaison de populations mixtes et pures dans l'aire sympatrique

Après avoir utilisé les marqueurs microsatellites pour l'étude des flux génétiques entre Q1 et Q2 autour du bassin méditerranéen, nous avons voulu comparer des situations de sympatrie entre ces deux cytotypes et des situations où ceux-ci sont en allopatrie afin de voir si le compartiment nucléaire variait entre ces deux situations. Nous avons donc génotypé d'autres individus : une population 100 % Q1 provenant de Moulès dans les Bouches-du-Rhône et une population 100 % Q2 provenant de Cheval-Blanc dans le Vaucluse. Nous avons aussi utilisé une lignée Q1 et une lignée Q2 du laboratoire. Quinze femelles pour chacune de ces populations ont été analysés avec les mêmes sept locus microsatellites que pour les deux cas précédent. Nous avons effectué une analyse en composante principale et utilisé le logiciel STRUCTURE pour déterminer la structure



de nos populations ainsi que le nombre de groupes à l'équilibre en incluant les populations européennes précédemment analysées.



**FIGURE 2.9** – Analyse en composante principale sur les fréquences alléliques de toutes les populations génotypées par microsatellites en Europe. Les axes représentent respectivement 6,6 et 3,9 % de la variabilité totale. Les individus Q1 en populations mixtes (ceux précédemment analysés) sont regroupés sous l'ensemble Q1mixte, de manière analogue pour les Q2 (Q2mixte). Q1-lab : Population Q1 de laboratoire, Q2-lab : Population Q2 de laboratoire. Les deux populations de référence sont aussi indiquées : Q1purGr et Q2purIsr, qui sont respectivement les deux cytotypes de référence Q1 et Q2 en Grèce et en Israël. Les deux populations françaises où l'on retrouve seulement l'un des deux cytotypes sont désignées avec l'acronyme « allo », pour allopatrique.

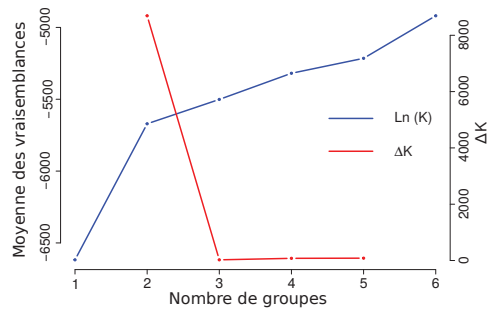
Les résultats de l'analyse en composante principale montrent toujours le gradient sur le premier axe depuis la population Q2 pure en Israël jusqu'aux populations sympatriques en France et en Espagne, en passant par le Q1 pur en Grèce. Les nouvelles populations génotypées se démarquent sur le deuxième axe. Le long de celui-ci, en partant des populations précédentes en sympatrie, on trouve la population Q2 allopatrique en France, puis la population Q1 allopatrique en France, la lignée Q2 de laboratoire et enfin la lignée Q1 de laboratoire. Ainsi les populations allopatriques françaises semblent aussi différentes des populations sympatriques que des populations grecques purement Q1.

Après cette analyse préliminaire, le logiciel STRUCTURE a été utilisé avec les mêmes données. Le nombre  $K$  de groupes génétiques variant de 1 à 6, la vraisemblance des résultats ainsi que le  $\Delta K$  de Evanno *et al.* (2005) sont sur la figure 2.10. Les probabilités d'assignement sont représentées sur la figure 2.11.

Ces données nous indiquent un nombre optimal de groupes génétiques égal à deux pour la méthode de Evanno *et al.* (2005), par contre la vraisemblance des résultats calculés par STRUCTURE indique un signal moins clair puisqu'elle augmente encore beaucoup après  $K = 2$ .

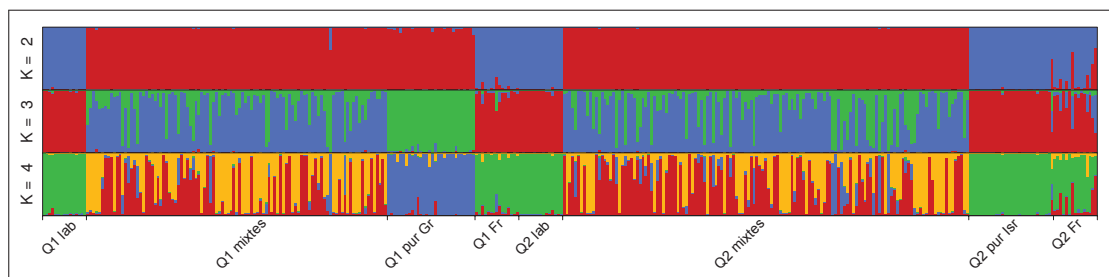
Si on observe les résultats de STRUCTURE pour un nombre de groupes génétiques égal à 2,





**FIGURE 2.10** – Vraisemblance en fonction du nombre de groupes génétiques, calculée par le logiciel STRUCTURE (en bleu), ainsi que le  $\Delta K$  calculé selon Evanno *et al.* (2005) (en rouge).

les populations mixtes de l'aire sympatrique forment un ensemble avec les populations de référence en Grèce, toutes les autres populations en formant un deuxième. À  $K = 3$ , on a un premier ensemble formé par les populations grecques, un deuxième regroupant les populations allopatriques, les populations de laboratoire ainsi que la population pure Q2 en Israël. Le dernier groupe est formé par les populations sympatriques. Notons que les individus de ces dernières sont assignés avec des probabilités plus ou moins grandes au groupe grec ce que l'on avait vu précédemment. Si on augmente encore la valeur de  $K$ , on différencie le fond nucléaire grec du fond nucléaire des populations mixtes qui était assigné à ce même fond nucléaire grec.



**FIGURE 2.11** – Probabilité d'assignement des individus génotypés par microsatellites, de  $K = 2$  (en haut) à  $K = 4$  (en bas). Chaque barre verticale représente un individu, la couleur indiquant le groupe auquel il est rattaché. Les individus sont groupés par cytotype (Q1 ou Q2) mais aussi par type : *mixtes* populations où l'on retrouve les deux cytotypes. *pur* signifie que les aires en Grèce (Gr) et en Israël (Isr) sont occupées seulement par l'un des deux cytotypes, *lab* signifie laboratoire. Les deux populations Fr Q1 et Q2 sont des populations françaises où l'on retrouve seulement l'un des deux cytotypes.

Les marqueurs microsatellites semblent ici montrer leur limite puisqu'il n'y a aucune justification biologique au groupe formé par les populations allopatriques, de laboratoire et d'Israël. Cela semble plutôt indiquer un problème de reproductibilité des analyses microsatellites où même un problème inhérent au marqueur qui peut montrer de l'homoplasie. L'homoplasie est le fait que deux allèles microsatellites peuvent avoir la même longueur, mais avec des séquences différentes du fait d'insertion ou de délétions. Ce phénomène est largement démontré et même pour le modèle *B. tabaci* (Dickey *et al.* 2013).

## Conclusion sur le modèle Q1-Q2 en Europe de l'ouest

Jusqu'en 2007, Q1 et Q2 étaient allopatriques à l'échelle du bassin méditerranéen, Q2 étant présent seulement en Israël et dans les pays voisins alors que Q1 avait une plus large distribution allant de la Grèce à la France et l'Afrique du nord. Cette séparation spatiale explique probablement la forte différenciation génétique observée sur les microsatellites entre cytotypes de référence (Q2 en Israël et Q1 en Grèce) et le grand nombre d'allèles spécifiques du Q2 de référence. Cette forte différenciation observée est comparable à celle observée entre des espèces distinctes du complexe *B. tabaci*, par exemple entre MED et MEAM1 (Saleh *et al.* 2012 ; Tahiri *et al.* 2013). De très nombreuses études de génétique des populations sur *B. tabaci* montrent des fortes valeurs de différenciations au sein d'un cytotype, par exemple Q1 en Grèce (de 6 à 43 %, Tsagkarakou *et al.* 2007), Q1 dans le bassin méditerranéen (Jusqu'à 43 % aussi, Gauthier *et al.* 2014) ou encore Simón *et al.* (2007) qui trouvent des valeurs s'échelonnant de 5 à 45 % pour des Q1. À l'intérieur de l'espèce potentielle MEAM1 au Brésil les  $F_{ST}$  peuvent être élevés (de 5 à 29 % Fontes *et al.* 2012).

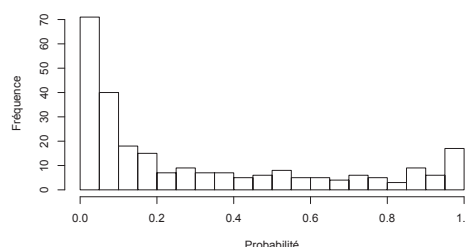
Étant donné le grand nombre d'allèles spécifiques au Q2 de référence, celui-ci n'est probablement pas la source des populations Q2 observées en France et en Espagne, mais cela n'exclue pas pour autant l'ensemble du territoire israélien comme source pour les populations ouest-européennes. Les effets de goulot d'étranglement liés à la colonisation (survie pendant le transport par exemple, voir l'introduction) pourraient expliquer une telle différenciation entre l'aire sympatrique et l'aire de référence.

Nous avons profité de cette nouvelle situation de sympatrie en France et en Espagne pour tester l'hybridation potentielle de ces deux groupes en condition naturelle (si l'on considère une serre en monoculture comme étant naturelle). L'analyse a mis en évidence trois groupes génétiques hautement différenciés comprenant respectivement les Q1 de référence, les Q2 de référence, et les populations de l'ouest avec les deux cytotypes en mélange. Ces résultats indiquent que les deux cytotypes sont homogènes lorsqu'ils sont en population mixte et démontrent ainsi l'hybridation totale ainsi que l'absence de barrières reproductives entre Q1 et Q2. Ils ne diffèrent pas non plus par la fréquence des allèles de résistance pour les deux gènes testés ici. Par contre, on a vu que certains individus des populations mixtes de l'aire sympatrique étaient plus apparentés aux Q1 de Grèce qu'au reste des Q1 de l'aire sympatrique (voir figure 2.7). La distribution des probabilités d'assignement au groupe grec pour tous les individus de France et d'Espagne (qu'ils soient Q1 ou Q2) est représentée sur la figure 2.12. On peut voir que ces probabilités sont réparties de façon assez continue de 0 à 1 (avec bien sûr beaucoup d'individus à zéro), ce qui montre des degrés d'introgessions assez variés et qui ne sont pas différents d'une population à l'autre. Cela suggère que des événements récurrents d'introduction de Q1 ont lieu dans ces serres, ce fond génétique étant assimilé indistinctement par les compartiments cytoplasmiques Q1 ou Q2 déjà présents.

Nous pouvons aussi faire l'hypothèse que le fond cytoplasmique Q2 est conservé mais le fond

nucléaire a été remplacé par celui de Q1. Ce résultat serait compatible avec ceux de Gauthier *et al.* (2014) qui montrent eux aussi une asymétrie dans l'hybridation au profit du fond nucléaire Q1 et du fond cytoplasmique Q2.

Les essais de croisements conduits au laboratoire confirment que Q1 et Q2 peuvent s'hybrider et produire une descendance fertile, même s'il faut rester prudent car les lignées utilisées (provenant de Lyon et Siegen) ont un fond nucléaire inconnu et sont donc probablement très différentes des cytotypes de référence en Grèce et Israël. On peut tout à fait imaginer un résultat différent avec ces populations dont le fond nucléaire est très différent. En tout cas, l'hypothèse la plus probable, sachant les résultats obtenus avec les croisements et les données génétiques des cytotypes en sympatrie est qu'il n'y a pas de barrières cytoplasmiques à l'hybridation et que ces deux cytotypes font partie de la même espèce. Le critère de différenciation mitochondriale (Dinsdale *et al.* 2010 ; De Barro *et al.* 2011) est pertinent dans ce cas là, contrairement à l'exemple des cytotypes en Afrique.



**FIGURE 2.12** – Probabilités d'assignement au groupe grec pour les individus des populations sympatriques, lorsque le nombre de groupe est fixé à trois.

## Discussion

### Les bactéries n'induisent pas d'incompatibilités

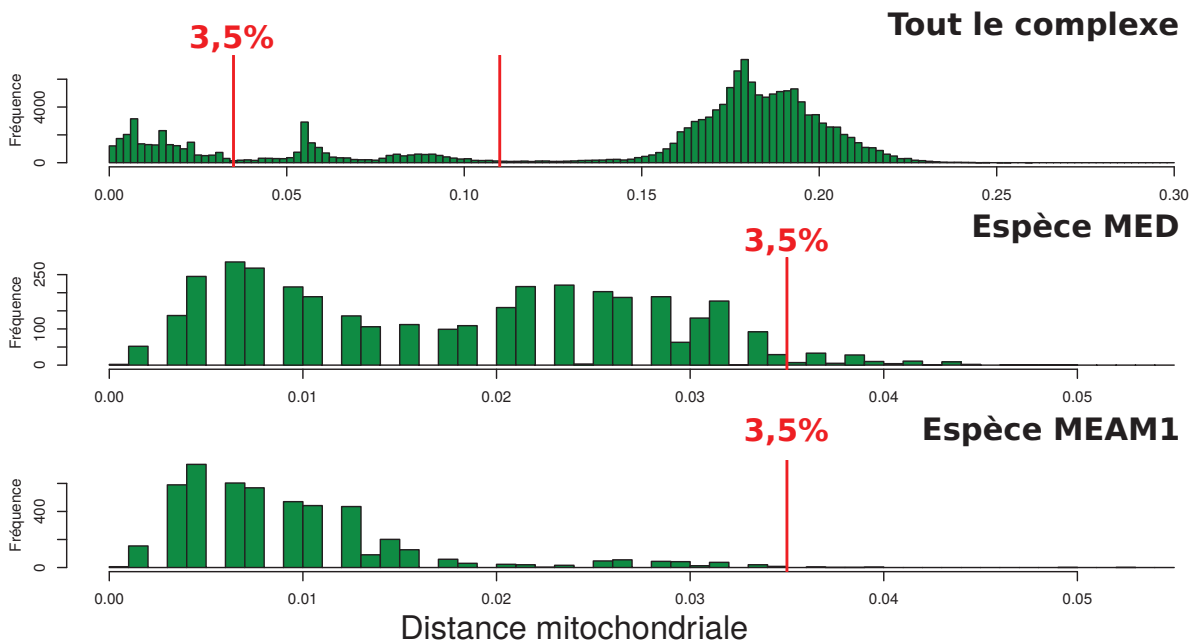
Nous avons vu que Q1 et Q2 peuvent s'hybrider malgré le fait qu'ils hébergent des communautés bactériennes différentes et à très forte prévalence. Cela suggère que ces bactéries n'induisent pas d'incompatibilité cytoplasmique chez *B. tabaci*, ou alors à un taux trop faible ne permettant pas de détecter ce phénomène dans nos conditions expérimentales de croisements contrôlés. Ces conclusions rejoignent celles que nous avons faites dans la première partie de ce chapitre traitant de la relation entre Q1 et ASL où nous ne suspicions pas la présence d'incompatibilité cytoplasmique. Néanmoins, des essais plus poussés en laboratoire avec une mesure plus précise du sexe-ratio dans les descendants des croisements contrôlés seraient nécessaire pour exclure ou déceler de l'incompatibilité cytoplasmique : en tout cas, s'il y a incompatibilité, elle est trop faible pour prévenir l'hybridation entre Q1 et Q2 ; et si tel était le cas entre ASL et Q1 ce serait un fait complètement nouveau pour ces bactéries.

### Taxonomie de l'espèce potentielle MED

Les résultats de ces deux études nous montrent les limites quant à l'utilisation du gène mitochondrial *COI* pour la systématique de *B. tabaci*, et notamment le seuil proposé de 3,5 % comme critère de séparation spécifique par Dinsdale *et al.* (2010). Le fait qu'il y a une rupture dans les distributions de paires de distances mitochondriales n'empêche pas que des barrières

## Chapitre 2. Analyse du compartiment nucléaire des cytotypes

reproductives peuvent exister dans certains groupes pour des valeurs de divergence plus faibles. Si on regarde la distribution des distances mitochondriales par paire dans l'espèce potentielle MED (voir figure 2.13), on observe aussi un décrochage, c'est-à-dire qu'il semble que cette distribution soit bimodale, ce qui pourrait indiquer des barrières reproductives dans ce clade, de la même manière qu'il y a une rupture dans la distribution des fréquences de tout le complexe *B. tabaci*. Par contre on ne retrouve pas cette distribution bimodale pour l'espèce potentielle MEAM1 (voir figure 2.13) qui est pourtant extrêmement diversifiée, bien que moins que MED (Firdaus *et al.* 2013).

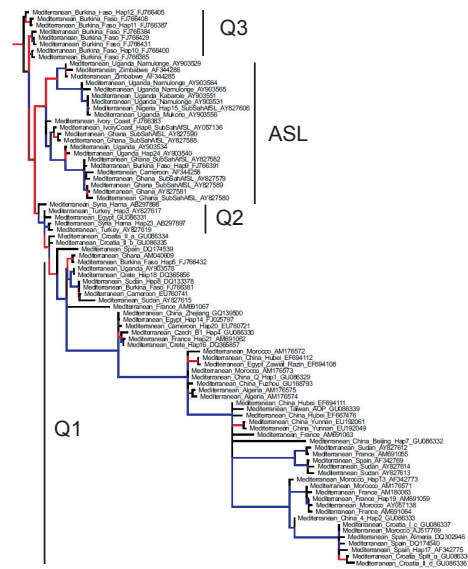


**FIGURE 2.13** – Distribution des distances mitochondriales entre paires d'individus dans le complexe d'espèces de *B. tabaci*, et dans les espèces potentielles MED et MEAM1. Le seuil de 3,5 % est indiqué en rouge. La distance est calculée selon Felsenstein (1984), méthode très proche de HKY85 utilisée dans Dinsdale *et al.* (2010)

Une étude avait déjà pointé les limites de ce seuil en montrant que certains haplotypes étaient divergents de plus de 3,5 % chez au moins 6 des espèces potentielles du complexe (Lee *et al.* 2013), montrant ainsi l'extrême diversité de certaines des espèces. A l'inverse, une distance supérieure à ce seuil n'implique pas forcément une barrière reproductive : l'hybridation est possible entre les espèces potentielles INDIAN OCEAN et MEAM1, pourtant divergentes de plus de 7 % sur le gène *mtCOI* (Thierry *et al.* 2011).

La figure 2.14 montre la diversité observée en 2013 dans l'espèce potentielle MED (Boykin *et al.* 2013) et par la même occasion on peut voir que selon cette phylogénie l'haplogroupe ASL serait plus proche du groupe formé par Q1 et Q2 que de Q3, qui pourtant semble s'hybrider avec Q1 en Afrique de l'ouest (données préliminaires sur microsatellites et croisements, *com. pers.*).

Cette phylogénie nous indique en plus l'extrême diversité observée dans le groupe Q1, on peut alors légitimement se poser la question des potentielles barrières reproductives existantes entre



**FIGURE 2.14** – Phylogénie de l'espèce potentielle MED (*mtCOI*), extraite de la phylogénie globale de *B. tabaci*, par Boykin *et al.* (2012). Tous les haplotypes disponibles dans les bases de données au moment de l'étude ont été utilisés. Les haplogroupes (Q1, Q2, Q3 et ASL) sont indiqués.

des populations Q1 du Burkina-Faso et d'autres présentes autour du bassin méditerranéen.

De plus, comme nous l'avons vu dans le paragraphe précédent, l'emploi des marqueurs mitochondriaux doit se faire avec précaution car ceux-ci sont soumis à des balayages sélectifs induits par les bactéries symbiotiques avec lesquelles ils sont en déséquilibre de liaison (Hurst & Jiggins 2005). En effet les gènes mitochondriaux subissent de la sélection indirecte en raison des symbiotes transmis verticalement qui envahissent les populations. Le déséquilibre de liaison entre mitochondrie et bactéries révèle donc l'histoire de l'infection plutôt que l'histoire évolutive du clade considéré. De plus, le transfert horizontal de ces bactéries dans une population suivi de leur invasion peut provoquer la diffusion de lignées mitochondriales. Ceci est dû à leur capacité à manipuler la reproduction (que nous n'avons pas détectée) et au rôle qu'elles jouent dans les populations de leur hôte.

Une approche plus juste serait d'utiliser l'ADN mitochondrial pour décrire la biodiversité et non pour des analyses phylogénétiques ou comme outil de diagnostic des barrières reproductives (Galtier *et al.* 2009). Pour ces dernières analyses, et comme pour l'identification des espèces du complexe *B. tabaci*, l'utilisation de marqueurs nucléaires est plus appropriée. Les marqueurs AFLP (Amplification Fragment Length Polymorphism) ont été utilisés historiquement dans l'étude de *B. tabaci*, notamment pour identifier le biotype présent sur Cassava (Abdullahi *et al.* 2003), qui se trouve maintenant être une des espèces SSA, ainsi que pour décrire la variabilité du complexe *B. tabaci* en général (Cervera *et al.* 2000 ; Guo *et al.* 2012). Par contre ce marqueur a montré ses limites dans la résolution qu'il apporte pour des études de génétique des populations (Zhang *et al.* 2005). Les microsatellites sont les plus couramment utilisés pour *B. tabaci*, aussi bien pour de la génétique des populations que dans certains cas pour l'identification de barrières reproductives, comme dans notre étude. L'avenir proche se situe probablement dans l'utilisation

des marqueurs d'outil de génotypage haut-débit tels que les marqueurs *RADSeq* (Restriction site Associated DNA sequencing), qui permettent de génotyper des individus à des (dizaines de) milliers de locus et ont de nombreuses applications dans les études phylogénétiques, de génétique des populations ou de QTL pour des organismes non modèles en particulier (Baxter *et al.* 2011). Nous verrons dans le chapitre 4 la mise en place et l'utilisation de ces marqueurs pour *B. tabaci*. Dans le paragraphe suivant nous allons voir en quoi l'utilisation des microsatellites peut s'avérer problématique ou plus simplement ne pas être suffisant pour certaines études.

### Limites des microsatellites, vers de nouveaux marqueurs

À l'aide de marqueurs microsatellites nous avons pu mettre en évidence des barrières reproductives entre les cytotypes en Afrique de l'ouest et nous avons pu montrer l'hybridation des cytotypes en Europe de l'ouest. Nos marqueurs ont donc été suffisant pour répondre à ces questions.

Un problème qui se pose par contre avec les marqueurs microsatellites est leur puissance à une échelle plus fine. En effet, dans ce chapitre on a pu discriminer des populations provenant de pays différents (Espagne *versus* France, ou encore plus facilement avec les populations de l'est méditerranéen). Par contre, il est très difficile de discriminer les populations provenant de deux villes différentes à l'intérieur d'un pays. Par exemple Saint-Laurent-du-Var n'est pas différenciable de la population de Fréjus, c'est le cas aussi pour Viladecans et Cabrils en Espagne, même si on ne peut exclure la présence de flux génétique entre les deux.

On peut se demander si pour des études précises de reconstruction de routes d'invasions ou de suivi temporel de certaines populations ces marqueurs sont pertinents surtout que nous n'avons que sept locus, en tout cas dans un contexte où se développent d'autres marqueurs tels que les *RADSeq*.

De plus, retracer les mouvements des populations n'est pas la seule chose qui est envisagée et envisageable lors de l'étude du compartiment nucléaire. On peut chercher à détecter des régions du génome sous sélection, notamment en contexte invasif, ou détecter de la coadaptation entre bactéries et gènes nucléaires. En effet, les bactéries symbiotiques des cytotypes Q1 et Q2 ne semblent pas induire d'incompatibilité cytoplasmique chez *B. tabaci*, puisque ceux-ci s'hybrident alors que leurs bactéries respectives sont différentes. Pourtant elles sont maintenues de façon stable dans les populations (et mêmes fixées). On peut donc se poser la question de la coadaptation entre ces bactéries et le fond génétique nucléaire. Une façon de détecter de telles co-adaptations et de chercher du déséquilibre de liaison entre les bactéries symbiotiques et certains locus dans le génome de l'insecte. Pour cela il faut pouvoir génotyper un très grand nombre de loci dans tout le génome. L'utilisation des microsatellites n'est pas envisageable car nous avons seulement sept locus, et même dans l'hypothèse où nous utiliserions tous ceux disponibles dans la littérature et en développerions d'autres, nous ne pourrions augmenter ce nombre que d'un ordre de grandeur au maximum. Les marqueurs *RADSeq* prennent alors tout

leur sens puisqu'ils permettent de génotyper des dizaines de milliers de locus à moindre coût. Nous verrons ces développements dans le dernier chapitre.





# Chapitre 3

## Émergence d'un nouveau cytotype

1	Problématique	109
2	Introduction	110
3	Identification de la souche de <i>Rickettsia</i> associée à Q1	112
3.1	Méthode : MLST	112
3.2	Résultats	113
4	Transfert expérimental	114
4.1	Transfert de <i>Rickettsia</i> lors de l'accouplement	114
4.2	Transfert de <i>Rickettsia</i> par l'intermédiaire de la plante hôte	116
4.2.1	Méthode	116
4.2.2	Résultats	116
5	Discussion	118
5.1	Le résultat d'un transfert horizontal	118
5.2	Fréquences des transferts	120
5.3	Stabilité de l'infection au niveau intra-individuel	121
5.3.1	Compétition intra-hôte entre bactéries	122
5.3.2	La capacité de <i>Rickettsia</i> à se transmettre à la génération suivante	123
5.3.3	Interactions avec l'hôte	123
5.4	Compétitivité du cytotype	124



## Problématique

Parmi les ravageurs de cultures, les complexes d'espèces sont potentiellement les plus nuisibles car le fait qu'ils soient composés de multiples entités procure autant d'occasions de recombinaisons à même d'augmenter encore leur potentiel adaptatif. L'admixture nucléaire est un phénomène aujourd'hui bien décrit en tant que facteur participant à la réussite de ravageurs. L'aspect qui nous intéresse ici est la présence de bactéries symbiotiques chez les espèces composant ce complexe. Ces bactéries se transmettent verticalement de la mère à la descendance (rarement de façon paternelle), mais parfois elles peuvent se transmettre horizontalement (ou latéralement) d'un individu infecté à un individu non infecté, ou infecté par une ou plusieurs autres bactéries. Si on considère plusieurs populations qui peuvent s'hybrider lors d'épisodes d'introductions, deux cas de figure se présentent. 1) Si les bactéries sont identiques et fixées dans les différents groupes qui peuvent potentiellement s'hybrider — il y a donc un seul cytotype —, les phénomènes d'hybridation ne permettront pas la création de nouvelles combinaisons. 2) Si ces bactéries ne sont pas les mêmes d'un groupe à l'autre, les échanges de bactéries entre différents hôtes peuvent créer de nouveaux cytotypes. C'est ce cas qui est bien sûr le plus intéressant.

Les mécanismes de transferts horizontaux de bactéries symbiotiques apparaissent comme nombreux et diversifiés : lors de l'accouplement chez le puceron *Acyrtosiphon pisum*, un mâle infecté par *Regiella insecticola* peut transmettre cette bactérie à la femelle (Moran & Degnan 2006). Des parasitoïdes peuvent transmettre *Regiella insecticola* et *Hamiltonella defensa* entre pucerons (*Aphis fabae*, Gehrler & Vorburger 2012) ou *Wolbachia* chez *B. tabaci* (Ahmed *et al.* 2015). Des parasitoïdes peuvent se transmettre *Arsenophonus* par l'intermédiaire d'un hôte (Duron *et al.* 2010). La plante hôte de certains insectes peut aussi servir de vecteur, par exemple chez *B. tabaci* (Caspi-Fluger *et al.* 2012). Ces transferts peuvent se produire entre des hôtes dont la parenté varie énormément : depuis le niveau intra-spécifique jusqu'au niveau inter-genre. C'est le cas de *Wolbachia* chez des guêpes (Vavre *et al.* 1999) ou des araignées (Baldo *et al.* 2008). *Arsenophonus* peut se transmettre entre différents aleurodes (Thao & Baumann 2004a ; Mouton *et al.* 2012) ou entre différentes guêpes parasitoïdes (Duron *et al.* 2010) .

En ce qui concerne les bactéries que l'on retrouve chez *B. tabaci*, toutes sont présentes dans de très nombreuses espèces d'arthropodes (majoritairement des insectes), montrant leur capacité à transférer. *Wolbachia* est présente chez 40 à 66 % des espèces d'insectes (Hilgenboecker *et al.* 2008 ; Zug & Hammerstein 2012 ; Weinert *et al.* 2015), *Cardinium* chez 6 à 13 % des espèces (Zchori-Fein & Perlman 2004 ; Weinert *et al.* 2015), *Rickettsia* infecterait environ un quart des espèces (Weinert *et al.* 2007, 2015). La fréquence d'*Hamiltonella* semble beaucoup plus restreinte mais est peu connue, comme celle d'*Arsenophonus* qui n'a pas vraiment été estimée mais qui infecte des taxons aussi divers que des guêpes, triatomines, psylles, aleurodes, pucerons, tiques, fourmis, etc (Nováková *et al.* 2009).

Du fait des nombreux transferts horizontaux observés et des nombreux mécanismes existants, on peut penser que les infections multiples sont très courantes chez les insectes. Ainsi la pré-

sence de multiples infections par différentes espèces de bactéries chez un même hôte est documentée chez les cicadelles (Chang & Musgrave 1972), cochenilles, aleurodes (Zchori-Fein & Brown 2002), pucerons (Unterman *et al.* 1989 ; Sandström *et al.* 2001), les mouches Hippoboscidae *Melophagus ovinus* (Chrudimský *et al.* 2012) et tsétsé ou encore les psylles (Subandiyah *et al.* 2000), mais est probablement sous estimée (Weeks *et al.* 2003).

On peut légitimement se poser la question de la fréquence de l'établissement de telles multi-infections, car ce phénomène met en jeu de fortes interactions entre les symbiotes, qui se situent pour la plupart dans le bactériome. Avec le modèle d'étude *B. tabaci*, et plus particulièrement les cytotypes Q1 et Q2 qui ont des cortèges symbiotiques très différents à transmission essentiellement verticale, nous allons pouvoir étudier la possibilité de transferts horizontaux grâce à la situation particulière trouvée en France où ces deux cytotypes se retrouvent en contact très étroit.

## Introduction

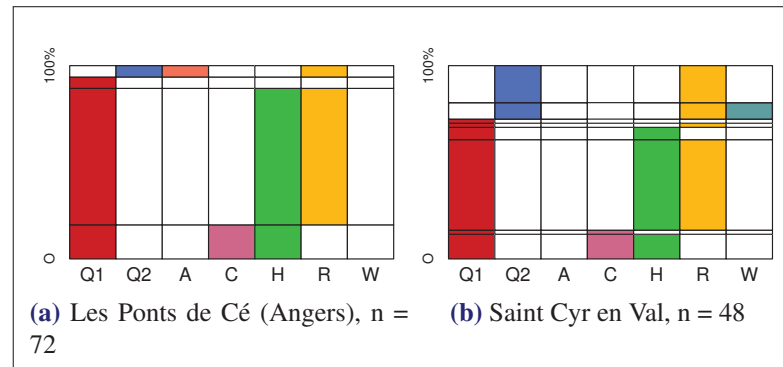
Dans les chapitres précédents nous avons décrit les cytotypes Q1 et Q2. Tous deux sont originaires du bassin méditerranéen, mais Q2 vient de l'extrême est alors que Q1 vient d'une large partie ouest. Q1 est associé à *Cardinium*, *Hamiltonella* et *Wolbachia*, alors que Q2 est associé aux bactéries *Arsenophonus*, *Rickettsia* et *Wolbachia*. En France, Q1 est le principal cytotype présent à partir de l'année 2003 jusqu'en 2007, avec l'apparition du cytotype Q2. Son arrivée ne provoque pas la disparition de Q1, mais plutôt une cohabitation car on observe de nombreuses populations dites mixtes où l'on retrouve les deux cytotypes. Cette situation originale de contact secondaire entre deux entités préalablement isolées géographiquement et différemment infectées crée des occasions de transfert de bactéries.

En 2012, un nouveau cytotype a été détecté simultanément en serre à Angers dans le département du Maine et Loire et à Saint-Cyr en Val dans le département du Loiret. Il faut noter que ces serres sont approvisionnées en plante par le même fournisseur, ce qui explique probablement la découverte simultanée de la nouvelle association dans ces deux lieux. Le cortège symbiotique découvert montrait l'association des bactéries *Hamiltonella* et *Rickettsia* à l'haplogroupe mitochondrial Q1, alors que *Cardinium* n'était plus détectable par PCR spécifique. Ces deux bactéries, *Rickettsia* et *Hamiltonella*, peuvent être rencontrées ensemble dans l'espèce potentielle MEAM1 (Chiel *et al.* 2007 ; Gueguen *et al.* 2010), c'est d'ailleurs pour celle-ci le profil symbiotique le plus courant.

En ce qui concerne le cytotype Q1, la présence d'*Hamiltonella* était habituelle, elle est même quasiment fixée comme nous l'avons vu dans le chapitre 1 ; par contre, la présence de la deuxième était nouvelle avec l'haplogroupe mitochondrial Q1. Quant à *Cardinium* sa présence n'est pas obligatoire dans le cytotype Q1, voir par exemple les populations de Grèce dans le chapitre 1.

Dans ces deux serres, la fréquence de cette nouvelle combinaison de bactéries parmi les haplo-

types Q1 est assez élevée, elle est de respectivement 75 % et 65 % . Dans ces mêmes populations, on trouve les cytotypes classiques Q1 et Q2 (voir figure 3.1).



**FIGURE 3.1** – Cortèges symbiotiques trouvés dans les deux serres où a été détectée la nouvelle association en 2012.  $n$  : nombre d’individus. Sur cette figure, l’haplogroupe mitochondrial est indiqué (Q1 ou Q2). A : *Arsenophonus*, C : *Cardinium*, H : *Hamiltonella*, R : *Rickettsia*, W : *Wolbachia*.

Notre hypothèse de travail est que Q1 était historiquement non infecté par *Rickettsia*. Sa présence résulterait donc d’un transfert horizontal depuis un hôte infecté. Or une étude récente a montré expérimentalement que cette bactérie peut se transmettre par la sève phloémienne chez *B. tabaci* depuis des hôtes infectés par *Rickettsia* vers des hôtes non infectés (Caspi-Fluger *et al.* 2012). La différence est que cette étude portait sur l’espèce potentielle MEAM1 et non MED comme dans notre cas, sachant que l’interaction entre *Rickettsia* et un autre fond nucléaire peut probablement donner des résultats phénotypiques très différents (Cass *et al.* 2015a). Une autre étude a montré la présence de *Rickettsia* dans le stylet, ce qui expliquerait pourquoi elle est transmise par la sève de la plante (Brumin *et al.* 2012). Nous pouvons donc faire l’hypothèse que *Rickettsia* a été transmise par l’intermédiaire de la plante.

Ensuite nous nous sommes posé la question de la provenance de *Rickettsia*, c’est-à-dire l’hôte porteur de l’infection initiale, et de nombreux candidats sont possibles. Par exemple les guêpes parasitoïdes du genre *Eretmocerus* utilisées en lutte biologique contre *B. tabaci* sont suspectées de transmettre *Rickettsia* (Chiel *et al.* 2009). Un autre agent de lutte biologique *Nesidiocoris tenuis* (Heteroptera : Miridae) utilisé contre *B. tabaci* est infecté par *Rickettsia* (Caspi-Fluger *et al.* 2014). D’autres insectes suceurs de sèves peuvent être porteurs de *Rickettsia* et potentiellement la transmettre par la sève : le puceron (Chen *et al.* 1996), la cochenille (Singh *et al.* 2013), les tétraniques (Hoy & Jeyaprakash 2005), les cicadelles (Davis *et al.* 1998) ou encore un cousin de *B. tabaci*, *Trialeurodes vaporariorum* même si la fréquence semble être très faible dans cette espèce (Skaljic *et al.* 2013).

Comme Q1 est en sympatrie avec Q2, infectée par cette même bactérie, on peut légitimement proposer celui-ci comme hôte donneur. Mais Q2 n’est pas le seul à porter *Rickettsia* dans ce complexe d’espèces même s’il nous paraît le plus probable : il y a à ce jour 10 espèces potentielles ou cytotypes qui sont concernés (voir figure 3.2). Pour vérifier que Q2 est bien le donneur et non pas un autre cytotype de *B. tabaci* ou encore tout autre espèce porteuse de *Ri-*

*ckettsia*, nous avons mis en place un protocole de MLST (Multi Locus Sequence Typing) afin de comparer différentes souches présentes chez différents hôtes. Le principe est de séquencer plusieurs gènes des bactéries que l'on veut comparer, afin de les différencier et le cas échéant de retracer leur histoire évolutive.

Pour le moment nous avons fait l'hypothèse que *Rickettsia* est transmise par la plante, mais des travaux ont montré que le transfert de bactéries symbiotiques était possible lors de l'accouplement de leur hôte, depuis le mâle infecté vers la femelle non infectée. Bien sûr dans ce cas là, ne se pose plus la question de savoir si c'est la même espèce ! Ce mode de transfert a été mis en évidence chez le puceron *Acyrtosiphon pisum*, avec les bactéries *Hamiltonella defensa* et *Regiella insecticola* (Moran & Dunbar 2006) ; chez la cicadelle *Scaphoideus titanus* avec la bactérie *Asaia* (Gonella *et al.* 2012), chez le moustique *Anopheles stephensi* avec une nouvelle fois la bactérie *Asaia* (Damiani *et al.* 2008). Plus récemment, le transfert de *Sodalis glossinidius* lors de l'accouplement chez la mouche tsétsé *Glossina morsitans morsitans* a été démontré (De Vooght *et al.* 2015). Nous testerons donc ce vecteur de transmission de *Rickettsia* à l'aide d'un essai au laboratoire.

Pour caractériser le mode de transfert, nous avons donc mis en place deux protocoles pour tester la possibilité du transfert. 1) Lors de l'accouplement 2) Par l'intermédiaire de la plante hôte Cette deuxième expérience a été divisée en deux sous parties. i) Nous avons testé la possibilité d'un transfert d'un individu infecté vers un individu non-infecté *via* la plante. ii) Nous avons vérifié qu'une telle infection peut se maintenir par transmission verticale au fil des générations.

## Identification de la souche de *Rickettsia* associée à Q1

### Méthode : MLST

Nous avons utilisé 31 femelles provenant de huit localités, sélectionnées pour identifier les *Rickettsia* provenant de différents cytotypes : Q1 (la nouvelle association détectée), Q2, Q3 (tous les trois appartenant à l'espèce potentielle MED), ainsi que les espèces potentielles SSA 2, et MEAM 1 (voir le tableau 3.1). Neuf gènes cibles (*atpA*, *dnaA*, *dnaG*, *dnaK*, *ftsY*, *gatB*, *gltA*, *recA* et *rpoB*) ont été amplifiés par PCR car classiquement utilisés dans ce but (Zhu *et al.* 2005 ; Leclerque *et al.* 2011). Des amorces de PCR ont été dessinées à l'aide des données du génome disponible de *Rickettsia bellii* qui est annoté, ainsi que des *contigs*<sup>1</sup> provenant de l'assemblage du génome de la *Rickettsia* de *B. tabaci* (com. pers.). En plus de ces neuf gènes, cinq régions intergéniques ont été séquencées car *a priori* plus variables. Cette dénomination de intergénique provient de l'annotation de la plateforme MicroScope (Microbial Genome Annotation & Analysis Platform) du génoscope.

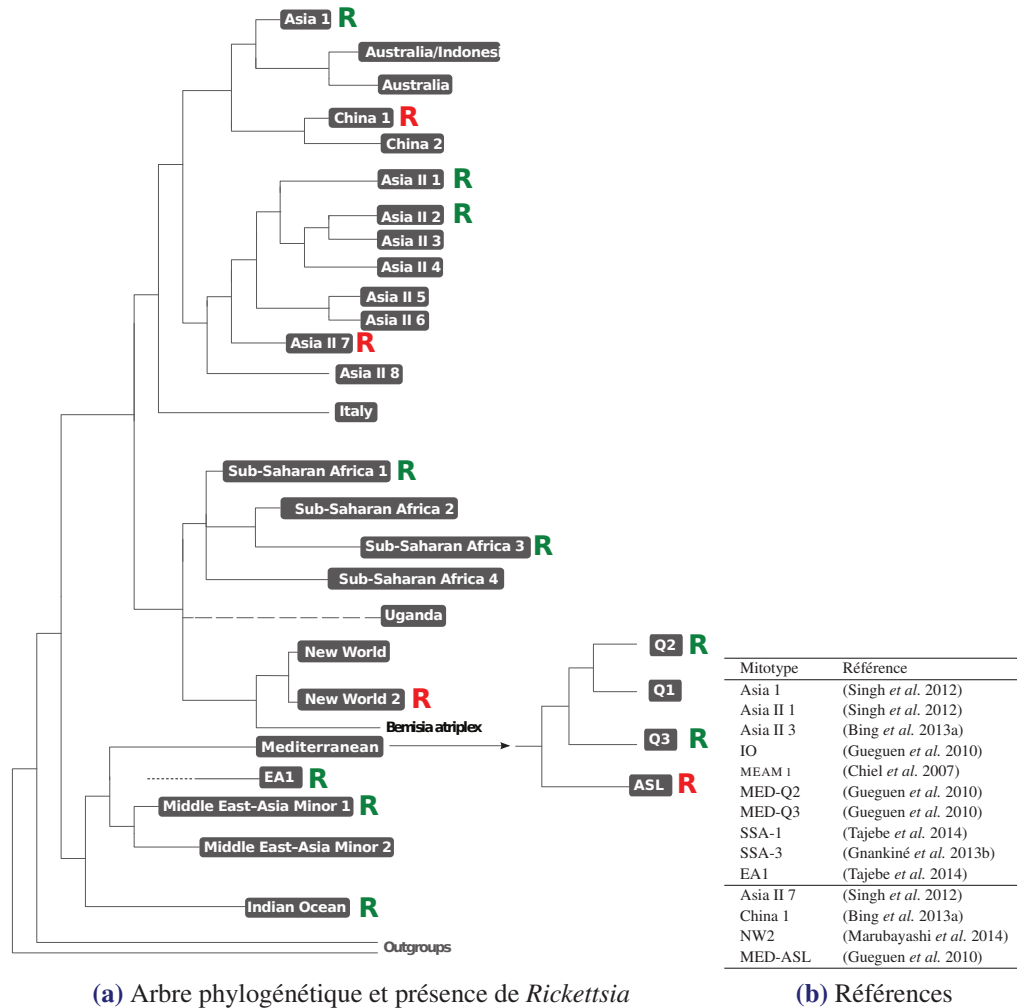
Une PCR a été réalisée pour chacun des gènes dans les conditions suivantes : 200  $\mu$ M de dNTP,

---

1. Un *contig* est un ensemble de séquences nucléiques qui forment une région consensus d'une partie d'un génome.



### 3. Identification de la souche de *Rickettsia* associée à Q1



**FIGURE 3.2** – Arbre phylogénétique du complexe d’espèces *B. tabaci*, réalisé sur le gène *mtCOI* indiquant la présence de *Rickettsia* dans le clade. En vert, *Rickettsia* a été trouvée, en rouge, elle n’a pas été trouvée, l’absence d’indication signifie qu’elle n’a pas été recherchée dans cette espèce. Les haplogroupes de l’espèce potentielle MED sont détaillés. Le placement de l’espèce EA 1 est incertain mais serait frère de l’espèce potentielle MED. La phylogénie est adaptée de De Barro *et al.* (2011).

200 nM de chaque amorce, 0,5 unité de Taq DNA et 2  $\mu$ L d’échantillon, avec le cycle suivant 94° C pendant 30s, 60° C 30s, 72° C 75s et une élévation finale pendant 10mn. Les produits de PCR ont été séquencés par méthode Sanger.

Les séquences obtenues ont été vérifiées avec le logiciel CLC DNA v6.0, et alignées avec l’algorithme MUSCLE (Edgar 2004) implémenté dans SEAVIEW (Gouy *et al.* 2010). Le placement phylogénétique de *Rickettsia*, réalisé avec l’algorithme PHYML aussi implémenté dans SEAVIEW, a été fait avec les 9 gènes puisque les séquences intergéniques ne sont pas disponibles pour les groupes externes.

## Résultats

Neuf gènes de *Rickettsia* ont été séquencés sur un total de 4 500 paires de bases, provenant de 31 individus appartenant à 5 cytotypes (Q1, Q2, Q3 de l’espèce MED, SSA 2 et MEAM 1).

Ville	Pays	Plante	Année	Espèce	Haplogroupe	R. Inter.	N
Ouagadougou	Burkina-Faso	Lantana	2013	MED	Q3	+	4
Les Ponts-de-Cé	France	Dipladenia	2012	MED	Q1	+	3
					Q2	+	3
Cheval-Blanc	France	Aubergine	2012	MED	Q2	+	3
Contamine/Arve	France	Lamium	2012	MED	Q2	-	3
Torre del Greco	Italie	Morelle noire	2012	MED	Q2	-	4
Mugla	Turquie	Tomate	2013	MEAM 1	-	+	4
Cabrils	Espagne	Chou	2011	MEAM 1	-	-	3
Montcada	Espagne	Maluve	2012	SSA 2	-	+	4

**TABLE 3.1** – Origine des femelles génotypées pour le MLST. Pour l'espèce potentielle MED, les cytotypes sont précisés. R. Inter. : régions intergéniques séquencées ou non, N : nombre d'individus.

Pour certains individus, nous avons aussi séquencé 5 régions intergéniques pour essayer de trouver plus de variabilité, c'est pourquoi ces séquences ne sont pas disponibles pour toutes les populations mais seulement pour certaines, représentant tout de même tous les cytotypes et espèces (voir tableau 3.1).

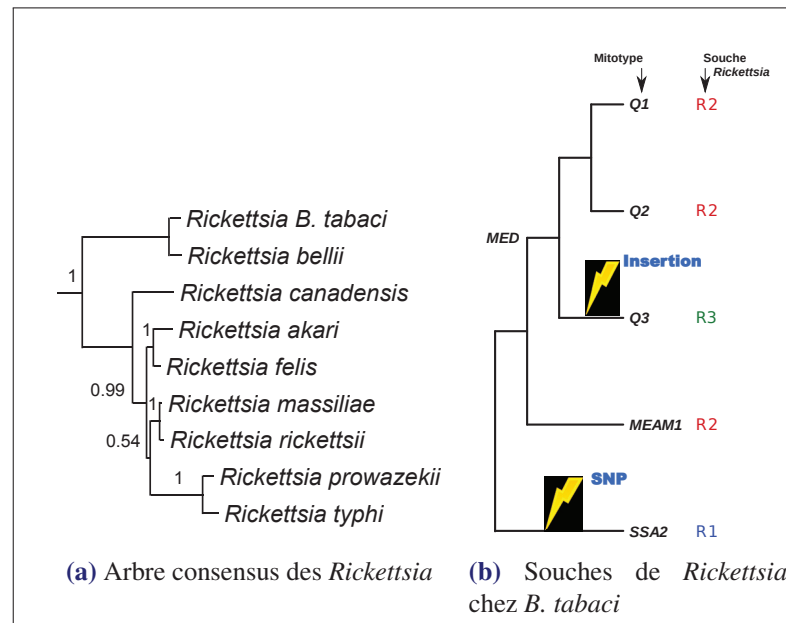
Nos résultats montrent que le groupe (*Rickettsia bellii* + *Rickettsia* de *B. tabaci*) est monophylétique (voire figure 3.3a), comme attendu (Gueguen *et al.* 2009).

Malgré le nombre relativement grand de bases séquencées, il y a très peu de variation à l'intérieur du clade *Bemisia*. On peut différencier seulement 3 haplotypes par une insertion de 15 bases dans le gène *atpA* de la *Rickettsia* du cytotype Q3 (confirmée avec les 4 individus) et un SNP détecté dans le gène *gltA* de la *Rickettsia* de SSA 2 (confirmé aussi sur les 4 individus). Cette très faible variabilité ne nous permet pas de discriminer toutes les souches comme nous aurions pu le souhaiter. En résumé nous avons trois souches de *Rickettsia* : SSA 2, Q3 et le dernier regroupant Q1, Q2 et MEAM 1 (voir figure 3.3b). La figure 3.3b indique que les phylogénies de l'hôte et des *Rickettsia* sont incongruentes. Une explication possible est la suivante (se reporter à la figure 3.3b) : la souche R2 est ancestrale et il y eu deux évènements de mutations : l'un dans la branche de SSA 2 (R1), l'autre dans la branche de Q3 (R3). Dans ce cas là, Q1 était historiquement infecté (impossible de dater cet état) et par la suite *Rickettsia* a été perdue (d'après nos données qui montrent son absence en populations naturelle) puis acquise de nouveau lors d'un transfert.

## Transfert expérimental

### Transfert de *Rickettsia* lors de l'accouplement

Le protocole suivant a été mis en place : une femelle Q1 non infectée par *Rickettsia*, mais infectée par *Cardinium* et *Hamiltonella* (désignée Q1CH) ainsi que deux mâles infectés par *Rickettsia*, sont récupérés à l'émergence, pour s'assurer de leur virginité, et sont placés dans une



**FIGURE 3.3** – A gauche, phylogénie des *Rickettsia* réalisée sur les neuf gènes. A droite, phylogénie des mitotypes de *B. tabaci* avec les évènements de mutation des *Rickettsia* correspondantes. Les souches (numérotées R1 à R3) de *Rickettsia* sont à droite.

enceinte fermée (un simple tube en verre fermé avec du coton) pour qu'ils s'accouplent. Deux cytotypes infectés par *Rickettsia* ont été utilisés pour les mâles. L'un provient de la nouvelle lignée Q1 d'Angers où a été détectée cette infection pour la première fois en 2012 et est infecté par *Rickettsia* et *Hamiltonella* (désigné Q1RH). L'autre cytotype provient de Bassens en Savoie, aussi prélevé en 2012, ce sont des Q2 infectés par *Rickettsia*, *Arsenophonus* et *Wolbachia* (désignés Q2RAW). Toutes ces lignées sont testées régulièrement par PCR pour vérifier le statut d'infection. Elles sont maintenues au laboratoire depuis au moins un an (environ 17 générations), sur *Hibiscus moscheutos* var. Galaxy à 25 °C. Les insectes placés dans le tube en verre ne disposent d'aucune source de nourriture pour éviter un potentiel transfert par le milieu nutritif (sève ou solution artificielle), vecteur qui sera l'objet d'un essai indépendant (voir partie suivante). Au bout de 24h, les femelles sont mises à pondre individuellement sur un disque foliaire d'Hibiscus, lui-même placé sur un milieu gélosé en boîte de Pétri. Au bout de 10 jours, l'ADN des femelles est extrait et la présence de *Rickettsia* testée par PCR spécifique (voir introduction). Cet essai préliminaire a été conduit avec 51 femelles Q1CH accouplées indépendamment à respectivement 33 et 18 paires de mâles Q1RH et Q2RAW. Parmi ces 51 femelles, 25 n'ont pas pondu et sont mortes assez rapidement, probablement à cause de la privation de nourriture. Sur les 26 femelles restantes, aucune n'était positive pour *Rickettsia*, infection qui n'a pas été détectée non plus dans la génération suivante. Cette expérience n'a pas été approfondie.

## Transfert de *Rickettsia* par l'intermédiaire de la plante hôte

### Méthode

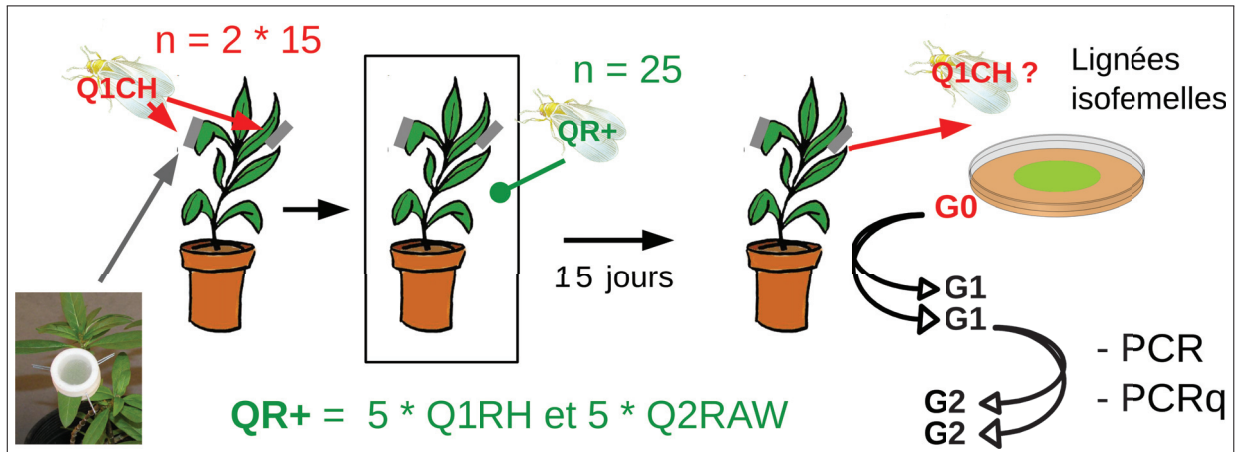
Trois lignées ont été utilisées pour cette expérience, une non infectée (les individus receveurs) et deux infectées par *Rickettsia* (appelés ici "donneurs"). Ce sont les mêmes que pour l'essai de transfert durant l'accouplement. Les receveurs sont de cytotype Q1 (infectée par *Hamiltonella* et *Cardinium* : Q1CH), les donneurs sont de cytotype Q1RH (infectés par *Rickettsia* et *Hamiltonella*), ou de cytotype Q2RAW (infectés par *Rickettsia*, *Arsenophonus* et *Wolbachia*). Trente femelles non infectées G0 (Q1CH) étaient réparties équitablement dans deux *clip-cage* (petit dispositif très léger, s'apparentant à une cage, que l'on peut fixer sur des feuilles, voir illustration en bas à gauche de la figure 3.4), chacune maintenue sur une feuille d'un même pied de coton. La plante est ensuite placée dans une cage avec 25 individus infectés par *Rickettsia*. Ceux-ci peuvent se déplacer librement sur toute la plante excepté dans la *clip-cage* contenant les individus non infectés, évitant ainsi tout contact direct. Deux cages contenant seulement des individus Q1CH servaient de contrôle. Deux semaines après, les individus initialement non infectés étaient retirés des *clip-cage*, les vivants dénombrés puis placés individuellement sur une feuille d'hibiscus, elle-même dans une boîte de Pétri sur milieu gélosé. Après cinq jours de ponte, l'infection des femelles G0 est vérifiée par PCR spécifique classique ainsi que par PCR quantitative (voir protocole ci dessous). Nous avons ensuite vérifié que l'infection pouvait se maintenir à la génération suivante et se transmettre de manière verticale : parmi les descendants, deux femelles G1 par boîte étaient mises à pondre (séparément) avec chacune deux mâles dans les mêmes conditions que précédemment, pour qu'elles soient fécondées. Après cinq jours, l'infection était testée sur ces femelles G1. A nouveau, deux descendantes G2 étaient testées, les positives étant conservées en lignées.

Cinq répétitions pour chaque cytotype ont été effectuées. Le protocole est schématisé sur la figure 3.4.

**Mesure de la densité symbiotique par PCR quantitative** La gamme étalon est établie grâce au Qubit et au réactif Quant-iT BR DNA Assay, dilué en cascades de  $10^6$  à  $10^0$  copies/ $\mu$ l. La PCR est réalisée dans un volume final de 10  $\mu$ l contenant 1x de *SSo Advanced Universal Probes Supermix* (Bio-Radn Hercules, California, USA), 500 nM de chaque amorce et 2  $\mu$ l d'ADN. Nous utilisons le fluorochrome SYBR green. Le programme est 95 °C 30 s, 95 °C 10 s, Tm 30 s 40 fois. L'amplification et la détection d'ADN sont effectués sur CFX96 (Bio-Rad).

### Résultats

Trois-cents individus du cytotype Q1CH ont été exposés à des individus infectés, pour moitié à des individus Q1RH et autant à des individus Q2RAW. Après 15 jours d'exposition, énormément d'insectes étaient morts dans les *clip-cage*, il restait en moyenne 6 individus (écart-type =



**FIGURE 3.4** – Test du transfert horizontal de *Rickettsia* entre individus par l’intermédiaire de la plante. Deux *clip-cage* (rectangle gris) sont placées sur un pied de coton, chacune contient 15 femelles non porteuses de *Rickettsia*. Cette plante est placée dans une cage pour éviter toute contamination extérieure. 25 individus porteurs de *Rickettsia* sont placés dans cette cage. Au bout de 15 jours les femelles non porteuses de l’infection par *Rickettsia* au début de l’essai sont retirées des *clip-cage* et placées individuellement sur un disque foliaire, lui-même sur milieu agar pour faire des lignées isofemelles. Après que les femelles ont pondues, leur ADN est extrait pour déterminer leur statut d’infection. Le protocole est répété à la génération suivante.

4,5) sur les 15 du début. 95 lignées isofemelles *G0* ont été réalisées, mais seulement 72 ont pu être testées par PCR quantitative. Parmi celles-ci, 20 se sont révélées positives pour *Rickettsia*, dont 12 provenant des cages avec Q1RH comme donneurs et 8 de cages avec Q2RAW comme donneurs. Cette différence entre les deux cytotypes donneurs est non significative (Test de  $\chi^2$  avec correction de continuité :  $\chi^2 = 1,2425$  ,  $ddl = 1$  et  $p = 0,27$ ). Sur ces 13 individus infectés par *Rickettsia*, 8 étaient aussi infectés par *Cardinium*, alors que les individus étaient tous au départ infectés par cette bactérie, certains ont donc perdu *Cardinium*. Nous avons testé si la probabilité d’infection simultanée par *Cardinium* et *Rickettsia* dans la *G0* était égal au produit des fréquences d’infection de chacune des bactéries de ces individus. Le test est significatif ( $\chi^2 = 11,315$ ;  $p < 0,001$ ) ce qui semble montrer qu’il y a moins de double infections qu’attendu, et qu’il y a une interaction négative entre les deux.

A la génération suivante, 8 *G1* provenant de donneurs Q1RH et 2 *G1* provenant de donneurs Q2RAW étaient toujours infectées par *Rickettsia*. Sur les trois qui n’étaient plus infectées par *Rickettsia*, une présentait le statut d’infection de départ, c’est-à-dire Q1CH, les deux autres étaient perdues par absence de descendance.

A la génération suivante, seulement 2 lignées étaient toujours infectées par *Rickettsia*. Ces deux infections ont pu être conservées ensuite pendant 10 générations environ en cage d’élevage, et seulement 1 au bout de 15. Ces résultats sont détaillés dans le tableau 3.2.

	G0		G1	G2	~G10	~G15
	R-	R+	R+	R+	R+	R+
Q1RH	46	12	8 (1CH +1†)	2 (3CH +3†)	2	1
Q2RAW	36	8	2 (1†)	0 (2†)	0	0

**TABLE 3.2** – Nombre de lignées infectées par *Rickettsia* à chaque génération testée. À la génération G0 les individus porteurs (R+) ou non (R-) de *Rickettsia* sont dénombrés pour chaque cytotype donneur (Q1RH ou Q2RAW). Pour les générations suivantes est indiqué le nombre de lignées encore porteuses de l'infection parmi les précédentes. Entre parenthèse est indiqué le nombre de lignées qui ne sont plus porteuses de l'infection, soit parce que il n'y a pas de descendants (†), soit parce que le statut d'infection est celui de départ (CH). Après la G2 les lignées infectées sont placées en cage à populations et à nouveau testées au bout d'environ 10 et 15 générations.

## Discussion

Dans ce chapitre nous avons décrit une nouvelle association bactéries-haplogroupe mitochondrial (Q1RH). Cette association étant très rare et nouvelle parmi toutes les données accumulées, nous allons discuter de l'opportunité de l'appeler cytotype.

### Le résultat d'un transfert horizontal

En introduction de ce chapitre nous avons vu que le cytotype Q1 était historiquement composé de l'haplotype mitochondrial du même nom associé aux bactéries *Cardinium* et *Hamiltonella*, nous l'appelons Q1CH. Quant au cytotype Q2, il est composé de l'haplotype mitochondrial Q2 associé aux bactéries *Rickettsia* et *Arsenophonus*. Ils peuvent aussi être infectés chacun par *Wolbachia* dans certaines populations. Lorsque nous avons découvert une association inhabituelle (Q1RH, voir figure 3.1 page 111) nous nous sommes posé la question de la provenance de *Rickettsia*. Sur la base de travaux récents (Caspi-Fluger *et al.* 2012) qui ont montré la capacité de *Rickettsia* à transférer *via* la plante chez *B. tabaci* MEAM 1, nous avons fait l'hypothèse d'un transfert horizontal récent depuis un hôte déjà infecté par *Rickettsia*, le plus probable étant Q2. Comme celui-ci est en sympatrie avec Q1 dans des populations mixtes depuis seulement quelques années, on peut faire l'hypothèse que cela a alors permis le transfert.

Pourtant, en regardant la figure 3.2 page 113, on serait tenté de dire que l'infection par *Rickettsia* est ancestrale dans le super-clade *Africa / Middle East / Asia Minor*. En effet, les 5 espèces potentielles de ce super-clade sont infectées par *Rickettsia*, et à l'intérieur de l'espèce potentielle MED, 2 des 4 haplogroupes le sont aussi (Q3 et Q2). L'hypothèse la plus parcimonieuse pourrait être une infection ancestrale suivi d'une perte dans les clades Q1 et ASL. Il n'y aurait pas forcément deux évènements de perte puisque la phylogénie de ces haplogroupes est peu résolue est rien n'exclue la monophylie Q1-ASL.

Néanmoins, deux éléments soutiennent l'hypothèse d'un transfert horizontal récent dans un Q1 non infecté par *Rickettsia*. i) Les données populationnelles ii) La faible divergence des souches



de *Rickettsia* comparé à la divergence des espèces de *B. tabaci*.

Tout d'abord, comme nous l'avons dit dans l'introduction de ce chapitre jamais Q1 n'a été retrouvé avec *Rickettsia* et ce dans de nombreux travaux sur de nombreux territoires : France (Gueguen *et al.* 2010 ; Terraz *et al.* 2014)<sup>1</sup>, Grèce (Tsagkarakou *et al.* 2012), Espagne (Terraz *et al.* 2014), Maroc (Gueguen *et al.* 2010 ; Tahiri *et al.* 2013), Tunisie (Gueguen *et al.* 2010 ; Gorsane *et al.* 2011), Italie (Parrella *et al.* 2013), Uruguay (Gueguen *et al.* 2010), Soudan (Gueguen *et al.* 2010), Cameroun (Gueguen *et al.* 2010), Bénin (Gnankiné *et al.* 2013a)<sup>2</sup>, Togo (Gnankiné *et al.* 2013a), Sénégal (Delatte *et al.* 2014), Chine (Bing *et al.* 2013a), la Réunion (Thierry *et al.* 2011, 2015) et le Japon (Fujiwara *et al.* 2015) ainsi que la méta-analyse de Zchori-Fein *et al.* (2014) qui ajoute à toutes ces données d'autres non publiées. Il y a tout de même une exception à cette parfaite image : en Chine, en 2009, Pan *et al.* (2012) ont montré la présence de *Rickettsia* dans 7% de leur individus Q1, ce qui fait tout de même bien peu au regard des données accumulées et bien loin de nos populations du bassin méditerranéen. Il faut noter que de nombreuses publications ne fournissent pas plus d'indications que « espèce MED », et ne permettent pas d'identifier le cytotype précis de cette espèce malheureusement. À l'aune de ces données on peut donc supposer que Q1 n'est pas infecté par *Rickettsia* et que le transfert est récent. Quand bien même, l'infection par *Rickettsia* serait ancestrale à l'ensemble du super-clade *Africa / Middle East / Asia Minor*, les données de génotypage accumulées sont en faveur d'une perte de *Rickettsia* dans l'haplogroupe Q1, elle aurait été perdue dans cette lignée puisque qu'elle est quasiment absente des individus Q1 échantillonnés de par le monde. Sous cette hypothèse de perte ancestrale de *Rickettsia* dans le Q1, notre hypothèse de transfert horizontal est toujours valide, il a lieu dans le groupe Q1 qui aurait ancestralement perdu son infection.

Le deuxième élément en faveur d'un transfert récent est la faible divergence parmi les *Rickettsia* que nous avons séquencées. Avec un total de 4 500 paires de bases séquencées, réparties sur 9 gènes, et 5 régions intergéniques de la bactérie *Rickettsia*, nous avons pu discriminer 3 souches : l'une propre à l'espèce potentielle SSA 2, l'une propre au cytotype Q3 de l'espèce potentielle MED et la dernière présente dans les cytotypes Q1 et Q2 de l'espèce potentielle MED ainsi que dans l'espèce potentielle MEAM 1. Ce patron peut s'interpréter de deux façons différentes au moins : 1) Il y a peu de variabilité entre les souches, nous ne l'avons donc pas détectée, peut-être parce que nos gènes sont peu variables *a posteriori* et qu'ils ne représentent au final que 4 500 bases sur les 1 500 000 du génome. Ceci nous amène à l'interprétation fautive qu'une seule souche de *Rickettsia* infecte Q1, Q2 et MEAM 1, 2) Ce patron est biologiquement vrai, les souches de *Rickettsia* sont extrêmement peu variables. Des travaux précédents avaient montré

1. Terraz G., Gueguen G., Arnó J., Fleury F., Mouton L. (2014); *Nuclear and cytoplasmic differentiation among Mediterranean populations of Bemisia tabaci : testing the biological relevance of cytotypes*. **Pest management science**, **70**, 1503–1513.

2. Gnankiné O., Mouton L., Henri H., Terraz G., Houndeté T., Martin T., Vavre F., Fleury F. (2013). *Distribution of Bemisia tabaci biotypes (Homoptera : Aleyrodidae) and their associated symbiotic bacteria on host plants in western Africa*. **Insect Conservation and Diversity**, **6**(3), 411–421.



que des souches de *Rickettsia* provenant de Q2 (espèce potentielle MED) et de l'espèce potentielle MEAM 1 étaient identiques (Gueguen *et al.* 2009). Par contre ce travail avait été mené sur deux gènes seulement (*Sca1* et *gltA*).

Il est difficile d'exclure la possibilité que ces souches soient en réalité différentes (autant il est facile de prouver une divergence en détectant des différences, autant prouver l'absence de différence s'avère plus difficile...), en tout cas nos données ne nous permettent pas de les discriminer. On pourrait utiliser des outils plus puissants pour les typer. Dans un premier temps on pourrait augmenter le nombre de gènes, ce qui peut-être long et fastidieux. Une autre possibilité est l'utilisation du *transposon display* qui utilise la variabilité d'insertion d'éléments transposables, utilisé avec succès pour discriminer des souches de *Wolbachia* notamment (Duron *et al.* 2005). Les nouvelles technologies de séquençage pourraient permettre de séquencer les génomes en entier pour résoudre de telles phylogénies où le signal est faible (Colston *et al.* 2014); surtout que le génome de la *Rickettsia* de l'espèce potentielle MEAM 1 est disponible (Rao *et al.* 2012).

Sous l'hypothèse que l'absence de différenciation observée entre les *Rickettsia* de Q1, Q2 et MEAM 1 reflète la réalité et n'est pas un artefact dû à notre outil (MLST), on pourrait y voir un état ancestral d'infection. En effet, la divergence entre les espèces potentielles MED et MEAM 1 semble être très faible d'après leur bactérie primaire (500 000 ans d'après Santos-Garcia *et al.* 2015), ce qui serait donc compatible avec une faible divergence des *Rickettsia* correspondantes. Mais par contre il est difficile d'accepter une seule mutation entre la *Rickettsia* de SSA 2 et celle de MED et MEAM 1 (voir figure 3.3b page 115) car SSA 2 a divergé beaucoup plus tôt dans l'histoire du clade, bien que le placement phylogénétique de cette espèce ne soit pas encore parfaitement clair (Hsieh *et al.* 2014).

### Fréquences des transferts

Après avoir fait l'hypothèse que la *Rickettsia* de Q2 avait infectée Q1, nous avons proposé que la plante soit le vecteur car cela a déjà été montré dans la littérature (Caspi-Fluger *et al.* 2012) et semblait très probable du fait des populations mixtes composées de Q1 et Q2. Le fait que *Rickettsia* puisse se trouver dans les glandes salivaires et le stylet (Brumin *et al.* 2012) explique sa transmission de l'insecte vers la plante et réciproquement, un fait décrit par ailleurs dans la littérature (Purcell *et al.* 1994 ; Mitsuhashi *et al.* 2002) et peut-être Sacchi *et al.* (2008) car les auteurs restent prudents.

Les expériences menées au laboratoire montrent que *Rickettsia* transfère d'individus infectés vers des individus non-infectés avec une relative facilité : la génération (G0) a été infectée avec un taux de 22 %. Cela paraît plutôt élevé par rapport aux situations naturelles que nous avons échantillonnées où seulement deux populations présentaient cette infection parmi une vingtaine où Q1 et Q2 sont dans des populations mixtes (et encore, ces deux populations provenaient probablement d'un même lot de plantes). On peut envisager que la plante hôte joue un rôle dans ce

transfert, mais aussi que celui-ci soit facilité par une niche disponible, c'est-à-dire l'absence de *Cardinium* préalablement au transfert. Dans notre essai, la lignée Q1 non infectée par *Rickettsia* est infectée par *Hamiltonella* et *Cardinium*, donc *a priori* nous ne sommes pas dans le cas d'une niche vide (absence de *Cardinium*).

Il semblerait que la détection simultanée de *Rickettsia* et *Cardinium* soit plus rare qu'attendue. En populations naturelles nous ne l'avons pas trouvé mais la fréquence de *Cardinium* était faible dans ces populations, il faut donc rester prudent. Dans notre expérience, nous avons moins de double infection par ces deux bactéries qu'attendu.

Le transfert en population naturelle aurait pu s'effectuer vers des individus non infectés auparavant par *Cardinium*, puisque de tels Q1 existent dans certaines de nos populations (En France : Chambéry, Moulès, Saint-Rémy de Provence ou en Grèce par exemple, où *Cardinium* est quasiment absente). Nous pourrions en déduire que le transfert s'est fait vers une niche libre (si on omet la présence de *Portiera* qui est fixée car obligatoire, et de *Hamiltonella* qui est quasiment fixée elle aussi).

Bien sûr cette affirmation est à interpréter avec prudence puisqu'on ne peut vérifier le statut d'infection des individus avant transfert. En effet, pour tester la présence d'une bactérie, l'individu doit être tué, on a donc une prévalence de la bactérie pour la population et non un statut d'infection pour chaque individu de l'essai. Dans la lignée de laboratoire utilisée, établie depuis de nombreuses années, cette fréquence est proche de 100 %, proche car on trouve régulièrement un individu non infecté. Ainsi lors du transfert, on ne peut exclure que certains individus ne soient pas infectés et donc plus sensible à un transfert. Néanmoins ce taux de non infection apparaît trop faible au regard de la fréquence du transfert mesuré en *G0* (13 %).

Malgré ce fort taux d'infection en *G0*, on voit que la nouvelle infection se perd assez rapidement dans les générations suivantes. En *G1*, deux des treize lignées sont perdues parce qu'il n'y a pas eu de descendants, et une autre car *Rickettsia* était absente. À la génération suivante, les chiffres sont de respectivement 5 lignées perdues par absence de descendants et 3 parce que *Rickettsia* n'est plus détectée. Si *Cardinium* est encore présente, cela suppose qu'elle coexiste avec *Rickettsia* pendant une période transitoire et qu'une compétition au sein de l'hôte a lieu. Le fait que certaines lignées soient perdues du fait d'une absence de fécondité porte à croire que l'association résultante n'a pas forcément une valeur sélective élevée. On peut donc supposer que l'association résultante du nouveau transfert n'est pas « stable », ce que nous allons discuter maintenant.

### **Stabilité de l'infection au niveau intra-individuel**

Trois facteurs pourraient influencer le devenir de la nouvelle association. 1) La compétition à l'intérieur de l'hôte entre les bactéries 2) La capacité de *Rickettsia* à se transmettre verticalement à la génération suivante 3) La réaction de l'hôte, notamment immunitaire.

### Compétition intra-hôte entre bactéries

Nous avons émis l'hypothèse que *Rickettsia* pouvait être transférée vers un individu précédemment non infecté par *Cardinium*, une niche libre en quelque sorte (en quelque sorte car il y a d'autres bactéries : *Hamiltonella* et *Portiera*), mais nos lignées sont infectées par *Cardinium* ! Au-delà de la simple présence ou absence préalable de *Cardinium*, on peut supposer que la densité de *Cardinium* est variable d'un individu à l'autre, et imaginer que les densités les plus faibles permettent le transfert de *Rickettsia*. Celle-ci est transférée dans l'individu infecté par *Cardinium*, à ce moment là s'établit une compétition entre les deux car elles coexistent dans un même individu. Cette compétition va être plus ou moins forte suivant le niveau d'interaction entre les bactéries, niveau d'interaction qui dépend de la localisation. En l'occurrence les deux bactéries peuvent être présente dans le bactériomes, nécessaires à leur transmission verticale. Par contre, pour cette nouvelle infection la localisation peut-être différente, nous n'avons pas les données pour cette nouvelle association Q1RH. Au contraire, la compétition sera diminuée si l'une est dans les bactériocytes et l'autre dans d'autres tissus. La localisation tissulaire joue un rôle important dans la transmission, puisque tous les symbiotes secondaires doivent coloniser les bactériocytes pour être transmis à la génération suivante (Brumin *et al.* 2012). Donc si *Cardinium* est déplacée des bactériocytes par compétition, elle ne peut se transmettre à la génération suivante. *Rickettsia* est différente car il semblerait qu'elle colonise les ovocytes très tôt dans le développement puis qu'elle en soit exclue petit à petit pour finalement se concentrer dans les cellules folliculaires et ainsi être transmise à la génération suivante (Brumin *et al.* 2012).

Si la densité de *Cardinium* est faible au départ de la compétition (introduction de *Rickettsia*), *Rickettsia* pourrait avoir un avantage en terme de densité et donc plus de chances de « gagner ».

Si effectivement *Rickettsia* est plus « virulente » que *Cardinium* dans le fond nucléaire et mitochondrial de nos lignées, cela pourrait être du au fait qu'elle transfère souvent (Sakurai *et al.* 2005 ; Oliver *et al.* 2006 ; Goto *et al.* 2006). Une plus grande description des souches de *Rickettsia* à l'intérieur du clade *B. tabaci* pourrait nous donner des informations sur l'historique des transferts horizontaux à l'intérieur du clade.

Pour le moment, nous avons considéré seulement *Cardinium* et *Rickettsia*, mais il ne faut pas oublier que l'association est beaucoup plus complexe avec la présence d'*Hamiltonella*, et de *Portiera*, le symbiote primaire. La question se pose donc de l'arrivée de *Rickettsia* dans cette communauté car elle peut aussi se trouver en compétition avec *Hamiltonella*. Dans nos données nous n'avons pas mis en évidence d'une diminution de l'infection par *Hamiltonella* lorsque *Rickettsia* « arrive ». Mais il paraît peu probable qu'*Hamiltonella* puisse être éliminée car elle serait indispensable à son hôte, avec *Portiera*, pour la synthèse de certains acides aminés essentiels, amenant à une triple dépendance entre ces deux bactéries et l'hôte (Rao *et al.* 2015).

### La capacité de *Rickettsia* à se transmettre à la génération suivante

En regardant attentivement les résultats de transferts expérimentaux, on voit une forte décroissance de l'infection dès la deuxième génération (voir tableau 3.2), l'association semblant stable par la suite sur de nombreuses générations. La question se pose de sa capacité à se transmettre verticalement à la génération suivante. *Rickettsia* se transmettrait verticalement en envahissant les cellules folliculaires des ovocytes et non grâce aux bactériocytes (Brumin *et al.* 2012), contrairement à ce qui est classiquement observé chez les aleurodes (Costa *et al.* 1996 ; Szklarzewicz & Moskal 2001) ou même à ce qui était pensé auparavant chez *Rickettsia* (Gottlieb *et al.* 2006, 2008). Ces résultats contradictoires montrent qu'il faut rester prudent quant au mode de transmission verticale de *Rickettsia*. On peut supposer alors qu'elle n'arrive pas à infecter les ovocytes afin de se transmettre à la génération suivante, ce qui est peut-être plus difficile que d'être transmis par les bactériocytes comme *Cardinium*. En plus, la transmission verticale de mère à descendants provoque un fort goulot d'étranglement, car seulement quelques unes pourront pénétrer dans les œufs. Les nombres transmis étant très faible de l'ordre de la dizaine pour des mitochondries chez l'homme, à quelques milliers de *Buchnera* chez *Acyrtosiphon pisum* (Mira & Moran 2002). A cause de cette réduction du nombre de bactéries, de façon aléatoire on peut imaginer que *Rickettsia* ne se transmet pas si elle n'a pas atteint une densité suffisante, mais nos données de PCR quantitative (non montrées ici) ne semble pas montrer une réduction de la densité de *Rickettsia*.

En situation naturelle, cette nouvelle association est toujours présente en 2014, dans la même serre que celle où elle avait été trouvée deux ans auparavant et elle fixée dans la population : il n'y a pas d'autres cytotypes, ce qui semble confirmer qu'une telle infection peut exister et n'est pas qu'un état transitoire. Le fait que cette association soit toujours présente plusieurs années après nous indique qu'elle est stable est que *Rickettsia* se transmet bien verticalement aux générations suivantes.

### Interactions avec l'hôte

En plus de la compétition entre bactéries, il ne faut pas oublier le fait que celles-ci sont hébergées par l'hôte, qui joue un rôle dans ces interactions. Celui-ci a une réaction immunitaire pour se protéger de l'infection, mais si cette réaction est trop coûteuse, il a intérêt à tolérer la bactérie plutôt que de lutter sinon sa valeur sélective serait gravement diminuée (Alizon *et al.* 2009). Ainsi un équilibre se crée entre la défense de l'hôte et la densité du symbiote, clé de sa transmission. Dans le cas de notre essai en laboratoire, les insectes ne sont pas soumis à une compétition puisque ce sont des lignées isofemelles, la réaction de l'hôte envers les symbiotes n'est donc pas contre-sélectionnée au niveau inter-individuel, ce qui n'est pas le cas pour des populations naturelles. Nous discuterons dans la partie suivante de la compétition entre les hôtes.

Si l'hôte et les bactéries sont co-adaptés du fait d'une longue histoire évolutive commune on peut faire l'hypothèse que l'hôte tolère mieux la bactérie et ne déclenche pas une réaction immu-

nitaine forte. Ceci est clair pour une relation entre un hôte et son symbiote primaire par exemple. On peut imaginer que le clade *B. tabaci* soit ancestralement infecté par *Rickettsia* ou qu'en tout cas la présence de cette bactérie soit courante dans les différentes espèces. Ainsi on pourrait émettre l'hypothèse que *B. tabaci* tolère *Rickettsia* car tous deux sont coadaptés. Par contre, le fait que *Rickettsia* soit perdue secondairement dans certaines espèces implique le relâchement de pression sur ces coadaptations au niveau du génome de l'hôte, qui peuvent donc être perdues, plus ou moins rapidement selon les nouvelles pressions de sélection.

Pour alimenter cette hypothèse il faudrait mieux connaître l'histoire commune entre *Rickettsia* et *B. tabaci*, c'est à dire identifier les événements d'infection et de perte de *Rickettsia* dans le clade, ou en tout cas mieux connaître la prévalence de *Rickettsia* dans chacune des espèces potentielles. Malheureusement pour le moment, ce sont surtout les espèces potentielles MED et MEAM1 qui sont fortement étudiées (car les plus largement distribuées) en ce qui concerne les informations sur la prévalence, au détriment des 30 et quelques autres espèces. Pour les autres, nous avons juste la présence ou non de *Rickettsia*, rarement la fréquence (voir figure 3.2 page 113).

Ainsi, on peut imaginer que des entités biologiques (cytotypes, espèces,...) à l'intérieur du clade *B. tabaci* aient des allèles ou des gènes co-adaptés à *Rickettsia* du fait de leur histoire commune. Si une *Rickettsia* arrive dans un fond nucléaire qui n'a pas connu une telle bactérie, l'hôte sera moins réceptif, dans le cas contraire, il le sera plus. A partir de ce constat on peut imaginer un scénario : Q1 n'aurait pas « connu » *Rickettsia* dans l'aire dite de référence en Grèce et réagirait fortement envers *Rickettsia*, mais comme Q1 et Q2 sont en situation de sympatrie en France, ils ont échangé des allèles et peut-être permis l'introggression depuis Q2 des allèles nécessaires à une plus grande tolérance de *Rickettsia* dans Q1. Pour tester de telles hypothèses il faudrait des outils de diagnostic nucléaire très puissant pour scanner les génomes. Dans le chapitre précédent nous avons vu la limite des marqueurs microsatellites pour scanner efficacement le génome. C'est donc dans notre dernier chapitre que nous verrons la mise au point de marqueur RADSeq<sup>1</sup>, qui permettrons peut-être de répondre à de telles questions.

### Compétitivité du cytotype

Dans le paragraphe précédent nous avons étudié la possibilité de l'émergence d'un nouveau cytotype. Une fois que les différents freins au transfert et à la transmission verticale des bactéries ont été franchis, en faisant l'hypothèse que cette association est stable au niveau intra-hôte on peut se demander si cette nouvelle entité, que l'on pourrait donc appeler cytotype, va persister en populations naturelles.

Dans la partie précédente nous avons vu que l'hôte était soumis à un dilemme lorsqu'il était infecté par la nouvelle bactérie : la tolérer et donc qu'elle prélève des ressources à ses dépens ou bien lutter mais à un certain prix. Ce prix va se traduire en terme de valeur adaptative,

---

1. Restriction Associated DNA sequencing



l'individu va-t-il être contre-sélectionné dans la population où il est présent ? Il est assez difficile de prédire l'avenir, et donc le devenir de ce cytotype. Néanmoins on sait que deux ans plus tard il est toujours présent dans la serre où il a été découvert, et ce en absence d'autres cytotypes ce qui pourrait nous inciter à penser qu'il a été plus compétitif que deux autres cytotypes : Q2RAW et Q1CH qu'on trouvait auparavant dans la même serre (voir figure 3.1 page 111).

Des expériences menées au laboratoire nous ont permis de mesurer différents traits d'histoire de vie de Q1RH face à Q1CH, après traitement insecticide ou choc thermique. Nous avons choisi ces deux facteurs car on sait que la température influence directement le développement des insectes du fait qu'ils sont ectothermes. Bien que des températures élevées accélèrent leur développement, un stress thermique peut se traduire par une diminution de la longévité ou de la fertilité. Dans nos régions, *B. tabaci* ne survit qu'en serre l'hiver, sauf peut-être dans la partie sud de la France où il pourrait se maintenir en dehors des serres. Des épisodes caniculaires pourraient donc avoir un effet très fort sur des populations en serre et dans une moindre mesure en extérieur, tout comme l'augmentation continue des températures. L'aspect insecticide a été traité car *B. tabaci* est soumis à d'intenses traitements insecticides.

Ce qui nous intéresse tout particulièrement est le fait que les bactéries symbiotiques peuvent être impliquées dans la thermorégulation ou la résistance aux insecticides, comme nous l'avons évoqué en introduction. Ceci a été démontré chez le puceron du pois *Acyrtosiphon pisum* où *Serratia symbiotica* est impliquée dans la tolérance aux fortes températures (Montllor *et al.* 2002). Et plus proche de nous, la présence de *Rickettsia* augmente la tolérance de *B. tabaci* MEAM1 au choc thermique (Mahadav *et al.* 2009 ; Su *et al.* 2014a ; Brumin *et al.* 2011). De même, *Rickettsia* jouerait un rôle dans la tolérance aux insecticides tels que les néonicotinoïdes, les pyridines ou encore les acides tétroniques (Ghanim & Kontsedalov 2009). Bien sûr ces résultats sont à prendre avec précaution parce que les effets phénotypiques des bactéries sont très contexte-dépendant, ce dernier pouvant être environnemental ou génétique (Cass *et al.* 2015a). Nous avons donc mis en place un protocole expérimental pour tester ces effets (voir l'annexe C page 225). Nos résultats suggèrent que la température n'influence pas la compétition inter-hôte dans le cas d'une population mixte Q1RH et Q1CH. Par contre les individus infectés par *Rickettsia* et *Hamiltonella* sont plus sensibles au traitement insecticide (Decis<sup>®</sup>) que les individus infectés par *Cardinium* et *Hamiltonella*. Même si on ne peut démêler l'effet de *Cardinium* et de *Rickettsia* (est ce que l'un augmente la susceptibilité, où bien l'autre la diminue ?), on peut dire que l'association prise comme un tout est plus ou moins sensible, en l'occurrence le cytotype Q1RH paraît plus sensible.

Bien sûr, il est dangereux de tirer des conclusions hâtives, mais la serre dans laquelle ce nouveau cytotype a été trouvé est gérée en protection biologique intégrée, on pourrait alors supposer qu'en absence d'insecticide le cytotype Q1RH est plus performant.





# Chapitre 4

## Développement et utilisation des marqueurs *RADSeq* pour le modèle *B. tabaci*

1	Introduction	129
1.1	Objectifs	129
1.2	Un outil pour la génomique des populations	129
1.3	Concrètement, qu'est ce que le <i>RADSeq</i> ?	130
2	Expérience pilote, multiplexer de nombreux individus à faible coût	131
2.1	Méthode	133
2.1.1	Plan expérimental	133
2.1.2	Analyse des données	134
2.2	Résultats du séquençage	135
2.2.1	Nombre de lectures par modèles biologiques	135
2.2.2	Estimation du nombre de locus chez <i>B. tabaci</i>	136
2.3	Conclusion sur cette expérience pilote	139
3	Développement de l'outil pour <i>B. tabaci</i>	140
3.1	Objectifs de l'essai	140
3.2	Matériel et méthodes	140
3.2.1	Populations testées	140
3.2.2	Extraction d'ADN	141
3.2.3	Analyse des données	141
3.3	Résultats	145
3.3.1	Compte-rendu du séquençage	145
3.3.2	Démultiplexage et filtrage des lectures	146
3.3.3	<i>Clustering</i> : combien de locus ?	147

## Chapitre 4. Développement et utilisation des marqueurs *RADSeq* pour le modèle *B. tabaci*

3.3.4	L'hétérozygotie est fonction du nombre de lectures	148
3.3.5	<i>Clustering</i> inter-individus, combien de locus partagés ?	149
3.3.6	Détection des symbiotes	150
3.3.7	Structure génétique	153
3.3.8	Déséquilibre nucléo-cytoplasmique	157
4	Perspectives	160

## Introduction

### Objectifs

Dans le chapitre 2 traitant du compartiment nucléaire des cytotypes nous avons mis en évidence le besoin de développer un outil plus puissant pour faire de la génomique des populations. Jusqu'à présent, nous avons utilisé des marqueurs microsatellites (comme l'ensemble de la communauté scientifique s'intéressant à *B. tabaci*), mais ils sont limités par le nombre de locus génotypés (sept dans notre cas). De plus, dans le cas de *B. tabaci*, on voudrait être capable de détecter des régions du génome sous sélection et essayer de déterminer s'ils existent des interactions entre le compartiment nucléaire et les symbiotes, c'est-à-dire de la coadaptation. Pour cela il faut développer un outil capable de scanner l'ensemble du génome, à des milliers de locus, nous avons choisi le *RADSeq*. L'utilisation de ce nouvel outil est plutôt récente et nécessite de très nombreux ajustements, autant dans les protocoles de biologie moléculaire que les analyses bio-informatique.

Les objectifs étaient donc 1) Être capable au sein du laboratoire de préparer les banques nécessaires au séquençage. 2) *B. tabaci* étant un très petit insecte, montrer que la faible quantité d'ADN permet un génotypage individuel. 3) Estimer le nombre de locus du génome de *B. tabaci* 4) Réaliser différentes analyses de génomique des populations. 5) Détecter les symbiotes secondaires dans les données génomiques.

Pour cela, deux essais ont été mis en place, un premier qui contenait divers modèles biologiques des projets en cours de développement au sein du laboratoire et un deuxième, propre à *B. tabaci*. Nous allons tout d'abord brièvement introduire la génomique des populations puis définir ce qu'est le *RADSeq*, ensuite nous détaillerons les résultats obtenus tout au long de ce chapitre.

### Un outil pour la génomique des populations

L'étude de l'information génétique la plus complète possible pour de très nombreux individus est rendue possible par l'émergence du séquençage haut débit, en parallèle de l'augmentation des capacités de calcul, qui ont permis de grandes avancées dans le nouveau domaine qu'est la génomique des populations. Il y a encore quelques années, l'étude de quelques dizaines de locus microsatellites était longue et coûteuse, alors que la technologie permet aujourd'hui de génotyper des dizaines de milliers de SNP dans le cadre des outils NGS (*Next Generation Sequencing*) (Davey *et al.* 2011 ; McCormack *et al.* 2013). La génomique des populations est possible grâce au séquençage de milliers de régions homologues du génome de plusieurs individus, à une échelle populationnelle ou même interspécifique. La génomique des populations permet de résoudre des problèmes classiques en écologie, en génétique des populations et plus largement dans le domaine de la biologie évolutive tels que l'analyse de parenté et filiation, l'analyse des flux génétiques et des migrations, la structure des populations, la phylogéographie, la phylogénie, et tout cela avec une puissance sans précédent (Hohenlohe *et al.* 2010 ;

Stapley *et al.* 2010). De plus, l'analyse fine du génome par cette technique permet l'identification d'*outlier* pour n'importe quelle statistique, et ainsi d'identifier des régions impliquées dans l'adaptation locale ou même la spéciation (Hohenlohe *et al.* 2010).

Ces techniques ont en commun ce qu'on appelle une approche de génotypage par représentation réduite (*Reduced Representation, RR* en anglais) ou aussi GBS (*Genotype By Sequencing*), qui consiste à échantillonner l'ensemble du génome à des locus homologues pour identifier des variants, sans connaissance préalable du génome. Ainsi la génomique des populations est accessible à tous les modèles biologiques. Dans la littérature, l'expression *Genotype By Sequencing* est souvent associée à l'utilisation des marqueurs *RADSeq* (*Restriction site-Associated DNA sequencing* Davey *et al.* 2011), qui sont les techniques utilisant des enzymes de restriction pour définir des locus homologues. Plusieurs approches moléculaires ont été développées pour séquencer massivement des positions du génome associées à des sites de restriction enzymatique, et nous allons utiliser le protocole classique de *RADSeq* (Baird *et al.* 2008).

Le *RADSeq* a été utilisé pour des problématiques variées comme nous l'avons vu mais aussi sur de nombreux modèles biologiques différents comme les fleurs du genre *Pedicularis* (*Orobanchaceae*, Eaton & Ree 2013), le ver de terre *Lumbricus rubellus* (Giska *et al.* 2015), le crustacé cirripède *Notochthamalus scabrosus* (Zakas *et al.* 2014), le complexe d'espèces de la grenouille *Craugastor augusti* (Streicher *et al.* 2014), l'épinoche *Gasterosteus aculeatus* (Hohenlohe *et al.* 2010) ou encore le bourdon *Bombus impatiens* (Lozier 2014).

L'enzyme de restriction est choisie en fonction de la fréquence de coupure requise pour l'étude en question : un *genome scan* pour la détection de sélection nécessitera *a priori* beaucoup plus de marqueurs que la génétique des populations (Baird *et al.* 2008). La fréquence de coupure dépend de la longueur du site de restriction, mais aussi du taux de GC respectif de l'enzyme et du génome, ainsi que de la position phylogénétique du modèle biologique (Herrera *et al.* 2015).

### Concrètement, qu'est ce que le *RADSeq* ?

Le *RADSeq* permet le séquençage des régions flanquantes des sites de restriction (voir figure 4.1). Le génome est tout d'abord digéré par une enzyme de restriction. Ensuite, des adaptateurs (P1) sont ligués aux extrémités cohésives créées lors de la coupure précédente. Ces adaptateurs contiennent trois parties, une pour le séquençage Illumina, un identifiant moléculaire ou barcode (terme que nous utiliserons) spécifique de chaque individu et enfin une partie complémentaire des extrémités cohésives créées par l'enzyme de restriction.

Les extraits de tous les individus sont mélangés — le multiplexage — (les séquences seront identifiables grâce au barcode), pour former plus qu'un seul « échantillon » : la banque.

Une étape de fragmentation est ensuite nécessaire pour raccourcir et homogénéiser la longueur de ces fragments qui peuvent être long de centaines de milliers de bases : on peut utiliser une enzyme ou plus souvent la sonication. On ajoute alors un deuxième adaptateur (P2) à ces extrémités franches nouvellement créées. Cet adaptateur est en forme de Y pour éviter l'amplification

## 2. Expérience pilote, multiplexer de nombreux individus à faible coût

de fragments composés de chimères de P2 et qui ne contiendraient pas de P1 (indispensable). Pour finir, une PCR permet d'amplifier les fragments comportant les deux types d'adaptateurs et donc seulement ceux flanquant les sites de restriction. Le séquençage de la banque ainsi créée a lieu, soit à partir des deux extrémités des fragments (*paired-end*), soit seulement depuis l'extrémité correspondant au site de restriction (*single-read*). Par la suite on peut séparer *in silico* les séquences appartenant à chaque individu grâce au barcode qui se situe en début de lecture, c'est le démultiplexage.

### Expérience pilote, multiplexer de nombreux individus à faible coût

Afin de mettre en place au sein du laboratoire le protocole de biologie moléculaire lié au *RAD-Seq*, et pour réaliser à faible coût le multiplexage de nombreux individus, une première expérience pilote a été mise en place, dont les résultats ont fait l'objet d'une publication disponible en annexe A page 213<sup>1</sup>. Pour cela des ajustements au protocole standard de *RADSeq* (Etter *et al.* 2011) ont été réalisés.

En ce qui concerne plus particulièrement le modèle *B. tabaci*, cette première utilisation des *RADSeq* allait nous permettre de répondre à la question de la faisabilité liée à la très faible quantité d'ADN du modèle, et de connaître le nombre de locus de l'enzyme de restriction *SbfI* dans le génome de *B. tabaci*, dont dépend le nombre d'individus multiplexables. L'enzyme *SbfI* a été choisie car son site de restriction est plutôt rare dans le génome (tous les 65 kilobases environ Davey *et al.* 2011) ce qui pour un même nombre de séquence obtenue au séquençage permet d'augmenter le nombre d'individus multiplexables, c'est une question de compromis.

En effet connaître le nombre de locus dans le génome permet d'optimiser le nombre d'individus à séquencer simultanément car ces deux variables sont intimement liées, à un facteur près : la profondeur. La profondeur est le nombre moyen de lectures obtenues au séquençage pour un locus. On a donc la relation suivante :

$$N_{\text{Individus}} = \frac{N_{\text{lectures}}}{N_{\text{Locus}} * \text{Profondeur moyenne}}$$

Dans le projet, il était prévu de génotyper de nombreux individus pour les approches de génétique des populations, il fallait donc séquencer le maximum d'individus sur une même ligne de séquençage pour optimiser les coûts.

Nous avons vu que pour assigner les séquences à des individus, elles contiennent un barcode individuel, compris dans l'adaptateur P1. En tout, cet adaptateur est long d'environ 70 bases et coûte environ 100 euros, à multiplier par le nombre d'individus à multiplexer dans la banque. Le séquençage simultané de 200 individus (nombre de barcodes retenus pour l'essai pilote) revient

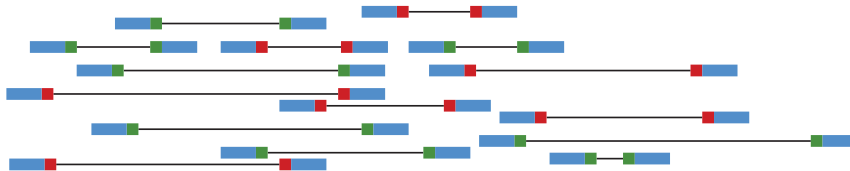
---

1. Henri H., Cariou M., Terraz G., Martinez S., El Filali A., Veyssiere M., Duret L., Charlat S. (2015). Optimization of multiplexed *RADSeq* libraries using low-cost adaptors. *Genetica*, 143(2), 139–143.

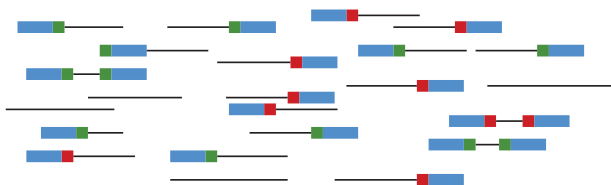
**1. Digestion et ligation des adaptateurs P1**



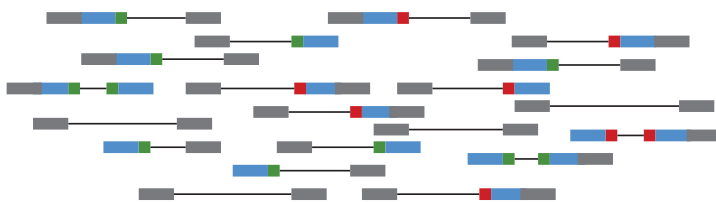
**2. Multiplexage**



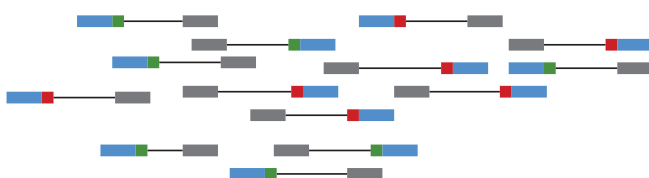
**3. Fragmentation**



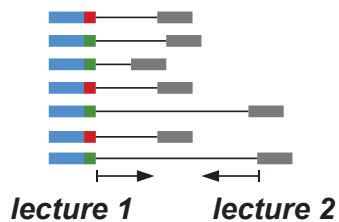
**4. Ligation des adaptateurs P2**



**5. Amplification entre P1 et P2**



**6. Séquençage**



**FIGURE 4.1** – Protocole d’obtention des marqueurs *RADSeq*. Les adaptateurs P1 sont composés de la partie commune, en bleu, du barcode propre à chaque individu (ici deux individus : rouge ou vert), ainsi qu’une partie complémentaire des extrémités cohésives créées par l’enzyme de restriction. A l’étape 4, tous les bouts francs sont ligués au P2 (en gris). L’étape 5 permet l’amplification sélective des fragments qui contiennent les deux adaptateurs. Le séquençage *single-read* ne comprend que la lecture 1, alors que le séquençage dit *paired-end* comprend les deux.

## 2. Expérience pilote, multiplexer de nombreux individus à faible coût

drait ainsi à 20 000 euros ! On pourrait bien sûr préparer des banques avec moins d'individus et les séquencer indépendamment, mais par contre on augmenterait d'autant le coût de préparation de ces banques et du séquençage (on gagne alors en profondeur de séquençage, mais ce n'est pas forcément nécessaire).

Nous avons donc synthétisé séparément la partie commune et la partie individuelle et les avons assemblées.

### Méthode

Pour limiter les coûts, l'approche qui a été utilisée au laboratoire est de synthétiser séparément la partie longue et commune à tous les adaptateurs et la partie individuelle (le barcode) qui est très courte. Ces deux morceaux sont ensuite ligés pour obtenir l'adaptateur complet. Le coût de la préparation de 200 adaptateurs tombe à moins de 6 000 euros, soit d'un ordre de grandeur par rapport au protocole classique ! Un problème qui a été envisagé avec ce protocole et la formation parasite de dimères P1-P1 lorsque les concentrations respectives d'extrait génomique et de cet adaptateur ne sont pas optimisées. L'astuce a été d'utiliser des adaptateurs dépourvus de phosphate en 5', ce qui les empêche de se lier entre eux. Nous n'allons pas détailler ici tout le protocole de préparation de la banque, détails qui sont disponibles dans l'article en annexe A page 213, mais plutôt nous concentrer sur quelques résultats obtenus, et détailler un peu les informations propres à *B. tabaci*.

### Plan expérimental

Des échantillons de plusieurs espèces d'arthropodes ont été sélectionnés, pour un total de 77 individus. Ils correspondent aux modèles biologiques des différents projets envisageant le *RAD-Seq* au laboratoire. Ces échantillons sont très hétérogènes, autant en terme biologique (taille et composition en bases du génome), que du point de vue de l'ADN : les méthodes d'extraction ainsi que la qualité des extraits disponibles ne sont pas forcément les mêmes. En ce qui concerne *B. tabaci*, nous avons choisi 18 individus provenant de nos lignées de laboratoire Q1 et Q2, celles que nous avons déjà évoquées dans les chapitres précédents. Dans le tableau 4.1 est indiqué le nombre d'individu de chaque espèce représentée. Le plan expérimental nous permettait de tester nos nouveaux adaptateurs, l'influence de la concentration d'ADN, et la reproductibilité de la méthode en assignant l'ADN d'un même échantillon, dilué ou non, à plusieurs barcodes. Des échantillons de *Drosophila melanogaster* ont servi de contrôle positif puisque leur génome est connu, cela permettait donc de connaître la proportion de locus retrouvés dans les données par rapport à l'attendu. Finalement, la banque a été séquencée avec la technologie Illumina (HiSeq 2000) en *single read* sur 50 bases. Le protocole détaillé est disponible dans le papier décrivant cet essai page A.



Espèce	Nb	C
<i>Leptopilina boulardi</i>	15	2.53
<i>Proasellus cavaticus</i>	15	4.94
<i>Aedes albopictus</i>	11	4.93
<i>Curculio elephas</i>	2	6.40
<i>Curculio glandium</i>	3	4.10
<i>Culicidae sp.*</i>	5	0.94
<i>Paratrechina sp.*</i>	5	0.24
<i>Cnaphalocrocis poeyalis*</i>	5	0.86
<i>Drosophila sukuzii</i>	5	5.04
<i>Bemisia tabaci</i>	18	0.54
<i>Drosophila melanogaster</i>	16	4.625

**TABLE 4.1** – Individus utilisés lors de l’expérience pilote. Nb : nombre d’individus ; C : concentration d’ADN moyenne en nanogramme. On voit ici que les échantillons *B. tabaci* sont les plus pauvres en ADN. \* : ces différentes espèces font partie d’un projet « Symbiocode », nom sous lesquelles elles seront désignées.

### Analyse des données

L’analyse des données comprend plusieurs étapes, elles ont toutes été effectués avec STACKS v1.16 (Catchen *et al.* 2013) qui a l’avantage d’être facile et rapide à utiliser.

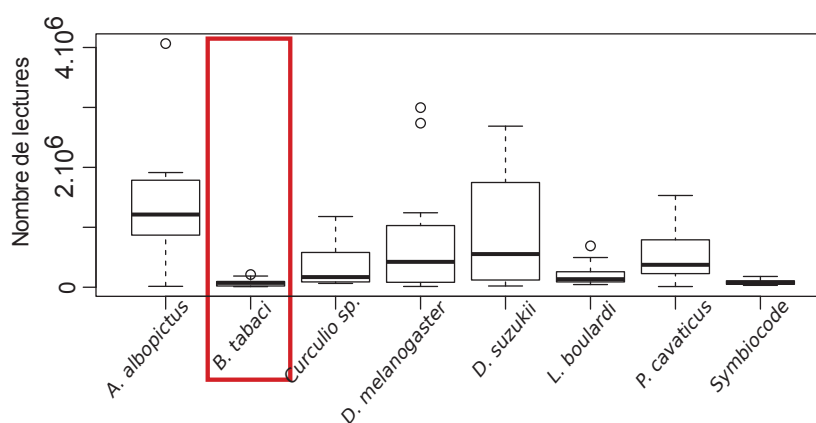
1. **Demultiplexage** : Tri des séquences en fonction de leur barcode et vérification de la présence du site de restriction. Lors de cette étape les séquences sont assignées à chaque individu en fonction des premières bases de la séquence. La vérification du site de restriction permet d’éliminer des fragments d’ADN qui se seraient liés à des adaptateurs sans être de véritables locus.
2. **Élimination des adaptateurs** : Le protocole de *RADSeq* crée malheureusement des chimères d’adaptateurs : ce sont des séquences qui contiennent un barcode, mais qui sont composées de polymères d’adaptateurs et ne sont donc pas exploitable. Elles ont été retirées du jeu de données par recherche de similarité avec BLAST.
3. **Clustering intra-individuel** : Cette étape permet de regrouper les séquences d’un même locus pour chaque individu. Deux paramètres sont très importants ici. 1) Le nombre minimal  $m$  de lectures pour créer un locus, que nous avons fixé à trois. 2) Le nombre de mésappariements  $M$  autorisés entre deux lectures pour les rassembler dans un même locus, fixé ici à deux. Il est important de noter que STACKS ne gère pas les insertions et délétions.
4. **Clustering inter-individus** : Cette étape est similaire à la précédente mais elle permet de regrouper les locus homologues entre les individus, et non plus les lectures d’un même locus au sein d’un individu. Ici, le paramètre important est le nombre  $n$  de mésappariement autorisé (en plus de ceux intra-individuels) entre deux locus de deux individus pour

les rassembler dans un même locus, nous l'avons fixé à trois.

## Résultats du séquençage

### Nombre de lectures par modèles biologiques

Le nombre total de lectures obtenues lors du séquençage est de 76 538 697, qui ont alors été démultiplexées. Parmi ces lectures, 1 440 083 sont attribuées à *B. tabaci*. La variabilité est très grande d'une espèce à l'autre (voir figure 4.2). Par exemple la médiane du nombre de lectures par individu de l'espèce *A. albopictus* est supérieure à un million alors que le maximum pour *B. tabaci* est de 211 977. Pour une même espèce la variabilité est aussi très grande : dans le cas de *B. tabaci*, le nombre de lectures varie de 10 155 à 211 977, soit de deux ordres de grandeur, avec une moyenne de 75 793.

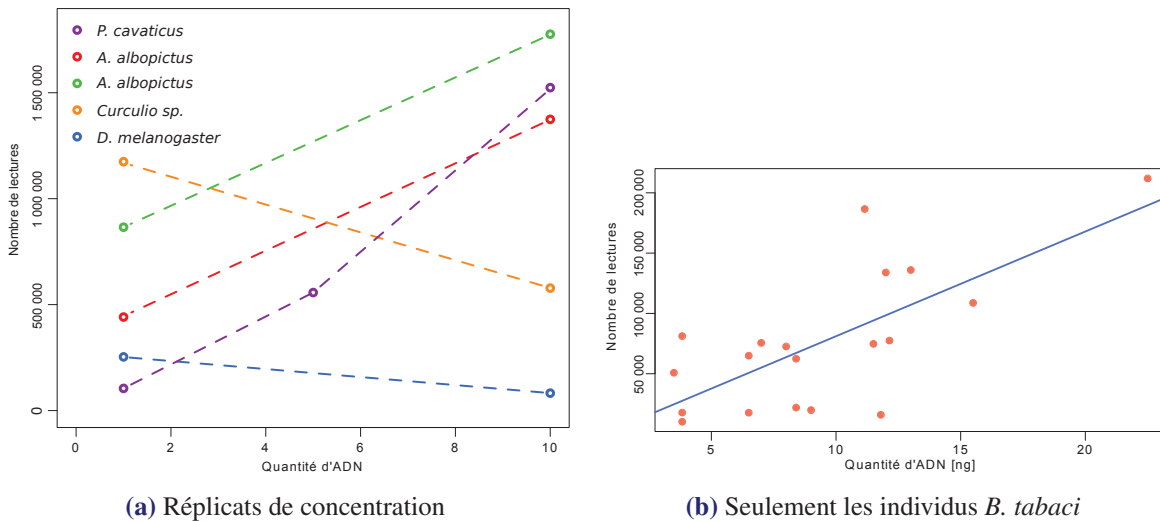


**FIGURE 4.2** – Nombre de lectures obtenues pour chaque espèce. Le modèle d'étude *B. tabaci* est encadré en rouge. Le genre *Curculio* regroupe deux espèces (voir tableau 4.1), *Symbiocode* regroupe 11 espèces différentes (voir tableau 4.1).

Comme, d'une part, la quantité d'ADN est très limitée chez *B. tabaci* (entre 3,5 et 22,5 nanogramme par individu, pour une moyenne de 9,3) et, d'autre part, on s'est demandé s'il est nécessaire de normaliser cette quantité d'un individu à l'autre pour la préparation des banques, nous avons regardé la relation entre concentration en ADN et nombre de lectures. De plus, certains échantillons avaient été dilués et affectés à différents identifiants moléculaires selon la concentration. Aucun des échantillons de *B. tabaci* n'a été dilué car la quantité d'ADN initiale était déjà très faible. Les résultats sont montrés sur la figure 4.3a. On voit que la relation entre quantité d'ADN et nombre de lectures n'est pas forcément dans le sens attendu puisque pour deux modèles (*Curculio sp.* et *D. melanogaster*), le nombre de lectures est inversement proportionnel à la quantité d'ADN. Pour les autres, la quantité d'ADN a un effet positif mais pas dans les proportions auxquelles on pourrait s'attendre (Pour *A. albopictus*, en multipliant la quantité d'ADN par 10, le nombre de lectures est seulement triplé). L'hypothèse la plus probable est qu'un effet de saturation est observé, probablement causé par une trop faible quantité d'adaptateur.

Au sein de l'espèce *B. tabaci*, la relation est positive ( $p < 0,001$ ) entre la quantité d'ADN et le nombre de lecture (chaque nanogramme d'ADN « apporte » environ 8 500 lectures, voir figure 4.3b). Même s'il est *a priori* important de normaliser les quantités d'ADN entre individus pour homogénéiser le nombre final de lectures, cela paraît impossible pour *B. tabaci* dont les quantités initiales sont déjà très faibles. Comme les échantillons *B. tabaci* sont en présence d'échantillons d'autres espèces dont les quantités initiales en ADN et les propriétés génomiques sont très différentes, les conclusions pourraient être différentes dans le cas d'une banque composée seulement d'individus de l'espèce *B. tabaci* avec une préparation optimisée pour ces conditions (faible quantité d'ADN).

Nous allons voir maintenant si cette faible quantité de lectures disponibles pour *B. tabaci* nous permet tout de même de retrouver un nombre raisonnable de locus. Nous allons donc nous focaliser uniquement sur notre modèle d'étude.



**FIGURE 4.3** – Nombre de lectures en fonction de la quantité d'ADN initiale. (a) : Individus de différentes espèces utilisés à plusieurs dilutions. (b) : Échantillons *B. tabaci*, qui n'ont pas été dilués.

### Estimation du nombre de locus chez *B. tabaci*

**Clustering intra-individuel** Le nombre de locus retrouvé par individu *B. tabaci* est très variable, dans une gamme allant de 430 à 3500 et est fortement corrélé au nombre de lectures des individus, comme l'indique la figure 4.4a. Cette figure nous montre qu'on n'atteint pas un plateau, c'est à dire que tant qu'on « ajoute » des lectures, on retrouve d'autres locus, ce qui indique que le nombre de locus total chez *B. tabaci* n'a pas été retrouvé, sans pour autant nous donner une indication sur la proportion retrouvée.

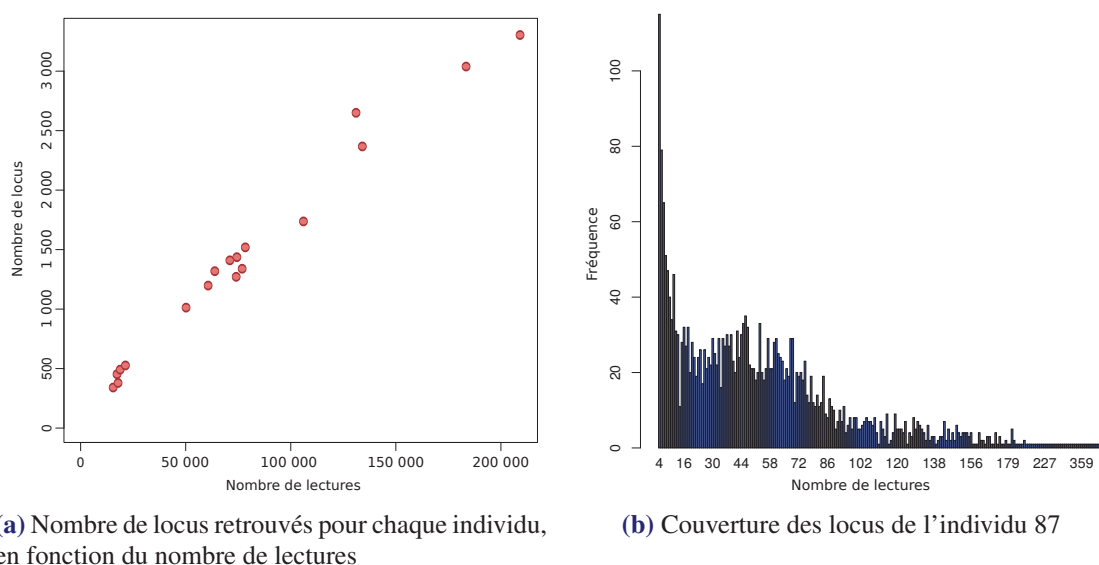
La profondeur moyenne de tous ces locus est de 52, nombre de lectures qui peut paraître suffisant pour définir un locus, même pour un hétérozygote cela fait 26 lectures par allèle ; mais cette valeur est trompeuse. Par exemple pour un individu avec environ 100 000 lectures, pour atteindre une profondeur de 50 il faudrait que le nombre de locus du génome soit d'environ

## 2. Expérience pilote, multiplexer de nombreux individus à faible coût

2 000, valeur incohérente au regard de la taille du génome (estimée à environ 700 Mb, Guo *et al.* 2015). L'explication est simple : la profondeur n'est pas du tout homogène d'un locus à l'autre, c'est ce qui est illustré sur la figure 4.4b. Sur cette figure est représentée la distribution de la profondeur des locus de l'individu 87, choisi pour son nombre de lecture moyen et assez représentatif du cas général. On voit que la profondeur a une queue de distribution très longue avec certains locus qui excède très largement 200x, à l'opposé d'autres locus qui ne dépassent pas quelques lectures.

Cette hétérogénéité a une conséquence très importante : il faut un nombre très élevé de lectures pour séquencer tous les locus (ou au moins une grande proportion), bien plus élevé qu'un simple tirage aléatoire de ceux-ci. En soit même ce n'est pas très important mais cela implique que le nombre d'individus que l'on peut multiplexer est beaucoup moins grand que ce que l'on aurait pu penser, et donc que le coût sera plus *a priori* plus élevé. Une explication possible de cette hétérogénéité dans la profondeur de séquençage est l'amplification aléatoire de certains fragments d'ADN pendant la PCR nécessaire à la création de la banque, cette différence d'amplification s'accroissant à chaque cycle de PCR. Lors des préparations ultérieures, plusieurs PCR en parallèle pourraient résoudre ce problème.

En conclusion de cette partie, nous avons vu que le nombre de locus retrouvé est plutôt faible, en partie à cause de l'hétérogénéité de la profondeur. Nous allons maintenant voir si pouvons retrouver les locus homologues entre les individus.



**FIGURE 4.4** – A gauche, nombre de locus en fonction du nombre de lectures pour chaque individu de *B. tabaci*. Les points sont alignés et ne forment pas un plateau. A droite : distribution de la couverture par locus pour l'individu 87, représentatif de l'ensemble des individus.

**Clustering inter-individus** Sur la figure 4.5a le nombre de locus retrouvés chez plusieurs individus est représenté. On voit que seulement 133 locus contiennent dix individus ou plus.

Pourquoi *seulement*? Si on se base sur la taille estimée du génome de *B. tabaci*, avec une densité de locus semblable à celle de *D. melanogaster*, par une simple règle de trois on obtient environ 23 000 locus. Bien sûr, la densité n'est pas la même dans tous les génomes et dépend fortement du taux de GC mais malgré tout, on peut dire qu'on passe à coté d'une majeure partie du potentiel des *RADSeq* avec seulement 133 locus.

Dans un deuxième temps nous avons cherché à estimer le nombre de locus total dans le génome de *B. tabaci* en utilisant les locus en commun par paire d'individus. Voici la méthode : Soit  $L_1$  et  $L_2$  le nombre de locus retrouvés respectivement chez les individus 1 et 2 et  $L$  le nombre de locus du génome de *B. tabaci*. Soit  $p_1$  et  $p_2$  les proportions de locus retrouvés chez les deux individus par rapport au nombre total de locus. On a donc :

$$L_1 = L * p_1$$

$$L_2 = L * p_2$$

Faisons l'hypothèse que l'échantillonnage des locus est aléatoire d'un individu à l'autre, le nombre de locus  $L_{12}$  retrouvé chez les deux individus vaut :

$$L_{12} = L * p_1 * p_2$$

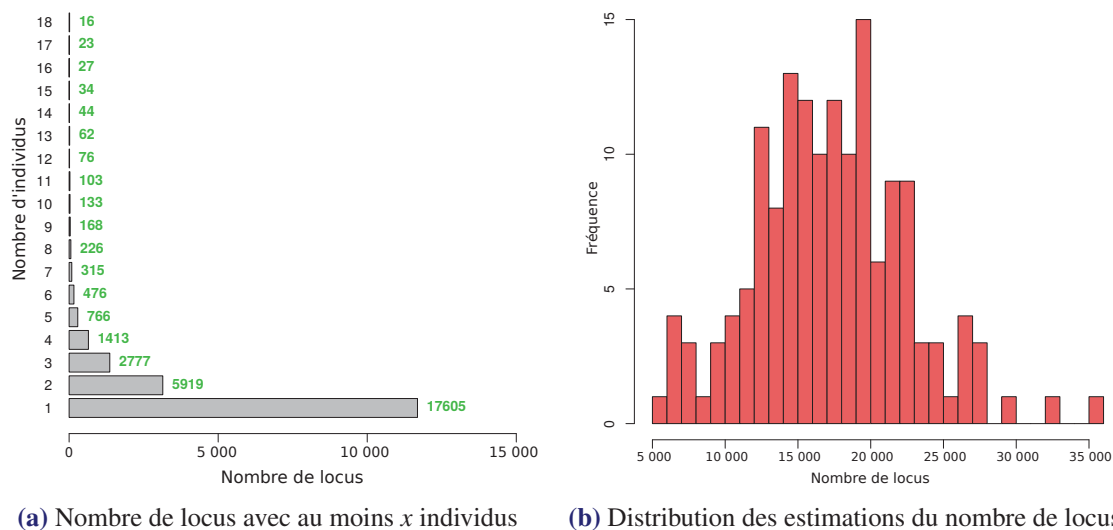
En remplaçant  $p_1$  et  $p_2$  par les deux premières équations, on en déduit facilement le nombre total  $L$  de locus dans le génome :

$$L = \frac{L_1 * L_2}{L_{12}}$$

Nous avons donc appliqué cette méthode à toutes les paires d'individus possibles soit 153 paires possibles pour 18 individus. La distribution de ces estimations est montrée sur la figure 4.5b. Les estimations varient de 5 000 à 35 000 avec un mode situé aux alentours de 15 à 20 000 locus. Avec la même méthode, le nombre de locus estimé chez *D. melanogaster* est d'environ 6 500 soit une légère surestimation du nombre réel. Cette méthode d'estimation par paire paraît donc correcte chez *D. melanogaster*, mais notre estimation pour *B. tabaci* n'est pas précise, probablement car nous retrouvons trop peu de locus pour chaque individu.

En regardant les données de *D. melanogaster* (données non présentées ici) il s'avère que pour retrouver environ 80 % des locus, il faut une couverture moyenne de 250x (à comparer aux 50x obtenus pour *B. tabaci*). On peut alors donner une estimation du nombre d'individus multiplexables pour obtenir la profondeur voulue. Dans le tableau 4.2 sont données quatre estimations du nombre d'individus séquençables sur une même ligne avec différentes hypothèses, sur le nombre de locus dans le génome de *B. tabaci* et sur le nombre de lectures utilisables. Soit la quantité de lecture obtenue dans les prochains séquençages sera la même (76 millions de lectures) ; soit on fait l'hypothèse que les données seront de meilleure qualité et on peut espérer 120 millions de lectures, valeur permise par la technologie. On voit qu'avec ces hypothèses,

## 2. Expérience pilote, multiplexer de nombreux individus à faible coût



**FIGURE 4.5** – (a) : Nombre de locus avec au moins  $x$  individus. Les barres représentent le nombre de locus pour lesquels on a retrouvé au moins  $x$  individus. Par exemple le nombre de locus chez lesquels on retrouve les 18 individus est de 16, et il y a 133 locus où on retrouve au moins 10 individus. Ainsi plus on descend sur la figure, plus on a de locus, mais plus ceux-ci sont susceptibles de contenir des données manquantes, (b) : Distribution des estimations du nombre de locus total chez *B. tabaci* à partir du nombre de locus en commun par paires d'individus.

même dans le meilleur des cas, on peut séquencer seulement 32 individus en même temps. Une possibilité d'amélioration est de diminuer la profondeur moyenne nécessaire (qui serait en l'état de 250x) pour retrouver plus de locus c'est-à-dire qu'il faudrait diminuer l'hétérogénéité de la profondeur entre les locus pour se rapprocher le plus possible d'un tirage aléatoire. Pour cela, on pourrait faire plusieurs PCR en parallèle lors de l'étape d'amplification en fin de préparation de la banque avec chacune un plus petit nombre de cycles.

Lectures (millions)	Locus	
	15 000	20 000
76	20	15
120	32	24

**TABLE 4.2** – Estimation du nombre d'individus multiplexables selon l'estimation du nombre de locus et du nombre total de lectures disponibles suite au séquençage.

### Conclusion sur cette expérience pilote

Tout d'abord nous avons vu qu'utiliser des marqueurs *RADSeq* est possible au sein du laboratoire. Nous avons aussi pu créer des adaptateurs en synthétisant séparément la partie commune et la partie individuelle, nous pouvons donc multiplexer de nombreux individus à moindre coût. En ce qui concerne *B. tabaci*, la faible quantité d'ADN ne nous a pas permis d'obtenir énormément de locus, et nous n'avons pas pu estimer convenablement le nombre total de locus du

génomique. Malgré des données insuffisantes en l'état, deux éléments en particulier nous permettent d'être optimistes. 1) Lors de cette expérience, d'autres modèles biologiques étaient présents dans la même banque avec des quantités d'ADN beaucoup plus élevées, ce qui ne sera pas le cas lors des essais ultérieurs. 2) L'hétérogénéité de la couverture observée d'un locus à l'autre augmente mécaniquement la profondeur moyenne nécessaire au séquençage d'un nombre intéressant de locus, mais la réalisation de plusieurs PCR en parallèle en diminuant le nombre de cycles pourrait être un moyen de résoudre en partie ce problème.

## Développement de l'outil pour *B. tabaci*

### Objectifs de l'essai

Suite à l'essai pilote comprenant plusieurs modèles biologiques, il a été décidé d'effectuer une première expérimentation propre au modèle *B. tabaci*. En plus d'une meilleure estimation du nombre de locus, l'inclusion de populations naturelles allait permettre de réaliser des analyses de génétique des populations proprement dites, contrairement à l'essai pilote. Ce nouvel essai comprenait donc trois populations naturelles, deux lignées de laboratoire, ainsi que deux témoins *D. melanogaster*. Ces derniers permettent de vérifier la réussite de la préparation de la banque et du séquençage car le génome est connu et on peut vérifier la quantité de locus retrouvée par rapport à l'attendu. Ces données allaient permettre de mettre en place un *pipeline* pour l'analyse bioinformatique, du demultiplexage aux analyses de structuration génétiques. De plus, la présence des bactéries symbiotiques allait être recherchée en comparant les lectures de chaque individu aux génomes de référence de ces bactéries. Obtenir le statut d'infection sans effectuer au préalable une PCR permet d'économiser l'ADN si on veut cette information.

### Matériel et méthodes

#### Populations testées

Nous voulions une population naturelle pour laquelle les cytotypes Q1 et Q2 sont en mélange (à Bassens en Savoie) et deux où seulement l'un est retrouvé (respectivement à Moulès dans les Bouches-du-Rhône pour les Q1 et à Publier en Savoie pour les Q2). Ces informations sur la présence des cytotypes proviennent de génotypages préliminaires de quelques individus de ces populations, généralement une dizaine. Ces génotypages ont été réalisés par PCR comme décrit dans le chapitre 1, où ces données ont été intégrées. Du fait du faible effectif de ces génotypages préliminaires, il s'avèrera que la population de Publier (censée être purement Q2) comprend en fait un individu Q1 ! En ce qui concerne les lignées de laboratoire, qui sont les mêmes que celles utilisées auparavant, nous avons inclus trois mâles en plus des dix femelles. En effet, comme *B. tabaci* est haplo-diploïde le génotypage des mâles donne accès aux haplotypes. Par contre les mâles sont encore plus petits que les femelles, diminuant d'autant la quantité d'ADN



disponible ! C'est pour cette raison que seulement 3 individus mâles, par lignée de laboratoire, ont été inclus en guise de test (voir table 4.3).

Acronyme	Localité	Plante	Département	Date de capture	Q1	Q2	♀	♂
MoubA	Moules	Aubergine	13	09-2012	8	0	8	0
PubH	Publier	Hibiscus	74	09-2012	1	7	8	0
Bas2P	Bassens	Poinsettia	73	10-2013	9	7	16	0
Q2Lab	Laboratoire	-	-	-	0	13	10	3
Q1Lab	Laboratoire	-	-	-	13	0	10	3
Dm	Drosophiles témoins						2	0

**Nombre total d'individus : 60**

**TABLE 4.3** – Provenance des individus inclus dans l'essai.

### Extraction d'ADN

Afin de récupérer la plus grande quantité possible d'ADN, et de bonne qualité, quelques essais préliminaires réalisés au laboratoire ont permis d'établir la meilleure méthode : l'extraction d'ADN en elle-même se fait au phénol-chloroforme puis la purification se fait avec le kit *NucleoSpin® Tissue XS* (MACHEREY-NAGEL). Le protocole complet est en annexe D page 227. En tout, l'ADN de 88 individus a été extrait et dosé afin de sélectionner les plus concentrés. L'haplogroupe mitochondrial de tous ces individus a été diagnostiqué par séquençage du *mtCOI*. Cela permet aussi de prendre une proportion équitable de chaque haplogroupe dans la population mixte. Cette PCR est la seule que nous nous sommes autorisées en amont de la préparation de la banque, pour économiser au maximum l'ADN. Nous espérons connaître le statut d'infection individuel grâce aux séquences obtenues.

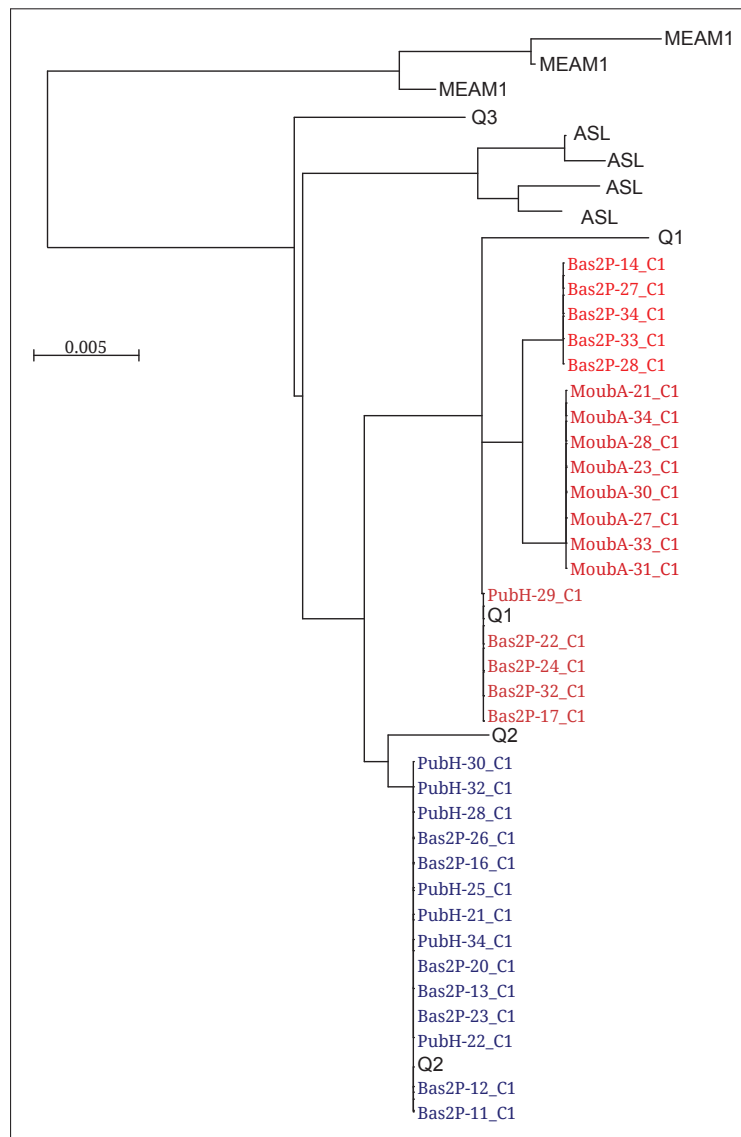
Au final, 58 individus ont été conservés (voir tableau 4.3) avec une quantité moyenne d'ADN restant de 30,5 ng pour les femelles et 14,8 pour les mâles, ce qui est tout de même deux fois moins !

Un arbre de distance réalisé sur le gène *mtCOI* est présenté en figure 4.6, il comprend seulement les populations naturelles. On voit qu'il y a un seul haplotype mitochondriale Q2 à Publier et Bassens, qui sont deux villes en Savoie, et 3 haplotypes Q1 : 1 à Moulès, 1 à Bassens et le dernier est présent à Bassens et Publier.

### Analyse des données

L'analyse des données est similaire à celle de l'essai pilote, car elle comprend les étapes de démultiplexage et de *clustering*, mais ces dernières n'ont pas été effectuées avec le même outil. De plus, des étapes supplémentaires ont été ajoutées, nous allons les expliquer.

**Demultiplexage** Comme dans la partie précédente



**FIGURE 4.6** – Arbre de distance *COI* des individus des populations naturelles. Les haplotypes de référence sont en noir, les haplotypes Q1 de l’essai sont en rouges, l’haplotype Q2 de l’essai est en bleu. L’arbre a été réalisé avec l’algorithme BioNJ (Gascuel 1997) avec la distance de Jukes-Cantor, implémenté dans SEAVIEW.

**Séquences uniques** Dans un souci d'éliminer un maximum de séquences parasites (erreurs de séquençage, erreurs lors de la PCR...), et divers artefacts que nous avons rencontré au fur et à mesure de notre analyse, nous avons conservé que les séquences présentes chez au moins deux individus. Pour cela, aucun mésappariement n'a été autorisé, les séquences devaient être strictement identiques (contrairement aux étapes de *clustering*). Pour réaliser cette étape, un script (en langage *bash*) a été développé.

**Détection des séquences cytoplasmiques** Toutes les séquences ont été comparées aux génomes des symbiotes connus chez *B. tabaci* ainsi qu'au génome mitochondrial afin de les mettre de côté, pour ne garder que les séquences nucléaires. Tous les génomes des symbiotes de *B. tabaci* ne sont pas disponibles. Nous avons donc utilisé le génome d'*Arsenophonus nasoniae* (Wilkes *et al.* 2010) ; le génome de *Wolbachia pipientis* provenant du moustique *Culex quinquefasciatus* (Klasson *et al.* 2008) ; les génomes de *Cardinium* et *Hamiltonella* et *Portiera* provenant de l'espèce MEAM 1. Pour *Rickettsia*, seuls des *contigs* étaient disponibles, provenant de l'espèce MEAM 1 à nouveau.

Ces recherches de similarité ont été réalisées avec l'algorithme BLAST. Nous avons retenu seulement les séquences où au moins 80 bases contiguës avaient une similarité avec l'un des génomes. Pour les séquences ayant une homologie avec deux symbiotes différents, la bactérie ayant le meilleur score a été retenue (*bit score* de BLAST).

Ces séquences seront ensuite traitées à part pour génotyper les symbiotes présents chez chaque individu.

Pour des raisons pratiques, toutes les chimères d'adaptateurs P1 et P2 ont été retirées à ce moment.

**Clustering intra-individuel** Dans l'essai pilote, nous avons effectué cette étape avec STACKS qui à l'avantage d'être très rapide, mais qui ne peut traiter les *indel* (insertions ou délétions), à l'inverse du programme *pyRAD* (Eaton 2014). Ce dernier utilise une méthode d'alignement global, l'algorithme USEARCH (Edgar 2010) alors que STACKS compare base à base les deux séquences en autorisant un certain nombre de mésappariement.

En conséquence, un locus contenant des *indel* sera séparé en plusieurs locus par STACKS (sauf si celui-ci est en fin de séquence), mais bien identifié en tant que locus unique par *pyRAD*. Dans l'exemple suivant, il y a une délétion dans la deuxième séquence : lorsqu'elle est en fin de séquence (à gauche) il y a seulement trois mésappariement, les séquences sont rassemblés dans le même locus (si et seulement si la distance autorisée est d'au moins 3 mésappariements). Lorsqu'elle apparait plus tôt dans la séquence le nombre de mésappariements est beaucoup plus élevé (ici 10), les deux séquences ne seront pas regroupées dans le même locus.

ATGCGATGCTAGATCG	ATGCGATGCTAGATCG
***	*****
ATGCGATGCTAGACGA	ATGCGAGCTAGATCGA

Pour vérifier à quel point ce phénomène pouvait influencer le nombre de locus retrouvés, le nombre de locus obtenus par les deux méthodes a été comparé (le résultat est visible plus loin, figure 4.9a), en essayant de choisir des paramètres similaires pour les deux méthodes. Le résultat semble montrer que le nombre de locus identifié par *pyRAD* est inférieur à celui de *STACKS*, comme attendu.

Nous avons aussi vérifié *a posteriori* le nombre de locus obtenus par *pyRAD* concernés par des *indel*. Parmi les 2 759 locus contenant au moins 42 individus, 732 contiennent au moins un *indel*, pour les 17 338 locus contenant au moins deux individus, 2 905 ont au moins un *indel*. D'après ces chiffres, entre 16 % et 25 % des locus contiennent donc des *indel* et les séquences correspondantes pourraient donc être assignées par erreur à plusieurs locus par *STACKS*.

Le *clustering* a donc été réalisé avec *pyRAD* en utilisant un seuil de 85 % d'identité minimum. Ce seuil est assez bas, mais il permet de rassembler des allèles potentiellement très divergent, sans pour autant rassembler par erreur des séquences paralogues (Communication personnelle), d'autant plus que d'autres filtres vont permettre d'écarter les locus paralogues. Par exemple, on s'attend à ce que la profondeur d'un locus observé, composé en réalité de plusieurs locus paralogues soit dans la queue de distribution. Le filtre mis en place par *pyRAD* est le suivant : la couverture ne doit pas dépasser la valeur la plus élevée entre 500 ou la couverture moyenne plus deux fois l'écart-type.

La couverture minimum pour former un locus est de 10 lectures, valeur qui permet d'écarter des « faux » locus, d'après des analyses préliminaires.

A la fin de cette étape, *pyRAD* estime de façon conjointe le taux d'erreur et le taux d'hétérozygotie. Certains paralogues peuvent être filtrés à cette étape : les locus contenant plus d'un certain nombre de sites hétérozygotes ainsi que ceux ayant plus d'haplotypes que ne le permet la ploïdie du modèle étudié, en l'occurrence deux pour les femelles *B. tabaci*.

**Clustering inter-individuel** *pyRAD* a aussi été utilisé, avec le même seuil de similarité minimum, c'est-à-dire 85 %. Au premier abord on pourrait être tenté de mettre une distance inter-individus plus grande qu'en intra-individu, mais puisque les individus ne sont jamais que des tirages d'allèles dans un ensemble (la population), il n'y a pas de raison que la distance soit plus grande entre deux individus qu'entre deux chromosomes.

**Export des données** Les locus obtenus au cours des étapes précédentes sont filtrés sur le nombre d'individus minimum dans un locus pour conserver celui-ci. En effet, on peut conserver les locus que nous retrouvons chez un maximum individus et obtenir ainsi des données complètes, où alors maximiser le nombre de locus, avec un nombre d'individus minimum plus réduit. Les conséquences de ces choix ne sont pas anodines mais restent très discutées (Rubin *et al.* 2012 ; Takahashi *et al.* 2014). Les données des locus retenus sont exportées grâce à *pyRAD* dans différents formats utilisables par des logiciels de génétique des populations classiques tels que *STRUCTURE*, *GENEPOP* ou encore *R* (R Core Team 2015).

**Analyses de génétique des populations** Nous avons fait une analyse en composante principale (ACP) sur les fréquences alléliques individuelles avec la bibliothèque *adegenet* (Jombart 2008) implémentée sous R (R Core Team 2015).

L'avantage de cette analyse est le faible nombre de pré-requis contrairement à d'autres logiciels comme STRUCTURE qui nécessite un modèle génétique comme l'équilibre d'Hardy-Weinberg (Jombart *et al.* 2009). Un autre avantage est la capacité des analyses multivariées à détecter des structures génétiques hiérarchiques ou des populations peu structurées (Jombart *et al.* 2009), s'apparentant plus à un gradient de connexions (voir par exemple Jombart *et al.* 2008). Dans le contexte des données de séquençage haut-débit et des analyses de type *RADSeq*, la rapidité de mise en œuvre (quelques secondes comparée à plusieurs jours !) est un atout majeur.

Deux aspects sont tout de même à surveiller : le premier est le nombre d'individus par population, il faut que l'échantillonnage soit équilibré pour éviter de déformer l'espace de projection (McVean 2009 ; Ma & Amos 2012) ; le deuxième est la proportion de données manquantes. En effet les données manquantes posent un problème aux analyses multivariées et doivent être remplacées. Généralement, on utilise la moyenne de la variable en question, au risque de « centrer » les individus, ce que nous vérifierons.

## Résultats

### Compte-rendu du séquençage

Les données sont en *paired-end*, ce qui signifie que chaque brin d'ADN est séquencé à ses deux extrémités, sur 100 paires de bases, dont l'une (*read 1*, comme pour l'essai pilote) démarre par le barcode puis le site de restriction. La banque obtenue a été séquencée deux fois car les premières données semblaient indiquer un problème technique (très peu de lectures, et le *read 2* de mauvaise qualité). Le problème est indéterminé à ce jour. On remarque que l'étalon interne (le phage *PhiX* dont le génome est connu) est retrouvé en plus forte quantité qu'il devrait l'être, comme s'il était mieux séquencé que la banque.

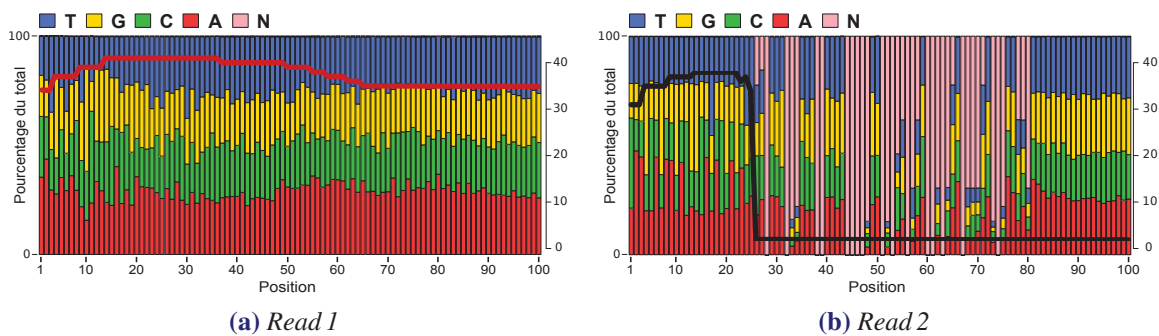
Séquençage	N	Q30 r1	Q30 r2	PhiX <i>pre</i>	PhiX <i>post</i>
1	65,85	88	30	15	43,7
2	67,88	85,6	26	10	30,6

**TABLE 4.4** – Données primaires du séquençage. N : nombre de lectures Q30 r1 et Q30 r2 sont des indices de qualité des deux *reads*. Le PhiX est de l'ADN de phage servant de contrôle interne mélangé avec la banque préalablement au séquençage. Les nombres indiqués sont les quantités injectées avant (*pre*) et les quantités retrouvées après (*post*).

Sur la figure 4.7 sont représentées la qualité et la composition en base des lectures obtenues. Pour le *read 1* la situation est conforme à l'attendu, avec une qualité qui décroît légèrement tout au long de la séquence et des proportions respectives pour les 4 bases s'approchant des 25 %.

Par contre le *read 2* n'est pas du tout conforme à cela puisque la qualité chute drastiquement peu après la 20<sup>ème</sup> base et il y a une sur-représentation de bases indéfinies (*N*), c'est-à-dire avec un signal de trop mauvaise qualité. Nous avons cherché à comprendre ce phénomène, les résultats sont en annexe E page 229. Cette mauvaise qualité est dommageable car le *read 2* peut aider à enlever des duplicats de PCR<sup>1</sup>, qui peuvent biaiser l'estimation de l'hétérozygotie. Comme ces *read 2* sont inexploitable nous ne détaillerons pas le problème.

La conclusion à tirer de ce séquençage est que nous ne pourrons pas utiliser le *read 2*, ramenant l'analyse à du *single-read*, comme dans l'essai pilote. Nous disposons d'environ 130 millions de lectures du fait des deux séquençages (Dans les fait, quand cela marche bien, un seul séquençage permet d'obtenir ce nombre de lectures !).



**FIGURE 4.7** – Qualité et composition en bases des lectures. Chaque rectangle représente une position de la lecture (de 1 à 100) dont les couleurs représentent les proportions de chaque base nucléique à cette position. Les courbes rouges et noires de chaque figure représentent la médiane de la qualité des bases.

Nous avons ensuite analysé les témoins drosophiles pour nous assurer de la qualité et de la quantité de données retrouvées. Les résultats complets sont en annexe F page 231.

Le nombre de locus obtenu approche les 80 % par rapport au génome de référence. De plus, la couverture observée est largement inférieure à celle de l'essai pilote qui était d'environ 250x, on a donc plus d'information avec moins de lectures ce qui est un très bon résultat.

Nous allons maintenant analyser les données de *B. tabaci* exclusivement.

### Démultiplexage et filtrage des lectures

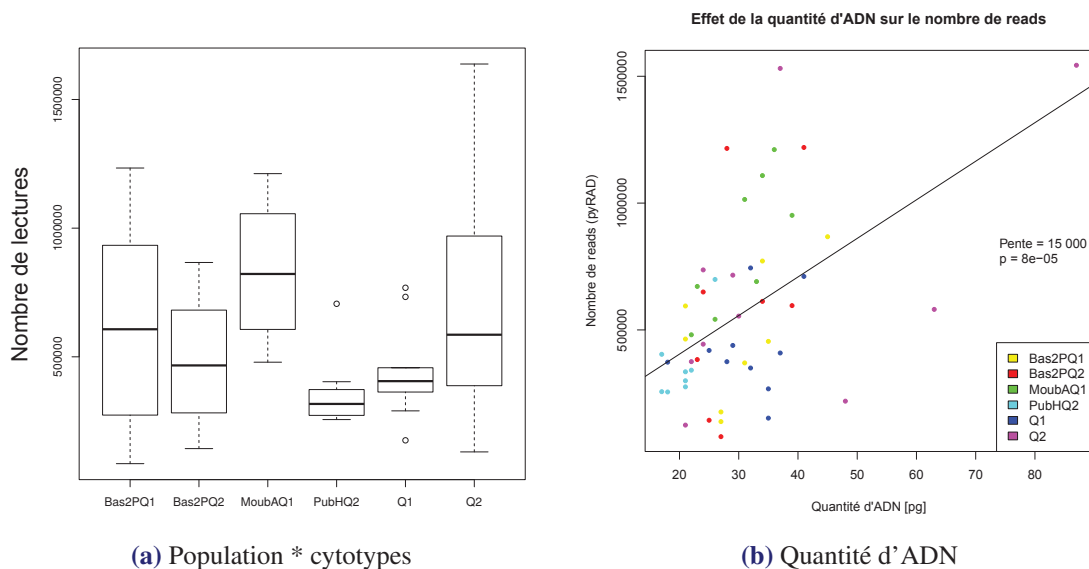
Au final, en rassemblant les données des deux séquençages, environ 32 millions de lectures sont utilisables (avec barcode, site de restriction et de bonne qualité), soit environ 24 %. Le détail est donné dans le tableau 4.5. Le nombre de lectures utilisables est en moyenne de : 510 000 pour les deux témoins drosophiles, 199 000 pour les 6 mâles *B. tabaci* et 577 000 pour les femelles *B. tabaci*. Pour rappel, ce nombre était d'environ 78 000 dans l'expérimentation pilote, soit une augmentation d'un facteur 7, ce qui est une très bonne chose.

1. Les duplicats de PCR sont des fragments issu d'un même brin d'ADN, ils sont donc des copies de la même information

	Total	Barcode	B. + Qualité	B. + Q. + ER
N	133 734 568	41 524 335	38 472 719	32 245 529
%, rap. au total	100	31	28,8	24,1
%, rap. au filtre précéd.	100	31	92,7	83,8

**TABLE 4.5** – Nombre de lectures utilisables après chaque étape de filtrage. B. : barcode, Q. : Qualité, ER : site de restriction, N : nombre de lectures. Les pourcentages sont donnés par rapport au total et par rapport au filtre précédent. La plus grosse perte est due à l'absence de barcode pour environ 90 millions de lectures.

Le nombre de lectures est très variable d'un individu à l'autre, et aussi d'une population à l'autre (voir figure 4.8a). Cette variabilité est fortement corrélée à la quantité d'ADN de chaque individu (figure 4.8b), elle-même influencée par la population : la lignée de laboratoire Q1 et la population de Publier ont moins d'ADN. Cet effet de la concentration sur le nombre de lectures avait déjà été mis en évidence lors de l'expérimentation pilote.



**FIGURE 4.8** – Influence de la quantité d'ADN, de la population et du cytotype sur le nombre de lectures par individu.

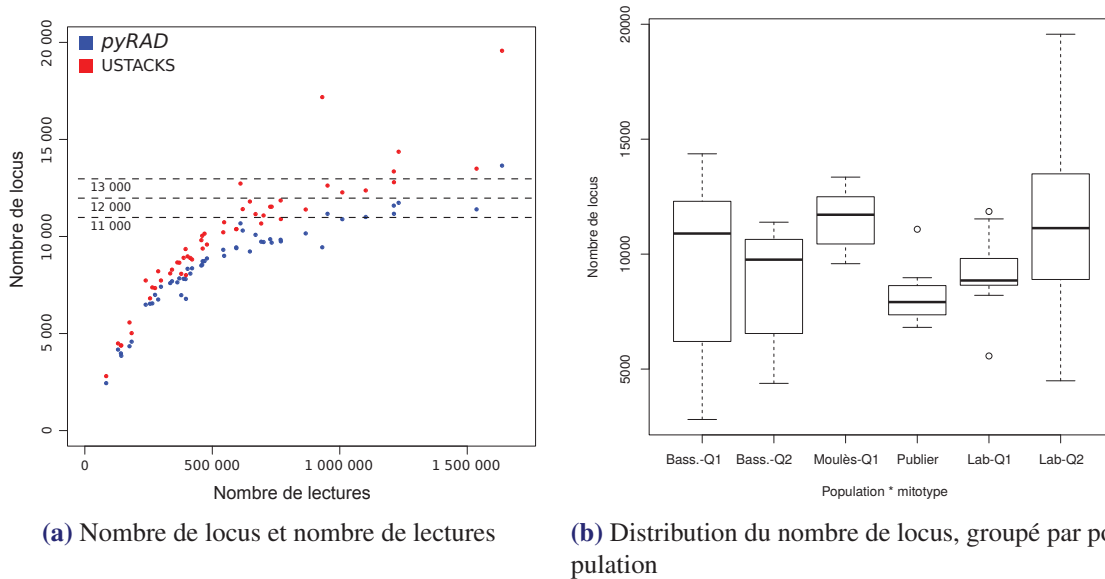
### Clustering : combien de locus ?

Le nombre de locus moyen est 9 833, avec un minimum de 2 807 et un maximum de 19 570. La courbe d'accumulation (nombre de locus en fonction du nombre de lectures, voir figure 4.9a) semble atteindre un plateau à environ 12 000 locus (estimé grossièrement). Pour les mâles, ce nombre varie entre 3 300 et 8 400, pour une moyenne de 5 900, ce qui est bien inférieur aux valeurs obtenues pour les femelles. La quantité d'ADN est probablement le facteur explicatif. La comparaison entre mâles et femelles pour les populations de laboratoire (les seules pour



lesquelles on a des mâles) est sans appel :  $q_{m\grave{a}le} = 14,8 \pm 5,7 ng$  et  $q_{femelle} = 34,9 \pm 7,5 ng$ . D'autres études montrent que le seuil de 10 ng est un minimum pour des analyses *RADSeq* (Communication personnelle), nous sommes donc à la limite. L'étude de mâles par *RADSeq* paraît donc encore difficile, et éloigne la perspective d'obtenir l'accès direct aux haplotypes, chose impossible avec les femelles.

Pour conclure sur cette partie, le nombre de locus retrouvé est assez important et semble atteindre un plateau, nous pouvons estimer le nombre de locus total à environ 12 000. Par contre, il semble difficile d'utiliser les mâles dans les analyses suivantes.



**FIGURE 4.9** – Nombre de locus obtenus lors du clustering intra-individuel. En a), le nombre de locus est représenté en fonction du nombre de lectures. Les résultats obtenus par *ustacks* ou *pyRAD* sont comparés. En b), distribution du nombre de locus par population. Bass. : Bassens

### L'hétérozygotie est fonction du nombre de lectures

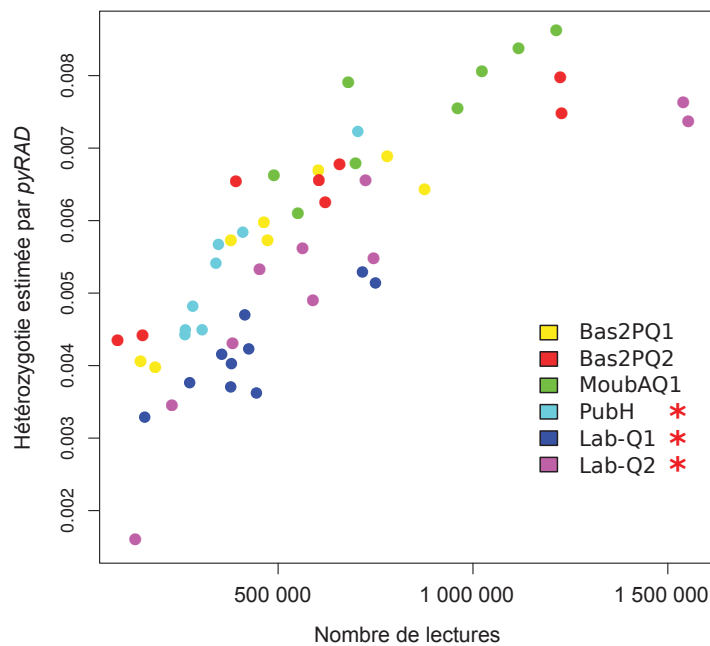
Une fois les locus créés, il faut détecter les bases variables, c'est le *SNP-calling*, réalisé avec *pyRAD*.

Il s'est avéré que la densité de SNP le long des locus était hétérogène : celle-ci augmente plus ou moins fortement en fin de locus, nous avons donc mené des analyses pour estimer ce problème (résultats présentées en annexe G page 233). La solution qui a été retenue pour résoudre ce problème est de couper les locus, mais seulement à la fin du processus, après le *clustering* et non de couper les séquence en amont. Les SNP dont la position était supérieure à 79 dans la séquence ont donc été écartés suite à ces tests, pour les toutes les analyses en aval.

Comme nous l'avons vu, les individus n'ont pas tous le même nombre de lectures, les valeurs sont d'ailleurs très variables s'échelonnant d'environ 150 000 à 1 500 000 pour les femelles. On s'attend donc à ce que pour les individus les moins couverts, certains allèles ne soient pas récupérés. Il y a plusieurs cas de figure : 1) Aucun des deux allèles est récupéré à un locus

et dans ce cas, il y aura une donnée manquante 2) Dans le cas d'un locus homozygote, seul un allèle est récupéré et le résultat ne change pas 3) L'individu est hétérozygote à ce locus et seulement l'un des allèles est récupéré. Dans ce dernier cas, l'erreur est de considérer le locus comme homozygote étant donné qu'on a qu'un seul allèle à notre disposition et aucun moyen de savoir qu'il y a en réalité deux allèles différents. Nous avons donc regardé l'évolution de l'estimation de l'hétérozygotie faite par *pyRAD* en fonction du nombre de lectures. Les résultats sont présentés sur la figure 4.10. Comme on pouvait s'y attendre, l'estimation de l'hétérozygotie augmente de façon significative avec le nombre de lectures (glm,  $p < 0,001$ ). Ce problème a déjà été mis en évidence (Gautier *et al.* 2013) et peut être aussi causé par des sites de restriction mutés et donc absents (allèles nuls). Sous ce terme se cache simplement le fait qu'à un locus, un des deux allèles contient le site de restriction et l'autre non.

Une solution qui pourrait être envisageable est d'échantillonner les lectures pour homogénéiser cette quantité d'un individu à l'autre. Tous les individus auraient donc le même biais de représentation en terme de nombre de séquences et donc le même biais de diversité. Dans notre cas, comme il y a une forte hétérogénéité dans le nombre de lectures, normaliser vers le bas tous les individus représenterait une très grande perte de données et n'est donc pas envisageable.



**FIGURE 4.10** – Hétérozygotie en fonction du nombre de lectures. Les populations marquées par une astérisque rouge sont significativement différentes des autres ( $p < 0,5$ ).

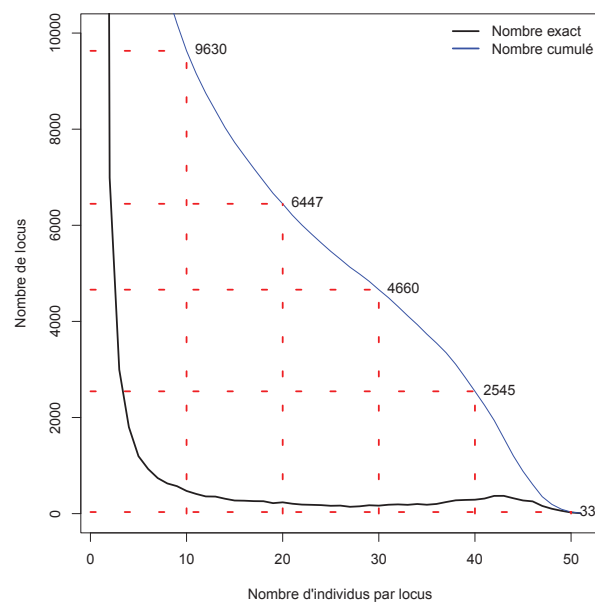
### **Clustering inter-individus, combien de locus partagés ?**

Après avoir fait un *clustering* intra-individuel, on réalise un *clustering* inter-individus pour trouver les locus partagés, l'objectif étant de ne conserver que des locus orthologues.

Aucun locus n'est retrouvé chez les 52 femelles et seuls 9 contiennent 51 individus, ce qui est très faible mais plutôt logique. En effet, il paraît peu probable de retrouver un très grand

nombre de locus chez tous les individus car il y a, en plus de la variabilité expérimentale, de la variabilité biologique individuelle et donc de la variabilité dans la présence d'une partie des sites de restriction. De plus, une plus grande matrice contenant énormément de données manquantes peut-être beaucoup plus résolutive qu'un petit jeu de données dont tous les individus sont génotypés à chaque locus, même si cela est discuté dans la littérature (Rubin *et al.* 2012 ; Takahashi *et al.* 2014). En baissant alors ce seuil du nombre d'individus minimum par locus à 40, on obtient 2 500 locus et environ 10 000 avec au moins 10 individus (voir figure 4.11).

En conclusion de cette partie, le nombre de locus utilisable s'élève à plusieurs milliers ce qui est un point très positif pour les analyses de génétique des populations, en comparaison des sept locus microsatellites utilisés dans le chapitre 2.



**FIGURE 4.11** – Locus partagés entre les individus. La courbe noire indique le nombre de locus qui contiennent exactement  $n$  individus, la courbe bleue indique le nombre de locus qui contiennent au moins  $n$  individus. Le nombre de locus ayant au moins 10, 20, 30, 40 et 50 individus est indiqué.

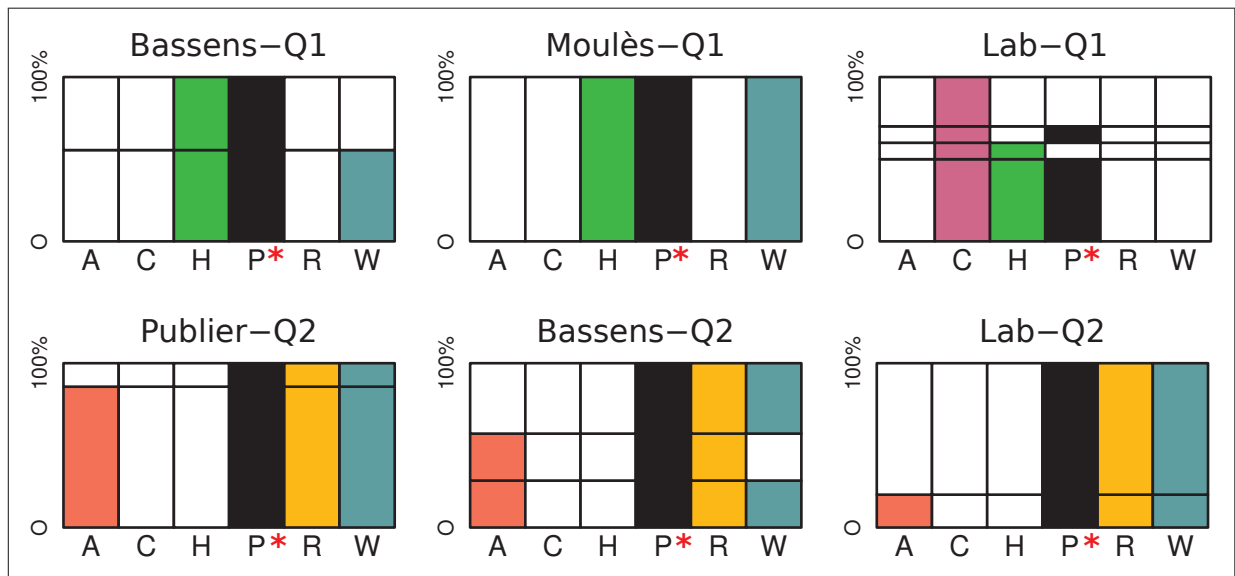
### Détection des symbiotes

Dans cette partie nous allons essayer de détecter les symbiotes dans les données de séquençage. Sur un total de 26 279 546 lectures, environ 590 000 ont une similarité avec au moins un des symbiotes, soit environ 2 %. Toutes ces séquences ont été rassemblées en locus grâce à la position sur le génome du symbiote renvoyée par *BLAST*.

Comme d'autres individus de toutes ces populations avaient été par ailleurs soumis à la détection de symbiote par PCR classique, on avait une très bonne idée de ce qu'on devait trouver potentiellement dans chaque population. La détection des symbiotes est globalement très bonne avec nos données de séquençage *RADSeq* (Aucun symbiote est retrouvé là où il ne devrait pas être, voir figure 4.12), à deux remarques près. La première est que *Portiera* n'est pas détecté dans 4 individus alors qu'il est obligatoire, tous ces individus proviennent de la lignée Q1 de

laboratoire. La deuxième difficulté rencontrée est venue lors de la détection d'*Arsenophonus* car de nombreux individus Q1 semblaient contenir ce symbiote, ce qui n'a jamais été observé jusqu'à présent et pas plus dans les génotypages préliminaires de cette population sur d'autres individus (voir chapitre 1). En effectuant un *clustering* sur ces séquences, il s'avère que les locus apparentés à *Arsenophonus* des individus Q1 et Q2 ne sont pas les mêmes. Après de plus amples analyses d'homologie par *BLAST* sur *GenBank*, il se pourrait finalement que ces lectures appartiennent au symbiote de *Plautia stali*.<sup>1</sup>

On a donc peut-être affaire à un nouveau symbiote chez *B. tabaci*, et des amorces spécifiques pourront être utilisées ultérieurement pour chercher ce symbiote chez *B. tabaci*.



**FIGURE 4.12** – Infections multiples des populations analysées par *RADSeq*, les cyotypes Q1 sont sur la première ligne, les cyotypes Q2 sur la deuxième. Aucun symbiote n'est retrouvé où il n'est pas attendu et tous les symbiotes sont retrouvés dans chaque population où il avait été détecté par PCR. Les astérisques rouges rappellent que *Portiera* est obligatoire pour l'hôte.

Le nombre de lectures pour chaque symbiote est résumé dans le tableau 4.6. Aucune séquence ne ressemble à *Fritschea* (peu de données de référence disponibles de toute façon). Nous n'avons conservé pour un individu donné que les locus ayant au moins trois lectures, et pour les bactéries nous n'avons conservé que les locus retrouvés chez au moins deux individus.

Le nombre de locus retrouvé est assez proche de l'attendu (voir les lignes 3 et 4 du tableau 4.6) si on considère l'ensemble des individus, à l'exception d'*Arsenophonus* où seulement 5 des 22 locus attendus sont retrouvés. Étant donné que ce génome ne correspond pas exactement à celui retrouvé chez *B. tabaci*, ce résultat n'est pas forcément étonnant. Le nombre de locus moyen par individu est lui légèrement plus faible, et fortement dépendant du nombre de lecture de chaque individu. En effet, le nombre de lecture appartenant à un des symbiotes pour chaque individu est très bien corrélé au nombre de lectures totale pour cette individu, et ce malgré la forte

1. *Plautia stali* (Scott, 1874) [Hémiptère : Pentatomidae] est une punaise polyphage qui attaque les arbres fruitiers dans le nord est de l'Asie.

	ARS	CARD	HAM	PORT	RICK	WOL	Total
Lectures	18	8	240	10	140	150	566
Nb. Ind	12	10	24	48	24	35	-
Loc. total	5	11	45	2	12	29	-
Attendu	22	8	44	2	12	30	-
Locus / Ind	3,0	7,9	37,5	3,4	11,0	16,9	-
Couverture par locus	14	74	210	1034	502	110	-

**TABLE 4.6** – Nombre de lectures pour chaque symbiote (en milliers), ainsi que le nombre de locus moyen par individu, en excluant les individus n’ayant pas de lectures pour ce symbiote. *Attendu*, nombre de locus *sbfl* retrouvé dans ce génome par recherche *in silico*. *Nb. Ind*, Nombre d’individus infectés. *Couverture*, couverture moyenne par locus.

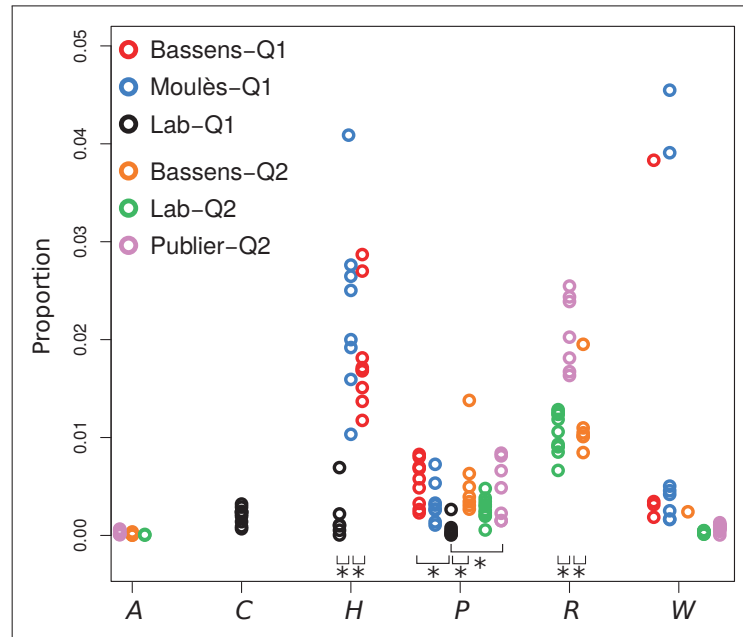
hétérogénéité de couverture entre locus que nous avons évoquée dans les parties précédentes. Cette corrélation est significative (au seuil de 5%), sauf pour les bactéries *Arsenophonus* et *Wolbachia*. Ces deux dernières sont justement celles pour lesquelles il n’y a pas le génome du symbiote chez *B. tabaci*. Cela laisse entendre un problème de reconnaissance de ces locus pour ces génomes, même si au final la détection des infections par cette méthode reste très bonne.

La couverture moyenne par locus est très variable d’un symbiote à l’autre et reflète très probablement leur densité dans l’hôte. On observe un facteur 10 entre *Arsenophonus* et *Wolbachia* et encore un facteur 10 entre cette dernière et *Portiera*. Nous avons donc normalisé le nombre de lectures de chaque symbiote par le nombre de lectures par individu pour obtenir un proxy de la densité. Les valeurs montrent que ce proxy de la densité de chacun des symbiotes est corrélée, c’est-à-dire qu’un individu qui a une forte proportion de lecture d’un symbiote aura aussi une forte proportion de lectures des autres symbiotes (les valeurs s’échelonnent de 40 à 80% de corrélation).

Nous avons aussi comparé ce proxy de la densité de chacun des symbiotes pour chaque individu afin de voir si des différences sont observables entre populations (tests de comparaisons multiples effectués avec la bibliothèque *multcomp* (Hothorn *et al.* 2008) implémentée dans R). On observe des différences significatives pour trois bactéries : *Hamiltonella*, *Portiera* et *Rickettsia* (voir figure 4.13). En ce qui concerne *Hamiltonella* et *Portiera*, dans les deux cas c’est toujours la population de laboratoire Q1 qui montre une moindre densité de symbiote comparée aux autres populations. Dans le cas de *Rickettsia*, la situation est différente puisque c’est une population naturelle qui montre la plus grande densité, comparée à la population de laboratoire et l’autre population naturelle de cytotype Q2.

Dans tous les cas il faut rester très prudent quant à l’interprétation de ces résultats qui mériteraient d’être confirmé par PCR quantitative pour s’assurer que le nombre de lecture du symbiote est un bon proxy de sa densité.

En conclusion de cette partie, nous avons vu que les symbiotes sont assez facilement détectables, mis à part pour *Arsenophonus* et *Wolbachia*. On retrouve bien *Cardinium*, *Hamiltonella*



**FIGURE 4.13** – Proportion de lectures d'un individu appartenant à chaque symbiote, en fonction de la population. Les six symbiotes recherchés sont en abscisse, les individus d'une même population sont sur une même ligne verticale et de même couleur. Les astérisques noires indiquent une différence significative de densité entre deux populations.

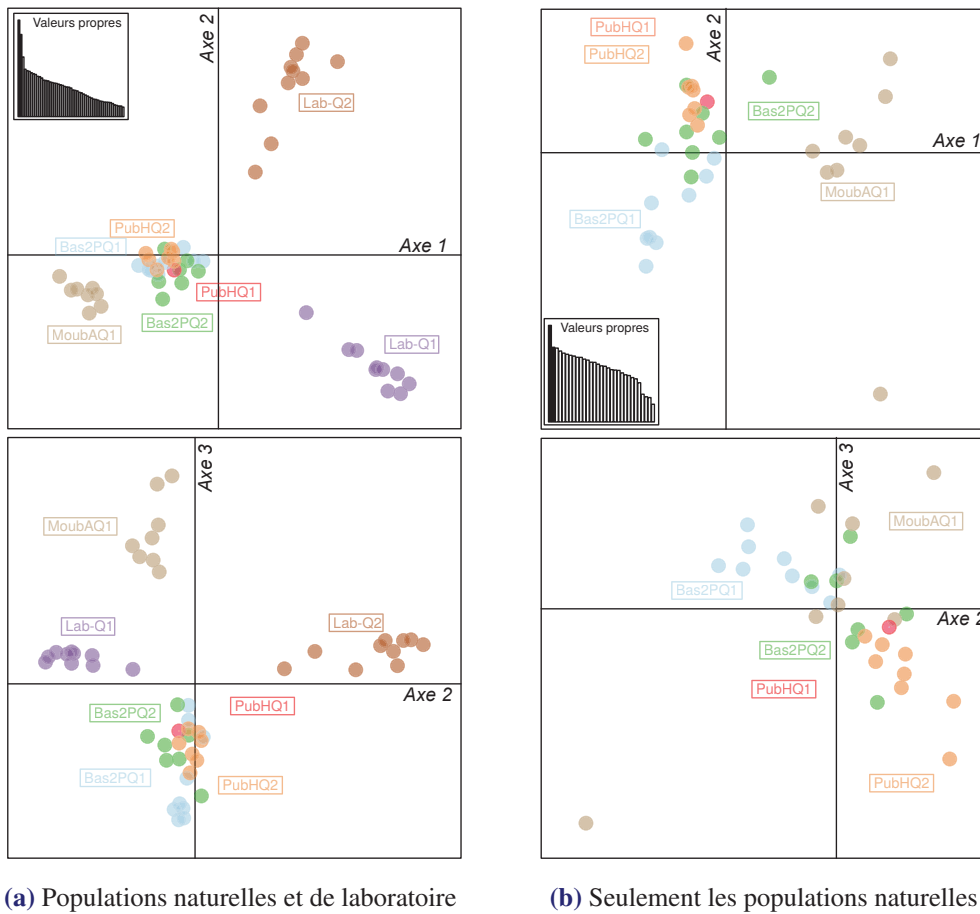
et *Wolbachia* dans le Q1 ; *Arsenophonus*, *Rickettsia* et *Wolbachia* dans le Q2. Dans le futur, avec la publication de tous les génomes des symbiotes de *B. tabaci*, cette recherche ne peut que s'améliorer. Cela permet d'accéder au statut d'infection individuel et ainsi on peut économiser les PCR de détection des symbiotes en amont de la préparation de la banque, et en faire une seule pour le génotypage de l'haplogroupe mitochondrial.

### Structure génétique

Comme nous l'avons vu, tous les individus n'ont pas le même nombre de locus car ils n'ont pas le même nombre de lectures. Lors des analyses subséquentes, il va falloir trouver un compromis entre utiliser seulement les locus génotypés chez tous les individus (ou au moins un nombre d'individus minimum assez grand), ou alors mettre la barre moins haute, et donc utiliser plus de locus au prix d'une plus grande quantité de données manquantes (voir le nombre de locus communs entre individus à la fin de la partie 3.3.5 et la figure 4.11).

**Analyse en composante principale** Plusieurs essais ont été conduits en changeant le nombre d'individus minimal par locus (2, 3, 13, 26, 31, 36, 42). Pour des seuils très bas, nous avons remarqué qu'il était difficile de différencier les populations naturelles car la variance observée dans certaines populations était très importante. Nous présentons donc ici les résultats pour un seuil minimum de 26 individus (tous les autres résultats sont visibles en annexe H page 235). La projection des individus est représentée sur la figure 4.14 et montre notre capacité à distinguer facilement les différentes populations. Sur la figure 4.14a, le premier axe distingue clairement

les populations de laboratoire des populations naturelles, et de façon moindre la population de Moulès (dans les Bouches-du-Rhône, 13). L'axe deux permet quant à lui de distinguer principalement les deux populations de laboratoire. La figure 4.14b a été réalisée sans les populations de laboratoire pour éclaircir les degrés de connexion des populations naturelles. On distingue clairement encore la population naturelle de Moulès (13) des autres, et sur la figure du bas, les axes deux et trois permettent de distinguer un gradient entre les trois autres populations, même si elles ne sont pas clairement isolées.



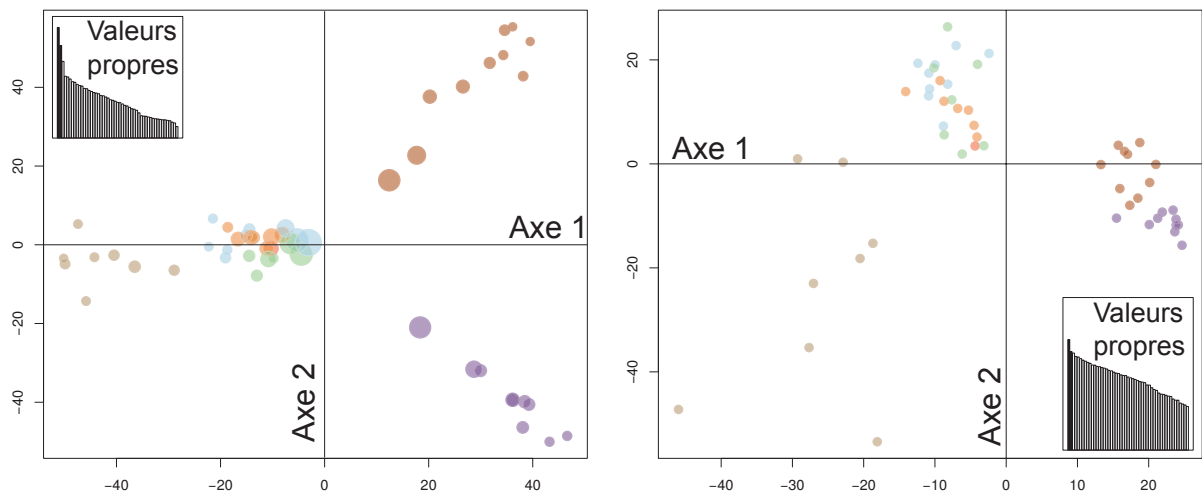
**FIGURE 4.14** – Plan factoriel de l'analyse en composante principale effectuée avec les locus présents chez au moins 26 individus. La figure de gauche est le résultat de l'analyse du jeu de données complet, la figure de droite avec seulement les populations naturelles. Pour chacune des deux figures, sont représentés en haut l'axe 1 et 2, en bas l'axe 2 et 3. La décroissance des valeurs propres est indiquée sur les diagrammes en bâtons pour chaque ACP.

Dans un second temps nous avons testé l'effet des données manquantes sur la structure de ce nuage de points. Nous avons utilisé le jeu de données dont les locus sont présent chez au moins deux individus, soit 7 235 locus en tout. Il y a beaucoup de données manquantes, l'individu en ayant le plus en a 80 %, soit seulement 1 354 locus utilisables pour celui-ci. Nous avons effectué une première ACP, visible sur la figure 4.15 de gauche, sur laquelle est représentée la quantité de données manquantes (plus le point est gros, moins il y a de données). Et en effet, tous les individus centraux sont ceux qui ont le moins de données, ce qui donne cette forme linéaire



pour les nuages de points des populations de laboratoire (en marron et violet sur la droite des graphiques).

Dans un second temps nous avons échantillonné des locus au hasard pour tous les individus de façon à ce qu'il n'en reste plus que 1 354 par individus, le nombre minimum de locus observé pour un individu. Avec cette nouvelle analyse, on peut toujours distinguer toutes les populations (de laboratoire ou naturelles) mais avec moins de puissance, les nuages de points sont plus chevauchant (figure 4.15, à droite). En ce qui concerne la population de Moulès (13), en bas à gauche sur l'image, les individus sont plus dispersés comme s'ils étaient peu homogènes en comparaison de l'analyse précédente.



**FIGURE 4.15** – Influence des données manquantes sur l'ACP. A gauche, l'ACP a été réalisée en conservant tous les locus, même ceux contenant peu d'individus. La taille des points est proportionnelle à la quantité de données manquantes pour un individu. A droite, même ACP mais réalisée en échantillonnant les locus pour que tous les individus en aient le même nombre. les couleurs sont les mêmes que sur la figure 4.14 : Marron foncé : Laboratoire-Q2, Violet : Laboratoire Q1, Bleu : Bassens-Q1, Vert : Bassens-Q2 , Orange : Publier-Q2, Beige : Moulès-Q1.

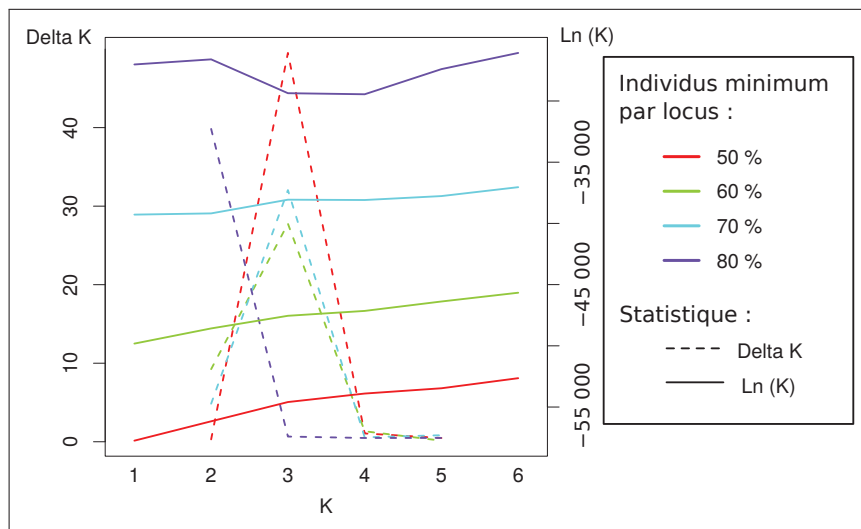
En conclusion de cette partie, d'une part sur le plan technique on voit que l'analyse en composante principale est facilement réalisable avec des données de *RADSeq*. D'autre part, comme attendu on différencie très bien les populations de laboratoire des populations naturelles. Puis, parmi ces dernières la population de Moulès (13) est fortement différenciée des autres qui se trouvent toutes les deux en Savoie. Nous allons maintenant utiliser STRUCTURE pour déterminer le nombre de groupes génétiques.

**Nombre de groupes génétiques** Nous avons cherché à estimer le nombre de groupes génétiques avec STRUCTURE (Pritchard *et al.* 2000), en incluant dans un premier temps toutes les populations, puis en retirant les populations de laboratoire, comme nous l'avons fait avec l'ACP. Les premières analyses ont montré que les populations de laboratoire se détachaient clairement des autres, comme attendu.

Nous allons donc présenter ici les résultats obtenus pour les seules populations naturelles. En-

core une fois, l'effet de la proportion de données manquantes sur la capacité du programme à retrouver une structuration a été testée. Les seuils de nombres d'individus minimums par locus testés sont : 50, 60, 70 ou 80 % des individus qui étaient présents au minimum dans un locus donné pour le conserver. Les nombres de locus étaient donc respectivement de 4 578, 3 946, 3 268 et 2 527 pour ces quatre jeux de données (sur  $\approx 95\,000$  locus en tout).

Pour chaque jeu de données (4), nous avons testé de  $K = 1$  à  $K = 6$  groupes génétiques. Pour chacun des 16 tests (4 jeux de données \* 4 nombres de groupes), 15 simulations indépendantes ont été effectuées, avec 500 000 MCMC dont les 50 000 premières ont été écartées. Cela fait donc 360 simulations en tout. Nous avons ensuite calculé le  $\Delta K$  (Evanno *et al.* 2005) qui est présenté sur la figure 4.16 avec la vraisemblance de chacune des simulations. Le  $\Delta K$  indique que le nombre de groupe optimum est atteint pour  $K = 3$  sauf pour le jeu de données comprenant le moins de données manquantes mais aussi le moins de locus (au moins 80 % des individus dans un locus pour le conserver) où le  $K$  est de 2. On augmente donc la puissance de l'analyse avec le nombre de locus. Avec ces données, le  $\Delta K$  semble plus indiqué pour déterminer le nombre de groupe, car on obtient un pic net à  $K = 3$  contrairement à la vraisemblance qui ne montre pas de pic ou de changement brutal ; contrairement aux microsatellites (voir chapitre 2).



**FIGURE 4.16** – Moyenne des vraisemblances des simulations obtenues par STRUCTURE ainsi que les  $\Delta K$  obtenus par la méthode de Evanno *et al.* (2005). Le taux de données manquantes est indiqué par la couleur.

Par contre, en regardant la figure 4.17, et plus particulièrement les sous-figures 4.17c, d et h, on voit que les simulations sont moins reproductibles lorsqu'il y a moins de locus : certains individus ne sont pas toujours affectés au même groupe. Cela semble indiquer qu'il vaut mieux ajouter plus de locus dans l'analyse, même s'ils ont plus de données manquantes. Cette hypothèse est à prendre avec précaution car en comparant les sous figures e et f, il semblerait que certains individus soient détectés comme hybrides en f mais pas en e. Le fait d'augmenter notre capacité à discriminer peut donc nous faire passer à côté d'autres informations.

En ce qui concerne l'interprétation des résultats, la population de Moulès (13) forme un groupe

génétique à elle toute seule, et c’est aussi la plus isolée géographiquement puisque dans le sud de la France, contrairement aux deux autres qui sont en Savoie. Un deuxième groupe est formé des individus de Publier (donc en Savoie), quant au troisième, il est formé par des individus de Bassens (en Savoie aussi). Dans cette dernière population, plusieurs individus sont des hybrides (à différents degrés) entre ce même groupe et la population de Publier. Parmi ces hybrides, la majorité sont de cytotype Q2, comme la population de Publier (sauf un individu qui est de cytotype Q1), ce qui pourrait laisser penser que le flux génétique vient de cette dernière population, vers Bassens. Comme il y a peu d’individus, on se gardera de conclure trop rapidement sur cette information.

Nous avons mis en évidence de la structuration génétique entre ces différentes populations, nous allons maintenant essayer de quantifier cette différenciation avec des mesures de diversité génétique.

**Estimation de la différenciation** Nous avons quantifié la différenciation entre ces populations en calculant des  $F_{ST}$  par paires de populations. Pour cette analyse, ce sont les haplotypes qui ont été utilisés comme allèles et non les SNP. Les résultats sont présentés dans le tableau 4.7. La distance génétique entre les deux populations savoyardes est d’environ 3 %, ce qui est inférieur à la distance entre ces deux populations et celle du sud, Moulès. Dans le chapitre sur les analyses microsatellites, nous avons trouvé des valeurs non significatives entre les populations d’un même pays, et des valeurs significatives de l’ordre de 1 à 2 % entre les populations espagnoles et françaises. Il semblerait donc que les valeurs trouvées soient supérieures à ce que l’on trouve avec des microsatellites, mais c’est attendu en augmentant le nombre de locus.

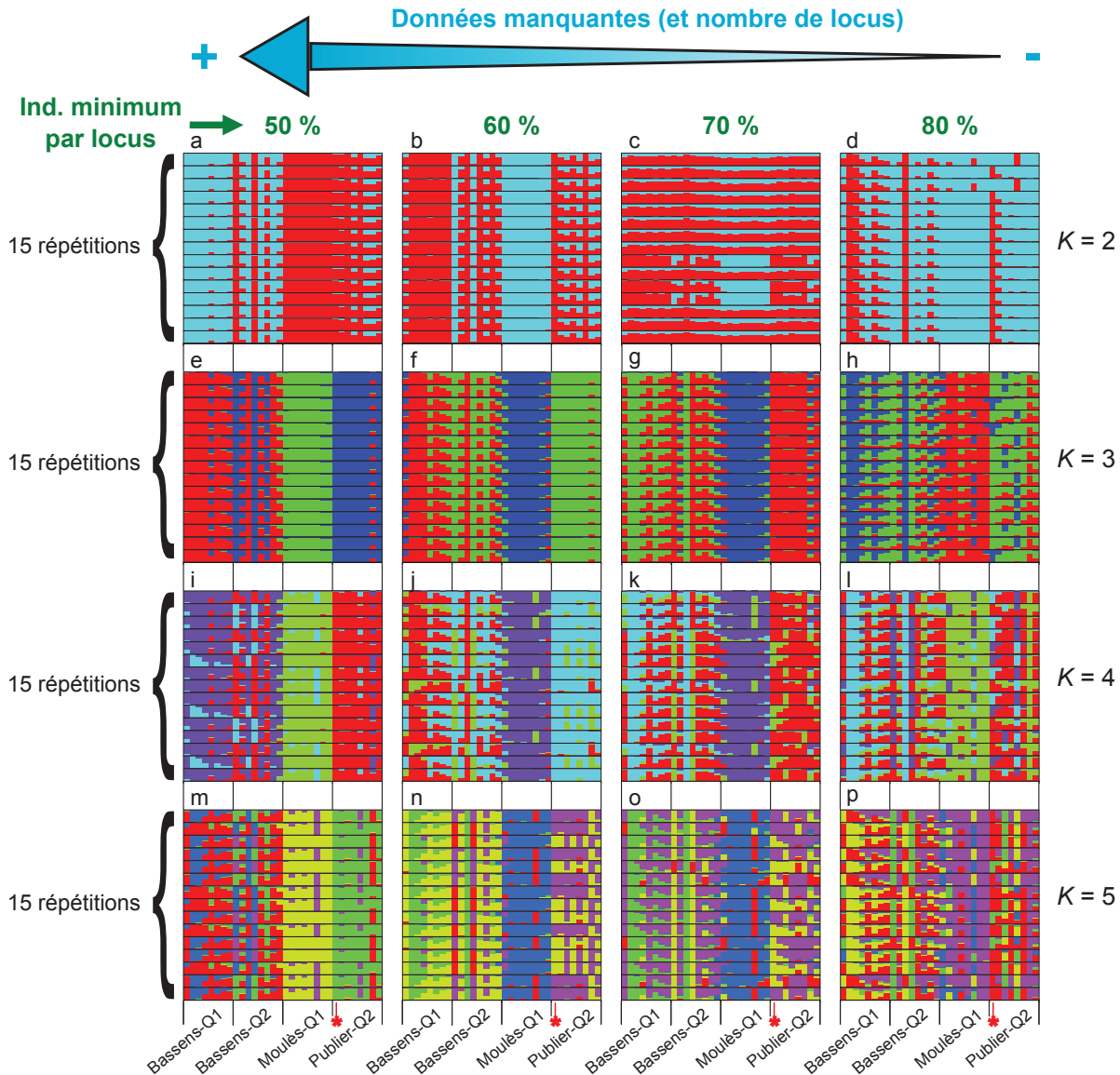
Pop	Bassens	Publier
Publier	0,0282	
Moulès	0,0439	0,0486

**TABLE 4.7** – Estimation des  $F_{ST}$ , avec les locus comprenant au moins 36 individus, pour les trois populations naturelles. Réalisé avec *Genepop* (Raymond & Rousset 1995). Toutes ces valeurs sont significatives ( $p < 0,001$ ).

Dans un second temps nous avons effectué une AMOVA (voir tableau 4.8) à l’aide de la bibliothèque *Poppr* (Kamvar *et al.* 2014), implémentée dans R. Le plus haut niveau hiérarchique considère deux populations (nord et sud) : l’une étant celle de Moulès, l’autre comprenant Publier et Bassens. Le deuxième niveau hiérarchique étant les trois populations naturelles. La variation entre la Savoie et le sud représente seulement 3 %. En fait la majorité de la variation se situe aux niveaux intra-population (38 %) et même intra-individuel (57 %).

### Déséquilibre nucléo-cytoplasmique

Dans les chapitres précédents nous avons décrit plusieurs cytotypes de l’espèce MED chez *B. tabaci*. La présence de certaines bactéries retrouvées tout le temps ensemble pose plusieurs ques-



**FIGURE 4.17** – Probabilités d’assignement des individus aux différents groupes. De haut en bas, pour respectivement 2, 3, 4 et 5 groupes. De gauche à droite en utilisant les locus retrouvés dans un minimum de 50, 60, 70 et 80 % des individus. Chacune des 16 images est composée de 15 itérations réalisées avec les mêmes paramètres. Les individus sont regroupés par population et par cytotype. L’astérisque rouge en bas indique le seul individu de Publier de cytotype Q1. Cette figure a été réalisée à l’aide de STRUCTURE HARVESTER vA.2, Juillet 2014 (Earl & VonHoldt 2011), CLUMPP v1.1.2 (Jakobsson & Rosenberg 2007) et R.

tions sur la coévolution entre leur hôte et elles-mêmes. Si l’infection est suffisamment ancienne, on peut imaginer que la coadaptation des génomes nucléaire et symbiotique ait sélectionné chez l’hôte certains allèles. Si de tels phénomènes existent, leur présence devrait être détectable dans le génome.

En effet, si une combinaison a été sélectionnée dans un cytotype qui se trouve au contact d’un autre, l’hybridation des deux compartiments nucléaires pourrait casser ces co-adaptations. Les génotypes issus de telles recombinaisons devraient donc être contre-sélectionnés et leur fréquence diminuer en conséquence. Ces combinaisons peuvent être à une fréquence plus élevée

Compartiment	$\sigma$	%	$p$
Savoie vs sud	11,78	2,96	0,66
Entre pop. en Savoie	7,98	2,01	< 0,001
Intra population	151,78	38,23	< 0,001
Intra individus	225,42	56,78	< 0,001
Variation total	396,97	100	

**TABLE 4.8** – Analyse hiérarchique de la variance moléculaire.  $\sigma$  : valeur de la variance, % : pourcentage de variance par rapport au total.  $p$  : *pvalue*

qu'attendue dans le cas d'épistasie positive, et moins élevée dans le cas d'interactions négatives entre les allèles.

La situation en France se prête à ce genre d'analyse puisque nous avons en présence les cytotypes Q1 et Q2 qui étaient préalablement séparés géographiquement. Cette séparation a pu permettre l'émergence d'adaptations des hôtes à leur symbiotes. Dans le chapitre 2 nous avons vu que l'hybridation entre les deux cytotypes Q1 et Q2 semblait asymétrique : le fond cytoplasmique Q2 est conservé au dépend de son fond nucléaire remplacé par celui de Q1, résultat comparable à celui de Gauthier *et al.* (2014). Dans ce cas précis, si les symbiotes de Q2 (par exemple *Rickettsia*) étaient co-adaptés au fond nucléaire de Q2, les hybrides avec un mitotype Q2 devraient présenter à certains locus nucléaires un déficit de diversité, avec des allèles issus du fond nucléaire Q2.

Le plan expérimental idéal serait deux populations ancestrales de Q1 purs et Q2 purs, ainsi que des populations avec les deux cytotypes en mélange (plusieurs situations mixtes indépendantes seraient un plus). Les populations que nous avons échantillonnées ne suivent pas ce plan parfait, mais néanmoins nous avons tout d'abord testé la présence de déséquilibre en utilisant les deux populations où seulement l'un des cytotypes est retrouvé (Moulès avec seulement du Q1 et Publier où l'on retrouve en fait un individu Q1 parmi les Q2). Ce déséquilibre a été recherché entre allèles nucléaires et haplogroupe mitochondrial Q1 ou Q2, ce qui revient à chercher aussi pour les bactéries *Hamiltonella* et *Rickettsia* dans le même temps puisque celles-ci sont fixées dans chacun des deux cytotypes respectivement et donc en liaison totale avec la mitochondrie. Après cette première étape, nous avons recherché parmi les locus trouvés en déséquilibre, lesquels se trouvaient aussi en déséquilibre dans la population de Bassens, composée d'un mélange des deux cytotypes.

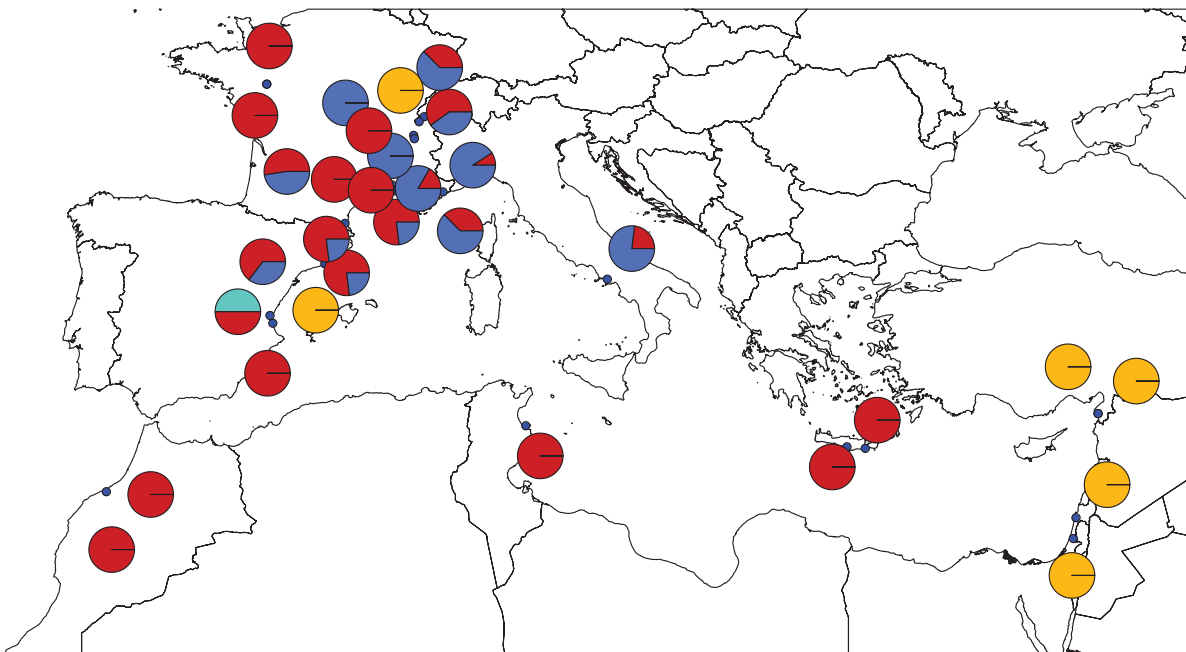
Le déséquilibre nucléo-cytoplasmique a été calculé avec le programme *CNDm* (Basten & Asmussen 1997). Pour cette analyse, seuls les locus avec au moins 36 individus ont été utilisés. Sur un total de 3 620 locus, 156 ont un test dont la valeur de  $p$  est inférieure à un pour mille entre les population allopatriques. Ce qui indique un déséquilibre entre allèles nucléaires et symbiote. Parmi ces 156 locus, 37 se trouvent aussi en déséquilibre dans la population où les deux cytotypes sont en mélange. Dans un second temps, nous avons recherché une similarité entre ces locus et les bases de données de séquences avec BLAST. Malheureusement, aucune de

ces séquences ne présente une similarité acceptable avec des séquences connues, même avec les transcriptomes connus de *B. tabaci*. Ceci n'est pas forcément étonnant puisque nos locus ne couvrent qu'une faible partie du génome de *B. tabaci* (environ 0,2 %) et ont donc une faible probabilité de se trouver dans des régions codantes. Toutefois, le génome complet de *B. tabaci* n'est pas encore disponible, le *mapping* de ces locus sur celui-ci pourrait mettre en évidence des régions fortement en déséquilibre ou la proximité de nos locus avec des gènes d'intérêts.

## Perspectives

Après ces deux essais, le génotypage de nombreuses populations a été prévu (voir la carte de la figure 4.18). Nous avons sélectionné de nombreuses populations en France où les deux cytotypes Q1 et Q2 (espèce potentielle MED) sont dans des populations mixtes, ainsi que d'autres de cette même espèce MED provenant du bassin méditerranéen, afin d'échantillonner au maximum la diversité génétique. Nous avons aussi prévu de génotyper les espèce potentielles MEAM 1 et SSA 2, les seules autres espèces que nous avons, pour traiter de questions interspécifiques : phylogénie, datation de divergence...

Pour le moment, l'ADN de 721 femelles provenant de 31 populations de l'aire méditerranéenne a été extrait. Parmi ces échantillons, 366 ont été sélectionné pour leur plus grande quantité d'ADN et une première banque a été séquencée. Les données n'ont pas encore été analysées.



**FIGURE 4.18** – Carte des populations qui vont être génotypées par *RADSeq*. Jaune : espèce potentielle MEAM 1, turquoise : espèce potentielle SSA 2, rouge : Q1 (espèce MED), bleu : Q2 (espèce MED).

# Discussion générale

Parmi les nombreux facteurs biotiques et abiotiques qui contribuent à la réussite des populations invasives, l'augmentation de la variabilité génétique est un processus maintenant largement reconnu dans la littérature. Les populations dont la variabilité génétique augmente échappent à la consanguinité et accroissent leur valeur sélective grâce aux multiples allèles présents et de leurs combinaisons potentielles, supports de l'adaptation. Bien que ce fait soit reconnu, les études présentes dans la littérature portent sur l'analyse du compartiment nucléaire, rarement sur la mitochondrie et encore moins sur les symbiotes.

Ce travail envisageait d'évaluer les rôles respectifs des compartiments nucléaire et cytoplasmique dans l'expansion démographique d'un ravageur de cultures, par ailleurs invasif à l'échelle de la planète. Nous allons reprendre les résultats décrits dans les différents chapitres et voir quels sont les éléments apportés par ce travail qui permettent une meilleure compréhension de ce phénomène.

Notre modèle d'étude est *Bemisia tabaci*, un insecte mondialement distribué dont certaines caractéristiques le rendent très intéressant dans le contexte de l'étude des introductions multiples. C'est tout d'abord un complexe d'espèces aux aires de répartition différenciées dont les limites reproductives des nombreuses entités qui le composent sont peu connues. Ce complexe d'espèces est aujourd'hui principalement appréhendé par un seuil de différenciation de 3.5 % sur le gène mitochondrial *COI*, ce qui permet de décrire 41 espèces potentielles. Bien que cette méthodologie soit pratique au quotidien pour la description du complexe, elle pose de nombreuses questions sur les barrières reproductives réelles des clades. Le deuxième aspect qui nous intéresse particulièrement ici est la richesse des associations symbiotiques secondaires que cet insecte forme avec de nombreuses bactéries, en plus de son symbiote primaire qui lui est nécessaire. Parmi les sept bactéries décrites, toutes ne sont pas retrouvées dans un même individu, sans pour autant être distribuées aléatoirement. Ces bactéries forment des combinaisons, elles-mêmes retrouvées en déséquilibre de liaison avec certains groupes haplotypiques mitochondriaux.

C'est cette dernière particularité qui fait tout l'intérêt du modèle : les nombreuses entités qui forment ce complexe se retrouvent en contact lors des épisodes d'introduction, principalement liés au commerce de plantes ornementales. Dans un premier temps nous allons discuter les résultats obtenus au cours de ce travail et voir quels sont les éléments qui soutiennent l'hypothèse de l'émergence d'un génotype très compétitif dans des conditions d'admixture telles qu'on les



rencontre en France. Dans une deuxième partie nous nous poserons la question de la présence d'interactions entre le génome de l'hôte et celui-des bactéries et nous proposerons des protocoles pour le tester.

### **Émergence d'un génotype très compétitif ?**

Dans le premier chapitre nous avons vu qu'en France la situation est particulière car deux cytotypes se trouvent en sympatrie, à la suite de l'introduction en 2007 du cytotype Q2 en France, région où le cytotype Q1 était déjà présent. Ces deux cytotypes appartiendraient à la même espèce potentielle d'après leur divergence mitochondriale qui est de l'ordre de 2 % (inférieur au seuil de référence de 3,5 %), et sont tous deux originaires du bassin méditerranéen : Q1 provient d'une large partie ouest tandis que Q2 provient d'Israël et des pays frontaliers. La littérature montrait qu'ils hébergeaient des bactéries symbiotiques différentes dans leur aires géographiques respectives. Le premier travail était donc de décrire l'évolution des compartiments nucléaire et cytoplasmique : est-ce que le fond nucléaire des deux cytotypes se mélange et est-ce que les combinaisons bactériennes changent ?

### **Les deux cytotypes s'hybrident**

Nous avons montré que, du point de vue mitochondrial, ces deux cytotypes peuvent coexister à l'échelle de la France, et ce dans de très nombreuses serres grâce à notre large échantillonnage ; en tout cas sur une durée de quelques années (durée que couvre notre échantillonnage). En effet depuis la première détection du cytotype Q2 en 2007 et jusqu'à la dernière campagne en 2014 on trouve toujours des populations mixtes dans de nombreuses serres. On ne peut bien sûr pas exclure que la coexistence soit une situation temporaire et que seules les introductions répétées permettent la détection régulière et simultanée des deux groupes. Dans nos données, il est rare qu'une même serre soit échantillonnée à plusieurs occasions (et alors les prélèvements ont au moins une année d'intervalle), soit pour des problèmes de logistique soit du fait qu'il y a peu d'aleurodes présents et donc qu'il est difficile d'en trouver. Le fait que certaines années les populations d'aleurodes soient en faible densité implique des goulots d'étranglement et donc une plus forte probabilité qu'un cytotype disparaisse par dérive. C'est un indice favorisant l'hypothèse d'introductions répétées. Pour montrer l'existence d'introductions répétées, il faudrait faire des suivis réguliers de populations en serre et les génotyper, probablement avec des marqueurs haut-débit car le signal avec les microsatellites serait trop faible. De cette façon nous pourrions suivre l'évolution du compartiment nucléaire au cours du temps et voir s'il évolue.

Lors de ces épisodes de coexistence nous avons vu, dans quatre serres différentes en France et en Espagne, grâce aux marqueurs microsatellites, que ces deux cytotypes peuvent s'hybrider. Dans le dernier chapitre, traitant des marqueurs *RADSeq*, nous avons trouvé à nouveau des preuves de l'hybridation dans une serre à Bassens en Savoie. Ces résultats renforcent ceux obtenus avec les

microsatellites puisque ceux-ci étaient « seulement » au nombre de sept et auraient pu être peu représentatifs du génome, contrairement aux marqueurs *RADSeq* qui sont au nombre de plusieurs milliers.

Cette hybridation entre les deux cytotypes est corroborée par des essais de croisements au laboratoire qui montrent l'interfécondité des deux groupes. Ces essais n'ont pas été conduits sur les mêmes populations que celles génotypées en serre mais sur des lignées de laboratoire établies depuis plusieurs années, qui n'avaient pas eu l'occasion de s'hybrider auparavant, renforçant la conclusion que ces cytotypes peuvent s'hybrider.

Puisque les deux cytotypes sont interféconds nous sommes dans une situation typique d'admixture où des groupes préalablement isolés se trouvent en contact secondaire. Nous avons donc testé si cette situation favorisait l'augmentation de la variabilité génétique dans l'aire d'introduction en France. Bien que les valeurs de richesse allélique soient supérieures en France et en Espagne par rapport à la Grèce et à Israël, ces valeurs ne sont pas significatives. De plus, nous n'avons pas pu rattacher le fond nucléaire des individus Q2 d'Israël à ceux des populations de l'aire d'introduction en France. Il faut toutefois noter plusieurs choses : ces deux aires de référence en Grèce et en Israël ne sont pas forcément les aires d'origine des populations introduites en France, mais seulement des zones où l'on trouve seulement l'un des cytotypes. De plus, notre échantillonnage est très faible dans ces zones supposées allopatriques, seulement une population en Israël et trois en Grèce, bien que pour ce dernier pays la situation soit bien moins critique en terme de capacité d'échantillonnage. C'est pourquoi nous n'avons peut-être pas génotypé tous les fonds nucléaires possibles pour les individus introduits en Europe de l'ouest. Ainsi, dans le cas de populations structurées on peut très bien passer à côté des populations à l'origine des introductions en Europe de l'ouest et la variabilité détectée n'être qu'un faible échantillon de la variabilité potentielle de l'aire géographique. Par exemple on sait que les populations françaises de Q1 en serre sont structurées, même si cette différenciation est faible (Dalmon *et al.* 2008).

Si dans le futur on voulait améliorer l'échantillonnage des cytotypes Q2 dits de référence dans l'est du bassin méditerranéen, cela s'avérerait extrêmement difficile pour plusieurs raisons. Tout d'abord, il semblerait que la densité de populations Q2 en Israël soit en baisse au profit du clade MEAM 1, comme nous l'avons vu dans le chapitre 1. La deuxième raison est que l'aire de répartition historique de Q2 (Israël, Chypre, Liban...) n'est pas forcément la région la plus facile d'accès, et encore moins pour la Syrie par exemple qui est limitrophe mais où les données sont quasiment absentes, et totalement depuis plusieurs années...

Enfin, il est possible que nos marqueurs microsatellites ne soient pas totalement appropriés car nous en avons seulement sept. À l'avenir, avec des marqueurs de génotypage haut débit, il se pourrait que ces conclusions soient différentes lorsque l'on prend en compte l'ensemble du génome.

Pour conclure sur cette partie, nous avons montré que les différents cytotypes s'hybrident en situation de sympatrie, mais nous n'avons pas pu mettre en évidence une augmentation de la variabilité génétique nucléaire.

## Quid des bactéries ?

Nous avons montré expérimentalement que la bactérie *Rickettsia* peut être transférée horizontalement d'un individu à l'autre *via* la plante, indistinctement depuis le cytotype Q1 infecté par *Rickettsia* et *Hamiltonella* ou le cytotype Q2 infecté par *Rickettsia*, *Arsenophonus* et *Wolbachia*. Ce résultat est corroboré par les données de la littérature qui montrent que les bactéries peuvent transférer d'un individu à l'autre, avec par exemple des preuves « phylogénétiques » lorsqu'il y a incongruence entre l'arbre de l'hôte et celui de la bactérie, où alors des preuves expérimentales directes. Plus particulièrement, chez *B. tabaci* — clade MEAM1 — la bactérie *Rickettsia* peut transférer d'un individu à l'autre par l'intermédiaire de la plante hôte (Caspi-Fluger *et al.* 2012).

Il est donc possible que dans notre situation de coexistence entre les cytotypes Q1 et Q2, certaines bactéries puissent transférer d'un individu à l'autre. Ce transfert pourrait permettre l'acquisition des traits phénotypiques apportées par la bactérie (négligeons pour le moment les interactions possibles entre les différentes bactéries, qui peuvent être négatives). Dans nos données, les centaines d'individus génotypés (dans le premier chapitre) montrent tous le même patron : Q1 est associé à *Cardinium* et *Hamiltonella* tandis que Q2 est associé à *Rickettsia* et *Arsenophonus* ; *Wolbachia* étant potentiellement associée à Q1 ou Q2. Ces données sont les mêmes que celles de la littérature, qui montrent une extrême stabilité des associations rencontrées (Chiel *et al.* 2007 ; Chu *et al.* 2007 ; Gueguen *et al.* 2010) ce qui semble indiquer que les recombinaisons bactériennes sont soit absentes soit rares (au mieux les associations symbiotiques chez *B. tabaci* sont étudiées depuis 15 ans !).

Il était donc surprenant, au premier abord, de trouver la bactérie *Rickettsia* associée à l'haplogroupe Q1 en 2012 simultanément dans deux serres. La détection simultanée s'explique probablement par le fait que les deux serres sont approvisionnées par le même fournisseur et les deux populations ne sont en fait le reflet que d'une seule. En effet cette hypothèse paraît plus parcimonieuse (un seul événement de transfert) que deux événements de transfert simultanés, transferts qui n'avaient encore jamais été détectés. La présence de cette bactérie provient vraisemblablement d'un transfert horizontal. Bien que nos analyses MLST n'ont pas pu identifier formellement la source, nous suspectons que Q2 soit l'insecte émetteur car on le retrouve dans les mêmes populations que les individus Q1 qui ont le cortège atypique nouvellement trouvé : Q1 associé à *Hamiltonella* et *Rickettsia*. Cette nouvelle association haplotype mitochondrial-bactéries-fond nucléaire est toujours présente dans cette serre deux ans après, dans une population dépourvue de tout autre cytotype ou autre association bactérienne. Cette situation où l'on retrouve qu'un seul cytotype est nouvelle dans cette serre car deux ans auparavant lors de la découverte de la nouvelle association, cette dernière n'était pas seule dans la serre puisqu'il y avait des individus Q1 classique ainsi que des individus Q2. On peut se poser la question des phénomènes qui ont mené à la disparition de Q1CH dans cette serre, au profit du Q1RH.

Tout d'abord, les données nucléaires (voir chapitre 2 sur les microsatellites) semblent mon-

trer une hybridation asymétrique, c'est-à-dire une pénétration du matériel nucléaire Q1 dans le fond cytoplasmique Q2. Ces résultats sont en accord avec d'autres de la littérature (Gauthier *et al.* 2014). Ce remplacement du fond nucléaire Q2 par celui de Q1 peut être expliqué par des bactéries manipulant la reproduction mais nous avons vu aucune n'est fixée. Bien que nous n'ayons pas la preuve que ce soit cette bactérie qui favorise la valeur sélective de Q2, d'autres auteurs prennent moins de précautions et impliquent *Rickettsia* dans l'invasion de Q2 en Italie (Parrella *et al.* 2013). Ce compartiment cytoplasmique contenant *Rickettsia* semble être sélectionné.

Par ailleurs, la littérature commence à s'étoffer quant aux effets phénotypiques de cette bactérie. Aux États-Unis, il a été démontré (Himler *et al.* 2011) qu'elle biaise le sexe-ratio en faveur des femelles tout en accroissant la fécondité et la survie, le tout dans un contexte d'expansion de l'infection à l'échelle de l'Arizona (ce qui fait tout de même 300 000 km<sup>2</sup>). Par contre ce phénomène est observé pour des populations de l'espèce potentielle MEAM1 aux États-Unis mais pas sur des individus MEAM1 provenant d'autres populations (Cass *et al.* 2015b), et pas non plus chez l'espèce potentielle MED. Il faut donc rester prudent, l'effet induit chez l'espèce potentielle MED est peut être très différent. D'autres données de la littérature montrent son rôle protecteur en cas de choc thermique (Brumin *et al.* 2011), bien que dans nos données expérimentales nous n'ayons pu mettre ce fait en évidence.

Ces données pourraient expliquer pourquoi le compartiment cytoplasmique contenant *Rickettsia* serait sélectionné. Mais dans cette même population de notre étude qui contenait des individus Q1RH et Q1CH il y avait des individus Q2 qui portent aussi *Rickettsia*. La question se pose de la disparition de ce cytotype Q2. La réponse pourrait venir de l'autre bactérie hébergée par le Q1 : *Hamiltonella*, qui conférerait un fort avantage sélectif. En effet, cette bactérie joue un rôle métabolique et serait indispensable à son hôte (Rao *et al.* 2015) car elle partage certaines voies de synthèse d'acides aminés essentiels avec *Portiera*, le symbiote obligatoire avec qui elle partage les bactériocytes (Gottlieb *et al.* 2008 ; Skaljic *et al.* 2010). Des données expérimentales montrent qu'*Hamiltonella* donnerait à son hôte une meilleure résistance en cas de stress nutritionnel (Su *et al.* 2014b).

Nous avons vu dans le chapitre 3 que *Rickettsia* pouvait transférer horizontalement, on pourrait donc se demander pourquoi *Hamiltonella* n'est pas transférée dans le fond cytoplasmique de Q2, ce qui donnerait à nouveau une combinaison comprenant *Rickettsia* et *Hamiltonella*. La première chose est que le fond cytoplasmique Q2 comprend aussi *Arsenophonus* qui bien qu'elle ne soit pas fixée pourrait compliquer le transfert d'*Hamiltonella* dans le Q2. Un deuxième indice vient des données génomiques qui montrent qu'*Hamiltonella* a perdu ses facteurs de virulence et donc sa capacité à transférer d'un hôte à l'autre (Rollat-Farnier *et al.* 2015). Cette perte de capacité à transférer est illustrée par la gamme d'hôte possible parmi les espèces potentielles du clade *B. tabaci* puisqu'on ne retrouve cette bactérie que dans les espèces MED, INDIAN OCEAN, EA1 et MEAM1 (ce qui est déjà beaucoup) mais surtout, ces espèces forment un grand groupe monophylétique, qui comprend les deux espèces les plus invasives (MED et MEAM1). Ainsi il paraît très peu probable qu'*Hamiltonella* puisse changer de fond cytoplasmique facilement et

*in fine* la capacité de *Rickettsia* à transférer vers un fond cytoplasmique contenant déjà *Hamiltonella* serait le moyen de « créer » cette combinaison.

Ainsi on pourrait donc se demander si cette nouvelle association (Q1-*Rickettsia*-*Hamiltonella*) ne serait pas une combinaison idéale rassemblant les effets bénéfiques nutritionnels d'*Hamiltonella* et ceux potentiels de *Rickettsia*, bien que ces derniers restent à éclaircir.

## Vers une génomique nucléo-cytoplasmique ?

Dans la partie précédente nous avons considéré les génomes nucléaires et les bactéries symbiotiques comme de simples réservoirs de gènes ou d'allèles et qui par le biais de la sélection vont être sélectionnés comme des entités indépendantes dans un individu constitué des meilleures éléments possibles. Bien que cette dernière phrase soit caricaturale, il est faux de considérer tous ces éléments comme indépendants car ils interagissent fortement. Nous allons voir quels sont les éléments en faveur de telles interactions, et quels sont les moyens envisageables pour décrire plus précisément ces interactions.

## Détection de covariance nucléo-cytoplasmique

### En populations naturelles

Le fait que les effets phénotypiques induits par les bactéries soient nucléaires-dépendant implique de l'épistasie entre compartiment nucléaire et cytoplasmique. De tels liens ont été démontrés entre la mitochondrie et le fond nucléaire avec certains modèles biologiques, mais rarement entre bactérie et allèles nucléaires. Par exemple Beck *et al.* (2015) ont montré que lors de l'hybridation entre *Drosophila yakuba* et *Drosophila santomea*, il y avait cointrogression d'allèles mitochondriaux et nucléaires, tous impliqués dans la phosphorylation oxydative, ce qui montre une coadaptation des différentes protéines impliquées.

Pour étudier l'épistasie entre gènes d'un même génome, le paramètre  $\Theta$  (Cockerham & Weir 1977) est utilisé, il mesure le degré de cotransmission de combinaisons de gènes. En effet, la descendance qui ne porte pas les combinaisons d'allèles qui interagissent positivement (épistasie positive), aura une valeur sélective inférieure à celle les portant. La fréquence des combinaisons positives augmente donc dans la population. Quand  $\Theta$  est grand, les deux gènes ne recombinent pas, comme deux gènes mitochondriaux par exemple, au contraire d'un  $\Theta$  petit qui indique que les deux gènes recombinent totalement, comme par exemple deux gènes indépendants dans une population panmictique. Le concept peut être élargi à l'étude de l'épistasie entre différents génomes (Wade & Goodnight 2006 ; Wade 2007). Ainsi les grandes valeurs de  $\Theta$  permettent à la sélection d'agir sur des combinaisons de gènes à travers le génome. La transmission verticale augmente  $\Theta$  mais ce phénomène est contrecarré par la reproduction sexuée qui casse les combinaisons. Pour qu'il y ait coadaptation il faut donc une histoire évolutive commune



assez longue entre la bactérie et l'hôte, ce qui implique que le symbiote transfère peu car cela casserait aussi les associations (Fitzpatrick 2014).

Dans le complexe *B. tabaci* on sait que certaines bactéries transfèrent d'une espèce à l'autre : *Arsenophonus* (Thao & Baumann 2004b ; Mouton *et al.* 2012 ; Ahmed *et al.* 2013) ; *Cardinium* et *Wolbachia* (Ahmed *et al.* 2013). Par contre, ces données ne nous donnent pas d'information sur la durée des relations entre les deux partenaires. Mais, comme les données de la littérature montrent que les effets phénotypiques des bactéries dépendent du fond nucléaire, cela suggère très fortement la présence de coadaptation et donc que la durée est assez longue pour que de tels événements puissent émerger.

Dans le chapitre qui traite de la mise au point des marqueurs *RADSeq*, nous avons cherché à détecter la présence de coadaptation en utilisant comme proxy le déséquilibre de liaison entre fond cytoplasmique et allèles nucléaires. Nous avons trouvé de nombreux allèles en déséquilibre avec certains fonds cytoplasmiques (une partie est probablement des faux positifs), mais nous ne pouvons pas rattacher les séquences de ces locus à des gènes ou des fonctions. Malgré la présence de plusieurs transcriptomes de *B. tabaci* dans les bases de données nucléotidiques, aucune similarité n'a pu être détectée. Il est très probable que la densité en marqueur ne soit pas assez élevée et que la probabilité que les locus en déséquilibre se trouvent dans un gène est faible sachant qu'avec cette enzyme de restriction nous n'avons probablement séquencé que 0,02 % du génome. L'apport d'un génome de référence pour *B. tabaci* serait un plus. En effet, l'évolution des valeurs de déséquilibre associées aux locus le long du génome permettrait de détecter des régions d'intérêt. En effet, si un gène est sélectionné, les régions adjacentes le sont avec lui (*selective sweep*). Un bémol à ajouter à ces résultats est que la covariance entre le génome de l'hôte et les bactéries dépend en plus fortement de l'histoire des populations, la structure géographique ou encore la sélection, de manière analogue à des gènes au sein d'un même génome (Fitzpatrick 2014). De plus, l'hybridation en elle-même crée artificiellement du déséquilibre cytonucléaire, qui ne sera érodé qu'à plus long terme (Asmussen & Arnold 1991 ; Arnold 1993). Il faut donc rester prudent dans le cas de situations naturelles où la date de l'évènement d'hybridation n'est pas connue et s'assurer que nous ne sommes pas en présence de *F1* ou *F2* (par exemple).

Nous pourrions utiliser certains outils bayésiens pour analyser ces données afin de détecter des locus sous sélection en populations naturelles, tout en s'affranchissant des effets démographiques. Justement, Foll *et al.* (2014) ont publié un outil qui permet de détecter de la sélection dans des populations en structure hiérarchique. Dans leur étude, ils ont analysé un grand jeu de données (SNP) sur des populations humaines de basse et haute altitude en Amérique et en Asie. Ils ont ainsi pu identifier plusieurs régions du génome qui sont sous sélection dans les deux régions, de façon convergente. Ici nous pourrions définir le fond cytoplasmique (les cytotypes Q1 et Q2 respectivement) comme niveau de structure le plus bas et ainsi trouver des locus sous sélection entre les deux fonds cytoplasmiques, mesure que nous pourrions coupler à celles de déséquilibre nucléo-cytoplasmique.

### Un protocole expérimental pour tester la covariance entre génomes

Dans un deuxième temps nous pourrions imaginer des expérimentations à même de tester la présence de coadaptation entre les symbiotes et les gènes de l'hôte. Ces essais associeraient des croisements, de l'évolution expérimentale et des outils de génotypage haut-débit. Tout d'abord il faudrait établir des lignées stables avec une infection donnée, dans l'idéal avec comme seul symbiote celui à tester (en plus du primaire), dans notre cas les symbiotes les plus intéressants seraient *Hamiltonella* et *Rickettsia*.

Une fois les lignées établies, on effectue un croisement puis un backcross. Ainsi les mâles de la F<sub>2</sub>, haploïdes (comme à toutes les générations bien sûr !) seraient génotypés par *RADSeq* ou toute autre méthode de génotypage par réduction du génome qui se révèle reproductible et abordable ! Si il y a coadaptation, certains haplotypes seraient plus rares que d'autres, et on pourrait ainsi localiser les régions du génome impliquées. Malheureusement, dans le chapitre sur le développement des marqueurs *RADSeq* nous avons vu qu'il était pour le moment difficile de génotyper des mâles pour des raisons de quantité d'ADN qui serait pourtant un meilleur outil que les femelles.

Une deuxième possibilité serait de l'évolution expérimentale en cage à populations. On injecte dans une cage des mâles de l'une des lignées et des femelles de l'autre, puis on attend plusieurs générations, afin de casser les associations entre cytoplasme et génome nucléaire de chaque lignée. Bien sûr, s'il y a coadaptation entre les deux génomes des individus, certaines associations auraient une valeur sélective plus élevée et donc retrouvées en plus forte fréquence. Les lignées seraient alors génotypées afin de déterminer quelles associations nucléo-cytoplasmiques se trouvent en excès, ou inversement, celles qui sont en faible fréquence. Plusieurs répétitions seraient nécessaires pour s'affranchir des effets statistiques et surtout de dérive qui pourraient fixer certaines combinaisons par simple hasard.

## Conclusion

*B. tabaci* est un ravageur largement étudié car il induit de sérieux problèmes à diverses cultures, dans presque tous les pays du monde. C'est probablement ces aspects négatifs qui en ont fait un modèle d'étude particulièrement étudié pour son écologie, dans le cadre de travaux plutôt appliqués. Entre temps des thématiques nouvelles ont émergé, sur des aspects plus fondamentaux comme les associations symbiotiques, l'hybridation, la phylogénie...

Aujourd'hui, des perspectives nouvelles s'ouvrent avec l'apparition des nouvelles technologies de séquençage haut-débit. Peut-être que nous allons pouvoir mettre en évidence les régions génomiques sous-jacentes à l'extraordinaire réussite de ce ravageur et mettre en évidence les processus évolutifs qui ont conduit à la radiation de ce complexe.



# Travaux publiés

Article 1 : Pertinence biologique des cytotypes	171
Article 2 : Distribution des cytotypes en Afrique de l'ouest	183
Article 3 : Caractérisation des haplotypes mitochondriaux par PCR-RFLP	195
Article 4 : Détection d'un cytotype génétiquement isolé en Afrique de l'ouest	205
Article 5 : Optimisation du multiplexage <i>RADSeq</i>	213



# Nuclear and cytoplasmic differentiation among Mediterranean populations of *Bemisia tabaci*: testing the biological relevance of cytotypes

Gabriel Terraz,<sup>a\*</sup> Gwenaëlle Gueguen,<sup>a</sup> Judit Arnó,<sup>b</sup> Frédéric Fleury<sup>a</sup> and Laurence Mouton<sup>a</sup>

## Abstract

**BACKGROUND:** The taxonomy of the species complex *Bemisia tabaci* is still an unresolved issue. Recently, phylogenetic analysis based on *mtCOI* identified 31 cryptic species. However, mitochondrial diversity is observed within these species, associated with distinct symbiotic bacterial communities forming associations, which here are called cytotypes. The authors investigated the biological significance of two cytotypes (Q1 and Q2) belonging to the Mediterranean species, which have only been found in allopatry in the Western Mediterranean to date. Sampling was done over a few years in Western Europe, and sympatric situations were found that allowed their reproductive compatibility to be tested in the field with the use of microsatellites.

**RESULTS:** The field survey indicated that, in spite of its recent introduction, Q2 is well established in France and Spain, where it coexists with Q1. Microsatellite data showed that, in allopatry, Q1 and Q2 are highly differentiated, while there is little or no genetic differentiation when they coexist in sympatry, suggesting a high rate of hybridisation. Crossing experiments in the lab confirmed their interfertility.

**CONCLUSION:** Q1 and Q2 hybridise, which confirms that they belong to the same species, in spite of the high degree of genetic differentiation at both the cytoplasmic and nuclear levels, and also suggests that their symbiotic bacteria do not prevent hybridisation.

© 2014 Society of Chemical Industry

**Keywords:** *Bemisia tabaci*; MED species; hybridisation; microsatellites; insecticide resistance gene; bacterial endosymbionts

## 1 INTRODUCTION

*Bemisia tabaci* is one of the most devastating pests in the world, both directly through plant sap feeding and indirectly by virus transmission.<sup>1</sup> Historically, based on significant variations between populations, biotypes have been identified using ecological traits such as host-plant range, virus transmission abilities, insecticide resistance, enzymatic variability and genetic markers such as the mitochondrial cytochrome oxidase gene (*mtCOI*).<sup>2–4</sup> Because some of these characters are shared between genetic groups or show within-group variability that is higher than between-group variations, the term biotype often appears to be inappropriate.<sup>5</sup> Based on a 3.5% *mtDNA* sequence divergence criterion, it has been proposed that *B. tabaci* is actually made up of at least 31 morphologically indistinguishable species.<sup>5–10</sup> Belonging to 11 major groups, these putative species regroup the previously described biotypes. This view is consistent with some genetic data accumulated on reproductive incompatibilities within the *B. tabaci* species complex.<sup>7,8</sup> Besides the number of species in the *B. tabaci* complex, several questions remain, such as the relevance of the *mtDNA* sequence divergence threshold used for species boundary identification (3.5%). It has been suggested that this value is probably underestimated. For example, Lee *et al.*<sup>9</sup> proposed that this threshold be raised to 4% because there is much variability within species, particularly within the Mediterranean (MED) species. The

MED species appears to be genetically diverse, not only because it includes the J, ASL, L and highly damaging Q biotypes<sup>5</sup> but also because high genetic variability has been observed within the Q biotype, leading to the recognition of 4 *mtCOI* haplotypes (Q1 to Q4).<sup>10,11</sup> This variability may regroup separate entities with high genetic differentiation and possible different species, challenging MED species status. Therefore, more information is clearly needed to determine what the true interbreeding limits of the putative species are within the *B. tabaci* complex, and whether reproductive incompatibilities may occur within the putative species that challenge the 3.5% *mtDNA* sequence divergence as a criterion for species delimitation.

In addition to the uncertainty of the *mtDNA* divergence threshold proposed for species recognition, another challenging point concerns the mitochondrial DNA marker, which is probably not the best marker to identify species limits. As in a number of other

\* Correspondence to: Gabriel Terraz, Laboratoire de Biométrie et Biologie Evolutive, UMR 5558 CNRS, Université Claude Bernard Lyon 1, Université de Lyon, Villeurbanne, France. E-mail: gabriel.terraz@univ-lyon1.fr

a Laboratoire de Biométrie et Biologie Evolutive, Université Claude Bernard Lyon 1, Université de Lyon, Villeurbanne, France

b Departament de Protecció Vegetal, IRTA, Centre de Cabriels, Cabriels, Spain

arthropods, *mtDNA* undergoes indirect selection owing to vertically transmitted microorganisms invading host populations.<sup>12</sup> Linkage disequilibrium between *mtDNA* haplotypes and bacterial endosymbionts may blur the clade history and rather reveal the history of infection. Moreover, horizontal transfers of bacterial endosymbionts followed by their invasion in the host population can result in the rapid spread of *mtDNA* lineages without any reproductive barrier.<sup>13,14</sup> This is achieved by their ability to manipulate host reproduction in different ways, especially by inducing cytoplasmic incompatibility (for a review, see Engländer and Hurst<sup>15</sup>). In addition to their ability to manipulate host reproduction, symbiotic bacteria may also play an important role in host ecology, host evolution and speciation (see the review in Ferrari and Vavre<sup>16</sup>). Symbiotic bacteria are involved in resistance to natural enemies,<sup>17</sup> resistance to insecticides and host-plant specialisation.<sup>18,19</sup> They can also increase their host fitness under certain environmental conditions.<sup>20</sup> Moreover, they can be involved in resource exploitation and specificity, which confer ecological isolation, leading to speciation.<sup>21</sup>

*B. tabaci*, like most phloem-feeding insects, harbours an obligatory maternally transmitted endosymbiont, *Portiera aleyrodidarum*, which provides necessary essential nutrients.<sup>22,23</sup> *B. tabaci* also hosts a wide diversity of facultative bacteria, all of which are maternally transmitted. To date, seven secondary endosymbionts have been reported, which form specific association in multi-infected individuals.<sup>24,25</sup> These symbiotic communities appear to be species-specific but also show variations within putative species.<sup>10,26–28</sup> For example, the four haplotypes identified within the MED species harbour distinct bacterial communities.<sup>10,26,27</sup> Q1 harbours *Hamiltonella* and sometimes *Cardinium*, whereas Q2 is infected by *Rickettsia* and *Arsenophonus*. Both Q1 and Q2 are infected by *Wolbachia*<sup>10</sup> but belong to different lineages. Q3 is free of *Wolbachia* but hosts *Arsenophonus* and *Rickettsia*. Q4 is a rare haplotype that has only been detected in Croatia, for which bacterial infection remains unknown. In the present paper, the term 'cytotype' is used to designate an identical *mtCOI* haplotype associated with a specific symbiotic bacterial community. As there is variability in infection frequencies both within and among populations, none of the individuals belonging to one cytotype harbours all of the members of this symbiotic community but rather a combination of symbionts. These cytotypes show different geographic localisations, as Q3 has been recorded only in Africa, whereas Q1 is localised in a large western part of the Mediterranean basin: Morocco, Tunisia, Portugal, France, Greece and Spain.<sup>10,29–31</sup> All data available on the *B. tabaci* MED species in France and Spain indicate the mere presence of the Q1 cytotype.<sup>29,32–35</sup> This cytotype has been found in several parts of the world, in China and America for example, while Q2 is restricted to the Middle East countries of the Mediterranean basin, Israel and Cyprus, where Q1 is absent.<sup>27,36,37</sup> Q1 and Q2 cytotypes overlap in neighbouring countries such as Turkey, Egypt and Syria (GenBank records).<sup>6,33</sup> As these two cytotypes harbour very different bacterial communities,<sup>10</sup> the MED species provides a good framework to test and to understand the role of these communities in the complex ecology and evolution of the *B. tabaci* species. This particular situation also provides a good opportunity to test whether the 3.5% *mtCOI* threshold reflects the true limit of species within the *B. tabaci* species complex in which bacterial endosymbiont infection is pervasive. This is particularly important from pest management perspectives.

In the present study, the authors report the presence of the Q2 cytotype in Western Europe in sympatry with Q1, which was

the only cytotype previously known in this area. The dynamics of colonisation of the Q2 cytotype in France and Spain, which coexists locally with the Q1 cytotype, was followed. These sympatric situations provided the opportunity to test for possible hybridisation or the presence of a reproductive barrier among Q1 and Q2 cytotypes via the use of microsatellite markers (expected to be neutral). Because of the extensive use of insecticides to control *B. tabaci* populations,<sup>38</sup> two genes that are known to be involved in insecticide resistance against pyrethroids and organophosphates were also studied. Resistant alleles and their possible different frequencies in both cytotypes may explain patterns of cytotype distribution. They also could drive the outcome of competition in the case of reproductive isolation.

## 2 MATERIALS AND METHODS

### 2.1 Insect lines and rearing

Two lines belonging to Q1 and Q2 were used for crossing experiments. The Q1 line, originating from Sigeac (Languedoc Roussillon, France), was infected by *Hamiltonella* and *Cardinium*. The Q2 line, originating from Lyon (Rhône-Alpes, France), was infected by *Rickettsia* and *Arsenophonus*. Both lines were fixed for bacterial infection. These lines were maintained in the lab for at least 1 year (about 17 generations) before the experiments were initiated. They were reared on *Hibiscus moscheutos* var. *Galaxy* plants at 25 °C with an L:D cycle of 16:8 and 70% relative humidity. Their cytotypes (*mtCOI* and associated symbionts) were regularly verified by PCR.

### 2.2 Crossing experiments

Pupae from the Q1 and Q2 lines were isolated, and adults were sexed from 1 to 4 h after emergence to ensure female virginity. All four possible crosses were performed (crosses between Q1 males and females, crosses between Q2 males and females, crosses between Q1 males and Q2 females and reciprocal crosses), with ten replicates each. Five females and males were placed on 30 mm diameter excised *Hibiscus moscheutos* var. *Galaxy* leaf discs on 1% agar medium in a 50 mm petri dish for 6 days. Laid eggs were counted 3 days after removing the females, as were emerging adults. In order to test for the fertility of F1 individuals, offspring were kept together, which allowed them to mate and produce progeny. Thus, the presence of both males and females in the progeny was checked.

### 2.3 *B. tabaci* field sampling

A total of 531 females from ten localities were collected in France and Spain from 2006 to 2010 for a temporal survey. Adults were collected in greenhouses on different hosts, including ornamental plants and vegetables (Table 1). Sampled adults were placed alive in ethanol (96%) and stored at –20 °C until DNA extraction. In order to have Q1 and Q2 'reference' cytotypes, use was made of samples from Greece and Israel, where only one Q cytotype had been reported to date: Q1 in Greece and Q2 in Israel. This was confirmed by the recent study of Gauthier and colleagues.<sup>39</sup>

### 2.4 Molecular assays

#### 2.4.1 DNA extraction

For each individual, total DNA was extracted in 26 µL of a buffer containing 50 mM of KCl, 10 mM of Tris-base pH 8, 0.45% Nonidet P-40, 0.45% Tween 20 and 50 mg mL<sup>-1</sup> of Proteinase K. After 3 h at 65 °C, samples were incubated at 100 °C for 15 min. A volume of 35

**Table 1.** Characteristics of the sample site locations

Country	Region	Locality	Acronym <sup>a, b</sup>	Year	Host plant	N <sup>c</sup>	
France	Midi-Pyrénées	Lavaur	LavO	2007	Ornamental	24	
		Castelmaurou	CastM	2007	Manoelila	15	
		Launaguet	LaunH	2007	Hibiscus	23	
	Languedoc	Sigean	SigT	2006	Tomato	20	
		Provence Alpes Cote d'Azur	St Martin de Crau	CrauP	2007	Sweet pepper	97
	Saint Laurent du Var		SlvE*	2010	Eggplant	45	
	Fréjus		FroH*	2010	Hibiscus	48	
	Rhône-Alpes		Lyon	LyoV	2007	Various plants	19
			Catalonia	Cabriels	CabV	2007	Various plants
	CabM*	2010			Melon	31	
Viladecans	CabT*	2010		Tomato	40		
	VilT*	2010		Tomato	31		
	VilC*	2010		Cucumber	47		
	Greece	Thessaloniki		TskC*	2010	Cucumber	10
Kasteli-Chania		KscE*	2010	Eggplant	10		
Ierapetra		IerC*	2010	Cucumber	10		
Israel		Hof Carmel	HfcV*	2010	Various	29	

<sup>a</sup> The first three or four letters of the acronyms correspond to the locality; the last capital letter indicates the host plant.

<sup>b</sup> \*: Populations used for microsatellite analysis (all the individuals of these populations have been genotyped) and insecticide-resistant allele detection (the number of individuals studied per population is specified in Table 3).

<sup>c</sup> N: the number of sampled individuals genotyped for mitochondrial haplotype and screened for the presence of endosymbionts.

µL of pure water was added to this extract, which was then stored at  $-20^{\circ}\text{C}$  until use.

#### 2.4.2 mtCOI haplotype determination

For all of the 531 individuals collected, mitochondrial haplotypes were determined using a PCR-RFLP diagnostics tool as described in Henri *et al.*,<sup>40</sup> which allowed the discrimination of all of the mitochondrial haplotypes described within the MED species. Some of the PCR products (2–10 individuals per population) were also sequenced by Biofidal (Lyon, France) for confirmation (GenBank accession numbers: KJ411776–KJ411814).

#### 2.4.3 Detection of endosymbiotic bacteria

The six most frequent secondary symbiotic bacteria found in *B. tabaci* (*Wolbachia*, *Cardinium*, *Hamiltonella*, *Rickettsia*, *Arsenophonus* and *Fritschea*) were screened using specific PCR

primers for each of the 531 individuals (Table 2). One simplex PCR was realised for each symbiont. All samples were tested for the presence of the obligatory symbiont *Portiera aleyrodidarum* to test for the quality of the DNA extraction. PCR reactions were performed in 25 µL volumes containing 200 µM of dNTP, 200 nM of primers, 0.5 IU of Taq DNA polymerase (Eurobio) and 2 µL of DNA template. PCR products were visualised in 1% agarose gels stained with Gelred<sup>TM</sup> (5 µL 100 mL<sup>-1</sup>) under UV illumination.

#### 2.4.4 Insecticide-resistant allele detection

Resistant (r) and susceptible (S) alleles were identified in two genes associated with insecticide resistance in *B. tabaci*: the *para*-type voltage-gated sodium channel and the *ace1* genes. Two mutations at two different sites in the *para*-type voltage-gated sodium channel gene, L925I and T929V (r1 and r2 alleles), were screened for,

**Table 2.** Characteristics of primers used for endosymbiont screening

Endosymbiont	Targeted gene	Primers	Primer sequence	T <sub>m</sub> (°C)	Reference
<i>Portiera aleyrodidarum</i>	<i>rDNA 16S</i>	28F 1098R	5'-TGCAAGTCGAGCGGCATCAT-3' 5'-AAAGTCCCCGCCTTATGCGT-3'	58	24
<i>Rickettsia</i>	<i>rDNA 16S</i>	Rb-F Rb-R	5'-GCTCAGAACGAACGCTATC-3' 5'-GAAGGAAAGCATCTCTGC-3'	58	73
<i>Hamiltonella</i>	<i>rDNA 16S</i>	Hb-F Hb-R	5'-TGAGTAAAGTCTGGGAATCTGG-3' 5'-AGTTCAAGACCGCAACCTC-3'	58	24
<i>Cardinium</i>	<i>rDNA 16S</i>	CFB-F CFB-R	5'-GCGGTGTAAAATGAGCGTG-3' 5'-ACCTMTTCTTAAGCAAGCCT-3'	56	74
<i>Wolbachia</i>	<i>wsp</i>	81F 691R	5'-TGGTCCAATAAGTGATGAAGAAAC-3' 5'-AAAAATTAACGCTACTCCA-3'	56	75
<i>Arsenophonus</i>	<i>rDNA 23S</i>	Ars-23S1 Ars-23S2	5'-CGTTTGATGAATTCATAGTCAA-3' 5'-GGTCTCCAGTTAGTGTACCCAAC-3'	60	76
<i>Fritschea</i>	<i>rDNA 23S</i>	Frit-F Frit-R	5'-GAGTTTGATCATGGCTCAGATTG-3' 5'-GCTCGCTACCACTTTAAATGGCG-3'	62	77

**Table 3.** Frequencies of sodium channel and *ace-1* resistant mutations in *B. tabaci* per host plant, cytotype and locality

Country	Locality	Cytotype	Sodium channel										<i>ace</i>				
			Locus 1 (L925I) genotypes				r1 allele resistance frequencies	Locus 2 (T929V) genotypes				r2 allele resistance frequencies	Genotypes			R allele resistance frequencies	
			n	r1r1	r1s1	s1s1		n	r2r2	r2s2	s2s2		n	RR	RS		SS
Greece	Variou	Q1	29	16	11	2	0.74	29	0	1	28	0.02	29	29	0	0	1
Israel	Hof-Carmel	Q2	23	12	8	3	0.70	27	3	11	13	0.29	27	27	0	0	1
Spain	Cabrilis	Q1	7	3	4	0	0.71	7	0	3	4	0.21	7	7	0	0	1
		Q2	25	6	16	3	0.56	25	0	15	10	0.30	25	25	0	0	1
	Viladecans	Q1	28	12	14	2	0.68	28	0	6	22	0.11	28	28	0	0	1
		Q2	9	6	2	1	0.78	9	0	3	6	0.17	9	9	0	0	1
France	Saint-Laurent du Var	Q1	12	11	1	0	0.96	12	0	0	12	0	12	12	0	0	1
		Q2	7	6	1	0	0.93	7	0	0	7	0	7	7	0	0	1
	Frejus	Q1	8	8	0	0	1	8	0	0	8	0	8	8	0	0	1
		Q2	15	12	3	0	0.90	15	0	2	13	0.07	15	15	0	0	1

along with one mutation in the acetylcholinesterase enzyme *ace1* (F331W, R allele). These mutations confer resistance to pyrethroids (Pyr) and organophosphates (OPs) respectively. L925I and F331W resistance mutations were identified using the PCR-RFLP tool developed by Tsagkarakou *et al.*<sup>41</sup> The T929V resistance mutation was detected by PCR amplification of specific alleles.<sup>41</sup> Screening was performed on 7–29 individuals per population, to a total of 111 (Table 3). Twelve ambiguous PCR products were sequenced for verification.

#### 2.4.5 Microsatellite genotyping

Seven microsatellite loci were used as neutral nuclear markers to study genetic diversity and differentiation of *B. tabaci* populations and cytotypes (Table 4). *B. tabaci* is a haplodiploid species, i.e. haploid males hatch from unfertilised eggs; therefore, genotyping was performed on females only. As only one Q cytotype has been described to date in Greece and Israel (Q1 and Q2 respectively), they were expected never to have been hybridised. In the present study, populations located in these countries were considered as references for Q1 and Q2 cytotypes. Around 30 individuals were genotyped from each population (Table 1). Among the 531 individuals collected in France and Spain between 2006 and 2010,

242 individuals were genotyped, originating from localities where Q1 and Q2 cytotypes were found in sympatry (two sites per country, indicated with an asterisk in Table 1).

PCR reactions were performed separately for each primer pair in 10 µL volumes containing 200 µM of dNTP, 200 nM of each primer (fluorescently labelled 5' primer) and 0.5 IU of Blue Taq DNA polymerase (Eurobio). Cycling conditions were initial denaturing at 94 °C for 2 min, followed by 35 cycles of 30 s at 94 °C for denaturation, 30 s at 52 or 57 °C, depending on the primer set (Table 4), for annealing and 1 min at 72 °C for elongation, with a final elongation at 72 °C for 10 min. Fluorescent amplicons were loaded on an ABIPrism 3100-Avant genetic analyser (Applied Biosystems, Foster City, CA) automated sequencer, and allele sizes were calculated using GeneMapper software (Applied Biosystems).

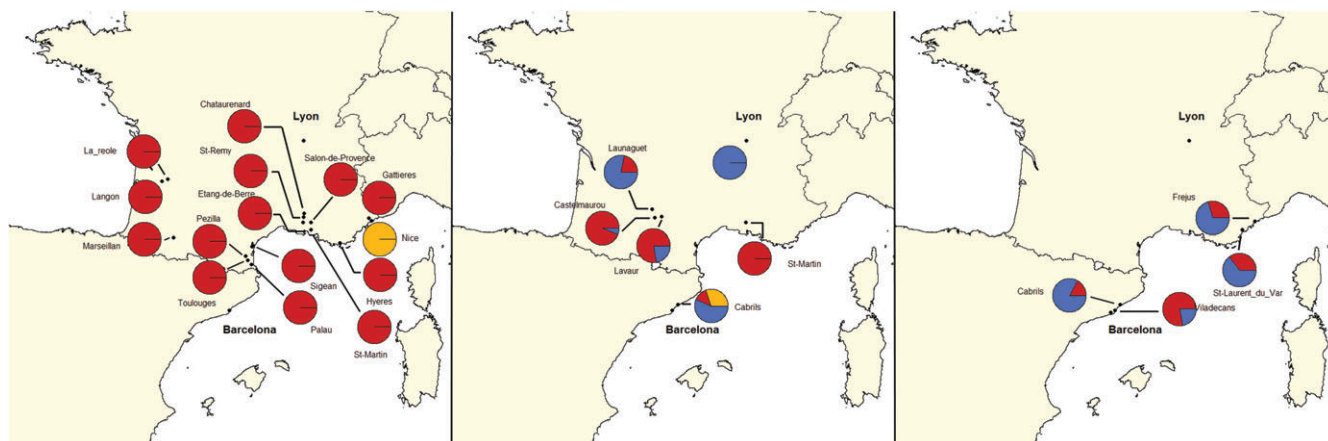
#### 2.5 Genetic data analyses

The mean number of alleles per locus ( $A^-$ ), the observed heterozygosity ( $H_o$ ) and the unbiased expected heterozygosity ( $H_e$ ) were calculated using Genetix.<sup>42</sup> Weir and Cockerham's estimators of  $F_{IS}$  within populations and genetic groups were calculated using Genepop, and exact tests were used for Hardy–Weinberg equilibrium.<sup>43,44</sup> Linkage disequilibrium between all pairs of loci

**Table 4.** Characteristics of the primers used for microsatellite genotyping

Locus	Primer name	Primer sequence	Reference	$T_m$ (°C)
BtIS2.3	BtIs2.3F	5'-CAGAACGACAGGTCGAG-3'	50	50
	BtIs2.3R	5'-CAAAATTAATGGTATTGACTC-3'		
BtIS1.13	BtIs1.13F	5'-CTAAGACCGATTCTCC-3'	50	52
	BtIs1.13R	5'-GAATACTACACCTTCAATTACC-3'		
Locus 11	11a	5'-CCAGAAAAGTGGACTTAAGA-3'	78	57
	11b	5'-GATCTGGGTGTTTTCTTCTA-3'		
BT83	BT83F	5'-GATGCCACAGTTGTCTGG-3'	79	57
	BT83R	5'-GCTTGCCAGGCACTTCTAG-3'		
BtIS1.1	BtIs1.1F	5'-CCCATAGAACACGCTCC-3'	50	57
	BtIs1.1R	5'-CATTGGAAGCCTCGAATAC-3'		
BtIS1.2	BtIs1.2F	5'-CTTACCTTCCATTCACC-3'	50	57
	BtIs1.2R	5'-ATCCCGAGTCTTATGTTG-3'		
Locus 145	145a	5'-CCTACCCATGAGAGCGGTAA-3'	29	57
	145b	5'-TCAACAAACGCGTTCTTCAC-3'		





**Figure 1.** Map of cytotypic distribution in France and Spain by year of sampling. Red: Q1 cytotypic (MED species); blue: Q2 cytotypic (MED species); yellow: MEAM1 species.

and  $F$ -estimates of differentiation<sup>43</sup> were calculated and tested with 5000 permutations using Genepop. The number of genetic groups was assessed using the software STRUCTURE,<sup>45</sup> which differentiates mixed populations on the basis of allele frequency at each locus. The following options were used: 500 000 Markov chain Monte Carlo with a burn-in period of 50 000, where admixture and correlated allele frequencies were allowed for; log-likelihood estimates were calculated for  $K = 1 - 15$ , with ten replicates each. STRUCTURE assumes Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium within each group, which is why both hypotheses were tested subsequently using exact tests implemented in Genepop 4.0. A principal component analysis (PCA) was also conducted, which did not make any assumption such as HWE and linkage disequilibrium, with R software,<sup>46</sup> *ade4* package.<sup>47</sup> In addition, a hierarchical analysis of molecular variance (AMOVA) was performed, using the *hierfstat* package,<sup>48</sup> implemented in R software, to test for geographical, host-plant, cytotypic and secondary symbiont effects. Significance was assessed with permutation tests (5000 permutations). Resistance allele frequencies were subjected to Fisher's exact test with simulated  $P$ -values based on 1000 replicates using R statistical software,<sup>46</sup> while HWE was tested by  $\chi^2$  with R statistical software.<sup>46</sup>

### 3 RESULTS

#### 3.1 Field survey in France and Spain

From 2006 to 2010, a survey of *B. tabaci* populations was performed in France and Spain to follow the geographical distribution of *B. tabaci* cytotypes. A total of 531 individuals from ten localities were collected (Table 1) and compared with the data of Dalmont *et al.*<sup>29</sup> and Gueguen *et al.*,<sup>10</sup> which were used as the historical baseline situation in Western Europe (Fig. 1a). These previous studies were performed from 2003 to 2006 and included 14 localities; only the Q1 cytotypic of MED species was observed. Here, almost all individuals collected belong to the MED species, except in two localities where the MEAM1 species has been previously observed, in a botanical garden in Nice (France) in 2004 and in Cabrils (Spain) in 2007. However, as this situation concerns very few individuals and is rather an exception, MEAM1 species will not be taken into account and discussed further in this study. In contrast to the situation in 2003–2006, with only the Q1 cytotypic, in 2007 the Q2 cytotypic was observed for the first time, in several localities, either in sympatry with Q1 individuals (four locations in France and Spain)

or alone in a greenhouse in Lyon (France) with a very high population density. The coexistence of Q1 and Q2 cytotypes was also observed in 2010 at all of the sampling sites (Fig. 1). The proportion of Q2 individuals in sympatry with Q1 varied, according to the collection, from 6 to 83% in 2007. They predominated in three (83, 71 and 64%) of the four localities in 2010 (22% frequency at the site where Q1 prevails).

#### 3.2 Analysis of reference Q1 and Q2 cytotypes

To date, only the Q1 cytotypic has been detected in Greece and Q2 in Israel.<sup>27,49,50</sup> In the present sampling, the determination of the *mtCOI* cytotypic of individuals collected in three localities in Greece (ten per locality) and one site in Israel ( $n = 29$ ) confirmed these data. They were therefore used as reference populations for Q1 and Q2 cytotypes.

##### 3.2.1 Endosymbiotic community

Individual detection of the seven main bacteria infecting *B. tabaci* revealed that Q1 cytotypes in Greece and Q2 cytotypic in Israel harbour a specific symbiotic community, as already reported.<sup>10,27</sup> The most common bacteria are *Hamiltonella/Wolbachia* for Q1 and the association *Arsenophonus/Rickettsia* for Q2 (Fig. 2).

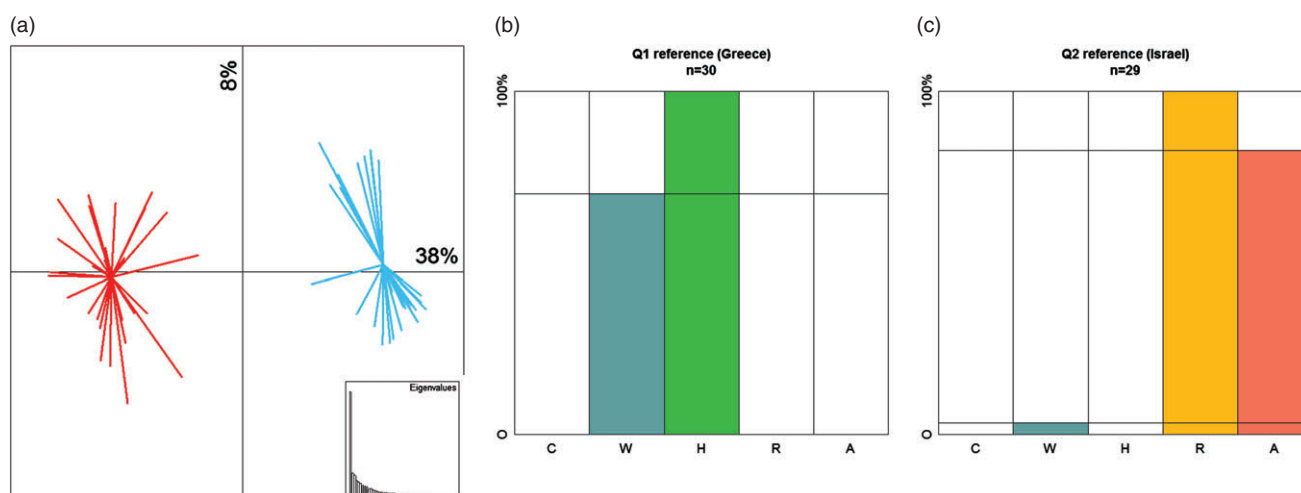
##### 3.2.2 Detection of sodium channel and *ace1* resistance mutations

A resistance allele in *ace1* (F331W) and two resistance alleles of *para*-type voltage-gated sodium (L925I and T929V) genes were detected in both reference cytotypes Q1 and Q2: F331W (R allele) was fixed in Q1 and Q2, while L925I (r1 allele) showed intermediate frequencies in the two cytotypes (0.70 and 0.74 respectively). Q1 and Q2 cytotypes differed significantly only in the frequencies of the T929V mutation (r2 allele; 0.02 and 0.29 respectively; Fisher's exact test:  $P < 10^{-4}$ ).

##### 3.2.3 Microsatellite analysis

The genetic analysis of the 59 individuals genotyped (30 from Greece and 29 from Israel) with seven microsatellites revealed a higher allelic richness in the Q1 than in the Q2 cytotypic (4.3 vs 3.4 alleles) (Table 5). Analysis carried out using STRUCTURE revealed a very high genetic differentiation between Q1 and Q2, which was confirmed by PCA (Fig. 2). The pairwise  $F_{ST}$  value reached 0.27, which is very high for two groups that belong to the same putative species.





**Figure 2.** Characteristics of Greek Q1 and Israel Q2 cytotypes: (a) Factorial map of the principal component analysis; (b) infection status of Q1 individuals; (c) infection status of Q2 individuals. C: *Cardinium*; W: *Wolbachia*; H: *Hamiltonella*; R: *Rickettsia*; A: *Arsenophonus*.

**Table 5.** Microsatellite genetic diversity of *B. tabaci* populations<sup>a</sup>

Country	Locality	Plant	Q1					Q2					Q1 + Q2				
			N	H <sub>o</sub>	H <sub>e</sub>	F <sub>IS</sub>	A <sup>-</sup>	N	H <sub>o</sub>	H <sub>e</sub>	F <sub>IS</sub>	A <sup>-</sup>	N	H <sub>o</sub>	H <sub>e</sub>	F <sub>IS</sub>	A <sup>-</sup>
Spain	Cabrils	Melon	5	0.41	0.42	0.14	2.3	26	0.38	0.49	0.25	4.4	31	0.39	0.51	0.24	4.6
Spain	Cabrils	Tomato	7	0.33	0.43	0.31	3.3	33	0.35	0.49	0.30	4.3	40	0.35	0.50	0.30	4.6
Spain	Viladecans	Tomato	21	0.45	0.58	0.25	4.6	10	0.43	0.57	0.29	3.3	31	0.44	0.60	0.26	4.9
Spain	Viladecans	Cucumber	40	0.41	0.59	0.33	4.1	7	0.44	0.55	0.27	3.3	47	0.41	0.60	0.30	4.3
France	Fréjus	Hibiscus	14	0.42	0.52	0.23	3.4	34	0.37	0.52	0.31	3.6	48	0.38	0.53	0.28	4.9
France	St Laurent du Var	Eggplant	30	0.39	0.52	0.30	3.9	15	0.40	0.55	0.33	4.1	45	0.39	0.55	0.32	4.9
Greece	Thessaloniki	Cucumber	10	0.28	0.56	0.52	3.4	0	-	-	-	-	-	-	-	-	-
Greece	Kasteli-Chania	Eggplant	10	0.38	0.56	0.34	3.4	0	-	-	-	-	-	-	-	-	-
Greece	Ierapetra	Cucumber	10	0.43	0.54	0.21	3.4	0	-	-	-	-	-	-	-	-	-
Israel	Hof Carmel	Various	0	-	-	-	-	29	0.4	0.52	0.37	4.3	-	-	-	-	-

<sup>a</sup> N: number of individuals; H<sub>o</sub>: observed heterozygosity; H<sub>e</sub>: unbiased expected heterozygosity; F<sub>IS</sub>: fixation indice from Weir and Cockerham; A<sup>-</sup>: mean number of alleles per population per locus.

### 3.3 Nuclear and cytoplasmic differentiation of sympatric Q1 and Q2 cytotypes

#### 3.3.1 Endosymbiotic community

*Fritschea* was never detected and will not be discussed further. The bacterial screening did not reveal any new association between a mitochondrial haplotype and a symbiotic community compared with the reference populations (Israeli and Greek), meaning that the cytoplasmic associations are stable (Fig. 3). In the Q1 cyto-type, *Hamiltonella* was almost fixed (93% of individuals infected), while *Wolbachia* and *Cardinium* were present at intermediate frequencies (28 and 38% respectively), with a high frequency of bi-infections (62%). In the Q2 cyto-type, *Rickettsia* was also almost fixed (95%), while *Wolbachia* and *Arsenophonus* were found at a lower frequency (69 and 62% respectively). In Lyon, *Wolbachia* was absent.

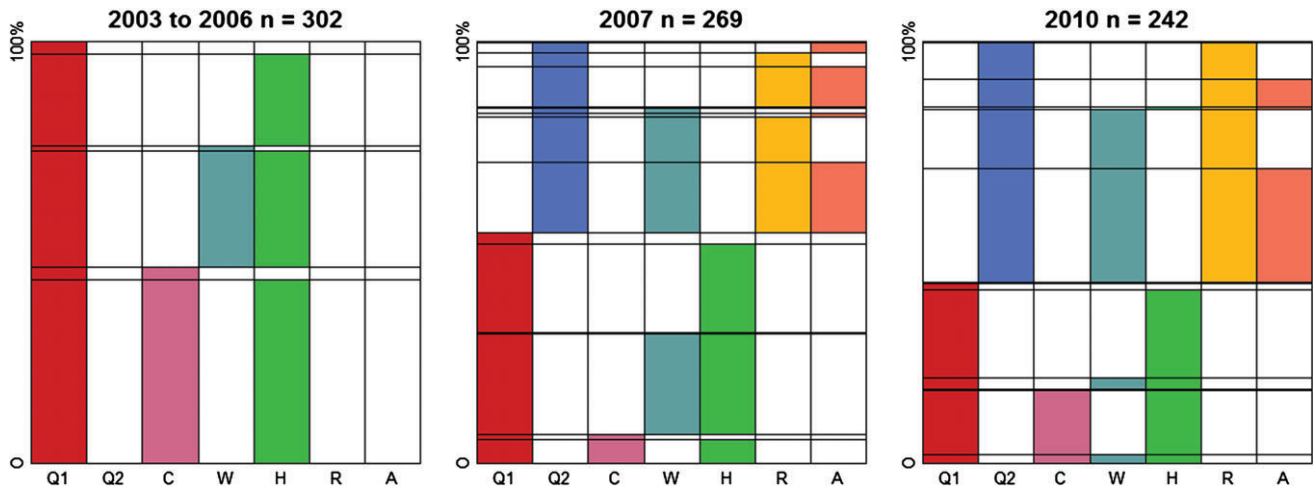
#### 3.3.2 Detection of sodium channel and ace1 resistant mutations

In the French and Spanish sympatric populations sampled in 2010 (111 females tested), all three resistance mutations were detected in both Q1 and Q2 cytotypes in all of the localities investigated, with a high variability in frequency (Table 3). All individuals were homozygous for F331W (R allele) of the *ace1* gene. L925I (r1

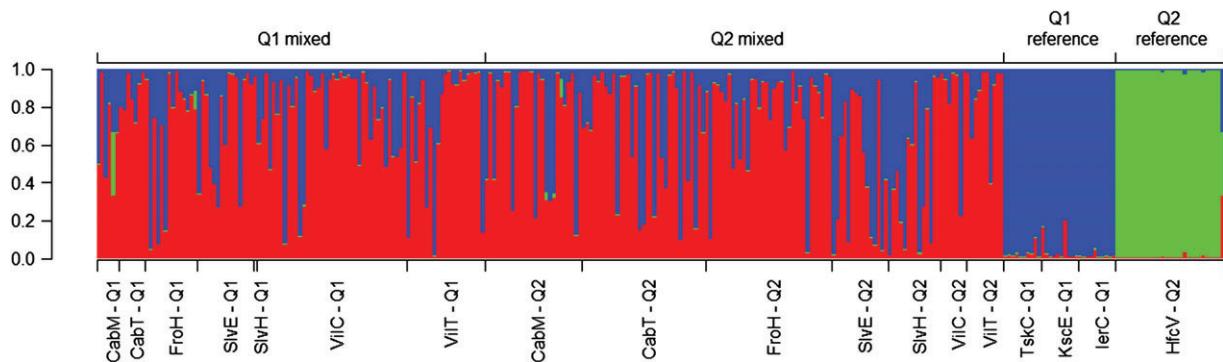
allele) showed a high frequency (0.56 to 1), and the frequencies of T929V (r2 allele) were low (0–0.30). These frequencies did not differ between Q1 and Q2 cytotypes within all four localities (Fisher's exact tests:  $P > 0.05$ ). No significant deviations from Hardy–Weinberg equilibrium were detected within three localities (regardless of the cyto-type involved) for both *kdr* mutations (Fréjus, Saint-Laurent du Var and Viladecans,  $\chi^2$ :  $P > 0.05$ ). Only the population from Cabrils deviated slightly from HWE for T929V ( $\chi^2$ :  $P = 0.01$ ).

#### 3.3.3 Microsatellite analysis

Field populations genotyped for seven microsatellite loci (Table 5) showed a significant deviation from HWE ( $P < 0.001$ ). Within all localities, a systematic heterozygous deficiency was observed that revealed high F<sub>IS</sub> values within localities (0.243–0.357). This deficiency was not due to a particular locus, as the observed heterozygosity was inferior to the expected heterozygosity in all loci. Within the four localities, the F<sub>IS</sub> was not higher when all individuals were considered than when cytotypes were considered separately (Table 5), suggesting that high F<sub>IS</sub> is not explained by a genetic structure that differs among cytotypes. This result was confirmed by STRUCTURE software analysis, as three clusters were



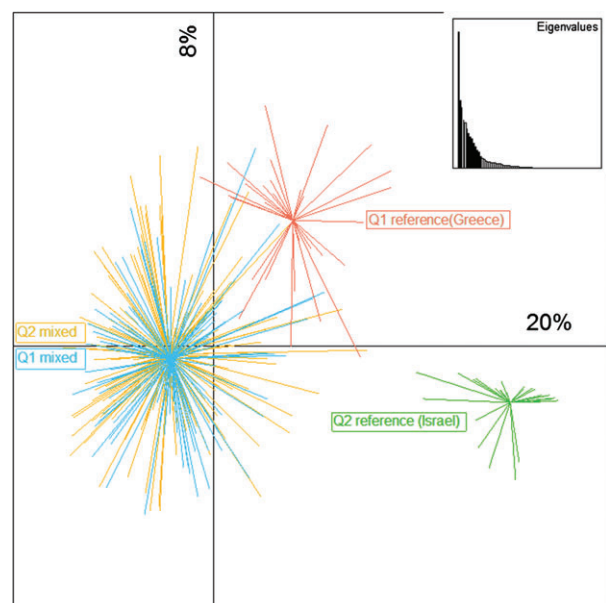
**Figure 3.** Infection status of females. C: *Cardinium*; W: *Wolbachia*; H: *Hamiltonella*; R: *Rickettsia*; A: *Arsenophonus*.



**Figure 4.** Clustering results from STRUCTURE for all samples.

identified when taking into account all sympatric and reference populations. The first cluster regrouped all of the Q2 reference individuals (from Israel). All Greek individuals (Q1 reference) were assigned to a second cluster. Q1 or Q2 cytotype individuals from sympatric populations were assigned, with probabilities more or less important, to the Greek Q1 reference cluster and a third cluster. Probabilities of sympatric individual assignment to each cluster are shown in Fig. 4. Sympatric Q1 and Q2 cytotypes thus showed a similar genetic composition that probably resulted from the hybridisation of two entities, one being close to the Greek population. The PCA (Fig. 5) of allele frequencies confirmed the STRUCTURE software Bayesian analysis. The first axis clearly separates the three groups: mixed Q1–Q2 (western part), reference Q1 (Greece) and reference Q2 (Israel). The second axis weakly discriminates between Greek Q1 and mixed populations. Pairwise comparisons between localities of these geographic groups are very high ( $0.09 < F_{ST(mixed-Greece)} < 0.14$ ;  $0.29 < F_{ST(mixed-Israel)} < 0.33$ ). The point is that Q1 and Q2 cytotypes are totally homogeneous when they are sympatric.

The authors have successively tested for the influence of plants and the symbiotic compartment on the genetic structure within the four localities. The AMOVA was not significant ( $P > 0.1$ ) for all of these factors. The only structuring genetic factor was geographical differentiation: *F*-statistics showed little but significant differentiation between pairs of these four French and Spanish localities ( $F_{ST}$  ranging from 0.01 to 0.02,  $P < 0.01$ ). Allelic richness was quite homogeneous among the four localities where sympatric cytotypes were observed (values ranging from 4.3 to



**Figure 5.** Factorial map of the principal component analysis with the entire dataset.

4.9), which is not significantly higher than the allelic richness detected within localities from reference Q1 (Greek) and Q2 (Israeli) cytotypes.

**Table 6.** Results of crossing experiments. Mean eggs laid per replicate ( $\pm$ SD), mean emerging adults ( $\pm$ SD) and mean percentage of emerging adults reported to the number of laid eggs ( $\pm$ SD)

	Cross <sup>a</sup>			
	Q1♀ × Q1♂	Q1♀ × Q2♂	Q2♀ × Q1♂	Q2♀ × Q2♂
Eggs	127* ± 43	75 ± 13	40 ± 13	65 ± 9
Emerging adults	32 ± 15	48* ± 9	32 ± 12	39 ± 8
Emerging adults (%)	24* ± 14	65 ± 13	79 ± 5	62 ± 16

<sup>a</sup> An asterisk (\*) denotes a significant difference from the others (Tukey's HSD test,  $P < 0.01$ ). The number of replicates in each cross was 10.

### 3.3.4 Crossing experiments

Viable offspring with both males and females were obtained in all of the crosses performed, suggesting that there is neither nuclear nor complete cytoplasmic incompatibilities between Q1 and Q2 cytotypes. The mean number of eggs laid and the percentage of non-hatched eggs are indicated in Table 6. The Q1♀ × Q1♂ cross produced more eggs than the other three (Tukey's HSD test:  $P < 0.001$ ), but the percentage of hatched eggs was lower (24%, Tukey's HSD test:  $P < 0.001$ ). The number of offspring was slightly higher in the Q1♀ × Q2♂ cross than in the others (Tukey's HSD test:  $P = 0.01$ ). Fertility of F1 individuals (males and females) was verified by keeping them together for mating, and the presence of a high number of males and females confirmed hybridisation among both cytotypes.

## 4 DISCUSSION

### 4.1 Cytotype distribution in Western Europe

*B. tabaci* is widely distributed in France and Spain, and the putative MED species largely predominates in these areas, such as in most of the Mediterranean basin. The MED species has been found in all of the Spanish and French localities where sampling was performed from 2003 to 2006. In very few areas, the MEAM1 putative species was observed in France (Nice) in 2004 and in Spain (Cabrils) in 2007, but this presence is anecdotal. Until 2007, the Q1 cytotype, which associates a particular mitochondrial variant with *Hamiltonella* and *Cardium* bacteria, was the only cytotype belonging to the MED species found in France and Spain, and more generally in the large western part of the Mediterranean basin.<sup>29,51</sup> In 2007, the Q2 cytotype was detected for the first time in four out of the six localities studied in France and Spain. Q2 is another mitochondrial variant infected by *Rickettsia* and *Arsenophonus* symbionts previously reported in the Middle East (Israel and Turkey exclusively). In 2010, Q2 was found in four other localities in France and Spain, always in sympatry with Q1. Because the initial situation in these countries was well known, this observation was interpreted as a recent introduction of the Q2 cytotype, which has also been recently observed in other localities in the Western Mediterranean basin.<sup>39,52</sup> This introduction is probably due to trade, which occurs intensively among Mediterranean countries. Indeed, in France, ornamental plants grow up in greenhouses, but the sowing and cloning is mostly performed in other countries such as Israel, Kenya, the Ivory Coast, Germany, Spain and Italy (private communication). It can be supposed that Q2 was introduced from the Eastern part of the Mediterranean basin (Middle East), as it has always been detected over there,<sup>10,33,36</sup> and is now spreading in Western Mediterranean countries. Extensive surveys of *B. tabaci* populations on a larger

geographical scale would make it possible to determine whether this situation is particular or whether it is occurring on a wider geographic scale in the Mediterranean basin. Nevertheless, these data raise the question of the ability of the Q2 cytotype to settle in these localities, whereas Q1 was already present, which is related to the question of cytotype coexistence discussed in Section 4.4. Some authors suggested a role of different interacting agroecological factors, such as insecticide use and high temperature (more suitable for Q2), as well as the extensive production of Solanaceae plants, in the invasion of Q2 in southern Italy.<sup>52</sup>

### 4.2 Q1 and Q2 differentiation and hybridisation

Until 2007, Q1 and Q2 cytotypes were known in allopatric conditions in the Mediterranean basin, with Q2 only reported in Israel, with its possible presence in bordering countries, while Q1 had a larger distribution from Greece to France and northern Africa. This separate distribution is consistent with the very high nuclear genetic differentiation that was measured in this study between the Q1 and the Q2 reference cytotypes (Greece and Israel respectively). The  $F_{ST}$  value (about 29%), measured with microsatellite markers, is comparable with those obtained between distinct *B. tabaci* species such as MED and MEAM1.<sup>53,54</sup> This divergence indicates that they probably did not hybridise recently, as they had not been in contact, leading the authors to consider these lines as reference cytotypes, even if the nuclear genetic compositions observed in these countries are probably not representative of all of the diversity that can exist in each cytotype. Moreover, this small sampling could explain the high differentiation values and the detection of Israeli Q2 specific alleles (many alleles from Q2 reference cytotype are not found anywhere else). The present Q2 reference is therefore certainly not the population source of the western Q2 cytotypes, as suggested by the microsatellite analysis comparison of allopatric and sympatric populations. Populations experiencing strong bottlenecks during colonisation processes or pesticide treatments inducing reduced population sizes could also explain this high differentiation.

This new situation with Q1 and Q2 cytotypes observed in sympatry in France and Spain provided the opportunity to test for their interfertility. The microsatellite analysis revealed three highly differentiated clusters: Israeli Q2 reference, Greek Q1 reference and the western mixed cytotypes. These results indicated that Q1 and Q2 are genetically homogeneous in sympatry (i.e. in France and Spain) and clearly demonstrate their full hybridisation and the absence of any reproductive barrier among them. They also do not differ for the insecticide-resistant allele frequency for the two genes considered here, i.e. the *ace1* (resistance is fixed) and the *para*-type voltage-gated sodium channel genes. Crossing experiments in the lab confirmed that Q1 and Q2 individuals can interbreed and produce fertile offspring, although care must be taken as the authors conducted crossing with Q2 and Q1 lines from Lyon and Sigean (France) respectively, the nuclear genetic backgrounds of which are unknown; therefore, a different result with Israeli Q2 and Greek Q1 reference cytotypes, which are strongly differentiated, cannot be excluded. All of these facts indicate that there is no nuclear incompatibility between them and that they belong to the same species, as suggested by phylogenetic analyses.<sup>5</sup>

### 4.3 Biological relevance of Q1 and Q2 cytotypes

#### 4.3.1 Cytotype stability

Q1 and Q2 freely interbreed but harbour different symbiont communities, even when they coexist in sympatry (Q1 is infected by

*Hamiltonella* and *Cardinium*, while Q2 is infected by *Rickettsia* and *Arsenophonus*). This means that the *mtCOI* haplotype and bacterial associations are relatively stable on a timescale of several years, thus giving biological relevance to cytotypes under the hypothesis of a role for bacteria in the *B. tabaci* phenotype. The absence of new mitochondrial haplotype/bacterial community associations can be explained either by the absence or very rare events of horizontal transfers or the inability of these bacteria to colonise new cytoplasmic and/or nuclear environments. Symbiotic, vertically transmitted bacteria are also known to be horizontally transferred. Horizontal transmission can occur through the diet between individuals that feed on the same host plant, as demonstrated in aphids and fruit flies,<sup>55,56</sup> or during mating.<sup>57,58</sup> Transfers can also occur between hosts and parasitoids.<sup>59,60</sup> Phylogenetic analysis revealed the existence of such transfers in *B. tabaci*.<sup>61</sup> Laboratory experiments demonstrated indeed that *Rickettsia* can be horizontally transmitted between *B. tabaci* individuals through the host plant.<sup>62</sup> Therefore, a total absence of horizontal transfer is unlikely, but it is possible that lateral transfer between cytotypes has not yet occurred (at the time of this study) because the contact of Q1 and Q2 cytotypes is new. Another explanation could be that new *mtCOI*/bacteria associations are unstable either because of a maladaptation of the transferred bacteria in the new environment or exclusion by the recipient cytotype. This exclusion can be achieved by the host nuclear genes or by the bacterial endosymbionts that are already present. As the two cytotypes interbreed, it can be supposed that their nuclear compartments are homogeneous. Therefore, the existence of horizontal transfer is possible, even if it cannot be ruled out that some nuclear genes specific to each cytotype could be involved in bacterial exclusion. Within-host, multiple infections could select for competitive symbionts that monopolise resources at the expense of the other, allowing their own vertical transmission. There are some examples of such cytoplasmic competition between bacteria.<sup>63,64</sup> Nevertheless, the stability of the *mtCOI*/symbiotic community associations indicates that the concept of cytotype has a real biological significance.

#### 4.3.2 Presence of bacteria

Q1 and Q2 hybridise, in spite of harbouring different symbiotic communities, suggesting that bacteria do not induce any cytoplasmic incompatibility in *B. tabaci* or low-level cytoplasmic incompatibility that was not detectable in the present experimental conditions. Clearly, laboratory experiments, such as precise measures of sex ratio, are required to affirm that these bacteria do not induce cytoplasmic incompatibility. However, it was shown in the present study that bacteria do not prevent hybridisation and gene flow between Q1 and Q2. As symbionts necessarily induce a physiological cost on their host,<sup>64,65</sup> they are probably involved in other phenotypic effects that allow them to spread and be maintained at high frequency in host populations (see the review in Ferrari and Vavre<sup>16</sup>). Some symbiotic bacteria manipulate host reproduction and increase their transmission, but this is apparently not the case in *B. tabaci*. Others have a mutualistic strategy and confer direct fitness benefits to their host, such as protection against natural enemies,<sup>17,66</sup> thermal tolerance<sup>67,68</sup> or host-plant specialisation.<sup>19</sup> For *B. tabaci*, there is little information about the phenotypic effects of the symbionts. It has been demonstrated that *Rickettsia* enhances host fitness by increasing fecundity and survival to adulthood,<sup>20</sup> and is also suspected to increase sensibility to insecticides.<sup>18</sup> *Wolbachia* could protect *B. tabaci* from parasitoids.<sup>69</sup> In Q1 and in Q2, one symbiont is fixed in natural populations, *Hamiltonella* and *Rickettsia* respectively, while others are

present at intermediate frequencies (around 38% for *Cardinium* in Q1 and 60% for *Arsenophonus* in Q2). It is possible that *Rickettsia* increases host performance in the Q2 populations of France and Spain, which is suspected in Italy;<sup>52</sup> however, this trait has only been demonstrated in Q2 populations in the United States, and has not been observed in Q2 populations in Israel. Future studies should focus on the phenotypic effects of the bacteria present in Q1 and Q2 in these mixed populations of France and Spain.

#### 4.4 Cytotype coexistence

The sympatry of Q1 and Q2 observed in France and Spain is relatively new (only since 2007), and the Q2 cytotype has spread throughout Western Europe without excluding Q1 until now. The question that arises is how this situation will evolve and whether these two cytotypes will still coexist or if one will be excluded. The coexistence of several *B. tabaci* cytotypes is uncommon and is generally a transient situation that leads to the displacement and/or exclusion of one of the competing cytotypes.<sup>51,70,71</sup> Each greenhouse can provide an environment (insecticide treatment, temperature, type of crop) that is more or less suitable to one cytotype, excluding the less adapted cytotypes. In this case there should be a mosaic of situations, even on a local scale, where cytotypes are allopatric from one greenhouse to another. Among the three previously cited factors, no data are available for temperature or for the use of insecticides. Moreover, most of the mechanisms known to confer resistance to insecticides are nuclear, while the two cytotypes freely interbreed and reference Q1 and Q2 populations show a similar composition of insecticide resistance alleles. However, *Rickettsia* (associated with the Q2 cytotype) could increase susceptibility to some insecticides<sup>18</sup> and thus disadvantage the Q2 cytotype in such an environment; however, a recent study proposed the opposite.<sup>52</sup> Parrella *et al.*<sup>52</sup> also suspect that host plants play a role in cytotype competition for the benefit of Q2, but the present authors did not find any difference in terms of host-plant preference between cytotypes.

However, it is possible that these two cytotypes coexist with a long-term equilibrium. Mixed populations could be maintained through plant rotation and the recurrent introductions of cytotypes. A high *B. tabaci* individual flow between glasshouses is supported by the fact that observed geographical differentiation is very low: only 1%  $F_{ST}$  between localities within a country, and 2% between countries. This low differentiation is in agreement with previous results.<sup>29,39</sup> This is probably due to the commercial exchanges of vegetable and ornamental crops (see above and Byrne *et al.*<sup>72</sup>). Clearly, a temporal survey coupled with an analysis of the nuclear compartment would provide more insight into the evolution of the Q1 and Q2 distribution in Europe.

## 5 CONCLUSION

The first detection of the Q2 cytotype of *B. tabaci* MED species in Western Europe (France and Spain) is reported here. A short survey over a period of 5 years indicated that Q2 seems to increase in frequency, but without the exclusion of the Q1 cytotype. Both cytotypes are always found in sympatry, except in one locality (Lyon, France). Genetic analysis showed hybridisation of the two cytotypes in the field, in spite of a very high differentiation observed among allopatric sampling. This interfertility is confirmed by crossing experiments in the lab. These results are consistent with the phylogenies based on *mtCOI*, which regroup Q1 and Q2 cytotypes in the same MED species, and indicate that symbionts do not prevent hybridisation between Q1 and Q2. The present data also show



that the stable associations of mitochondrial haplotypes with bacteria confer a biological reality to the Q1 and Q2 cytotypes.

## ACKNOWLEDGEMENTS

This work was funded by CNRS (IFR41-UMR5558) and conducted within the framework of the first call on Mediterranean agriculture carried out by ARIMNet, with funding from MOARD (IL), ANR (FR), INIA (ES), NAGREF-DEMETER (GR) and GDAR (TR). The authors are grateful to G Bout, A Tsagkarakou, E Zchori-Fein, J Vendeville and M Peterschmitt for sample collection. Gabriel Terraz is the recipient of a PhD studentship from the Rhône-Alpes region ('Program Cible' Grant). We thank the greenhouse technical platform (Lyon1) and especially Elise Lacroix.

## REFERENCES

- Jones DR, Plant viruses transmitted by whiteflies. *Eur J Plant Pathol* **109**:195–219 (2003).
- Boykin LM, Shatters RG, Rosell RC, McKenzie CL, Bagnall RA, De Barro P et al., Global relationships of *Bemisia tabaci* (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Mol Phylogenet Evol* **44**:1306–1319 (2007).
- Horowitz AR, Kontsedalov S, Khasdan V and Ishaaya I, Biotypes B and Q of *Bemisia tabaci* and their relevance to neonicotinoid and pyriproxyfen resistance. *Arch Insect Biochem Physiol* **58**:216–225 (2005).
- Perring TM, The *Bemisia tabaci* species complex. *Crop Prot* **20**:725–737 (2001).
- De Barro PJ, Liu S-S, Boykin LM and Dinsdale AB, *Bemisia tabaci*: a statement of species status. *Annu Rev Entomol* **56**:1–19 (2011).
- Dinsdale AB, Cook L, Riginos C, Buckley YM and De Barro P, Refined global analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodidae: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. *Ann Entomol Soc Am* **103**:196–208 (2010).
- Xu J, De Barro PJ and Liu S-S, Reproductive incompatibility among genetic groups of *Bemisia tabaci* supports the proposition that the whitefly is a cryptic species complex. *Bull Entomol Res* **100**:359–366 (2010).
- Liu S, Colvin J and De Barro PJ, Species concepts as applied to the whitefly *Bemisia tabaci* systematics: how many species are there? *J Integr Agric* **11**:176–186 (2012).
- Lee W, Park J, Lee G-S, Lee S and Akimoto S, Taxonomic status of the *Bemisia tabaci* complex (Hemiptera: Aleyrodidae) and reassessment of the number of its constituent species. *PLoS ONE* **8**:e63817 (2013).
- Gueguen G, Vavre F, Gnankiné O, Peterschmitt M, Charif D, Chiel E et al., Endosymbiont metacommunities, mtDNA diversity and the evolution of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. *Mol Ecol* **19**:4365–4378 (2010).
- Chu D, Hu X, Gao C, Zhao H, Nichols RL and Li X, Use of mitochondrial cytochrome oxidase I polymerase chain reaction-restriction fragment length polymorphism for identifying subclades of *Bemisia tabaci* Mediterranean group. *J Econ Entomol* **105**:242–251 (2012).
- Jiggins FM and Hurst GDD, Rapid insect evolution by symbiont transfer. *Science* **332**:185–186 (2011).
- Kriesner P, Hoffmann AA, Lee SF, Turelli M and Weeks AR, Rapid sequential spread of two *Wolbachia* variants in *Drosophila simulans*. *PLoS Pathog* **9**:e1003607 (2013).
- Turelli M, Hoffmann A and McKechnie S, Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations. *Genetics* **132**:713–723 (1992).
- Engelstädter J and Hurst GDD, The ecology and evolution of microbes that manipulate host reproduction. *Annu Rev Ecol Syst* **40**:127–149 (2009).
- Ferrari J and Vavre F, Bacterial symbionts in insects or the story of communities affecting communities. *Phil Trans R Soc Lond B Biol Sci* **366**:1389–1400 (2011).
- Oliver KM, Russell JA, Moran NA and Hunter MS, Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proc Natl Acad Sci USA* **100**:1803–1807 (2003).
- Kontsedalov S, Zchori-fein E, Chiel E, Gottlieb Y, Inbar M and Ghanim M, The presence of *Rickettsia* is associated with increased susceptibility of *Bemisia tabaci* (Homoptera: Aleyrodidae) to insecticides. *Pest Manag Sci* **64**:789–792 (2008).
- Tsuchida T, Koga R and Fukatsu T, Host plant specialisation governed by facultative symbiont. *Science* **303**:1989 (2004).
- Himler AG, Adachi-Hagimori T, Bergen JE, Kozuch A, Kelly SE, Tabashnik BE et al., Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. *Science* **332**:254–256 (2011).
- Brucker RM and Bordenstein SR, Speciation by symbiosis. *Trends Ecol Evol* **27**:443–451 (2012).
- Santos-Garcia D, Farnier P-A, Beitia F, Zchori-Fein E, Vavre F, Mouton L et al., Complete genome sequence of *Candidatus Portiera aleyrodidarum* BT-QVLC, an obligate symbiont that supplies amino acids and carotenoids to *Bemisia tabaci*. *J Bacteriol* **194**:6654–6655 (2012).
- Sloan DB and Moran NA, Endosymbiotic bacteria as a source of carotenoids in whiteflies. *Biol Lett* **8**:986–989 (2012).
- Zchori-Fein E and Brown JK, Diversity of prokaryotes associated with *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae). *Ann Entomol Soc Am* **95**:711–718 (2002).
- Bing X-L, Ruan Y-M, Rao Q, Wang X-W and Liu S-S, Diversity of secondary endosymbionts among different putative species of the whitefly *Bemisia tabaci*. *Insect Sci* **20**:194–206 (2013).
- Chu D, Jiang T, Liu G-X, Jiang DF, Tao YL, Fan ZX et al., Biotype status and distribution of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Shandong province of China based on mitochondrial DNA markers. *Environ Entomol* **36**:1290–1295 (2007).
- Chiel E, Gottlieb Y, Zchori-Fein E, Mozes-Daube N, Katzir N, Inbar M et al., Biotype-dependent secondary symbiont communities in sympatric populations of *Bemisia tabaci*. *Bull Entomol Res* **97**:407–413 (2007).
- Gnankiné O, Mouton L, Henri H, Terraz G, Houndeté T, Martin T et al., Distribution of *Bemisia tabaci* (Homoptera: Aleyrodidae) biotypes and their associated symbiotic bacteria on host plants in West Africa. *Insect Conserv Divers* **6**:411–421 (2013).
- Dalmon A, Halkett F, Granier M, Delatte H and Peterschmitt M, Genetic structure of the invasive pest *Bemisia tabaci*: evidence of limited but persistent genetic differentiation in glasshouse populations. *Heredity* **100**:316–325 (2008).
- Gorsane F, Ben Halima A, Ben Khalifa M, Bel-Kadhi MS and Fakhfakh H, Molecular characterisation of *Bemisia tabaci* populations in Tunisia: genetic structure and evidence for multiple acquisition of secondary symbionts. *Environ Entomol* **40**:809–817 (2011).
- Tsagkarakou A, Tsigenopoulos CS, Gorman KJ, Lagnel J and Bedford ID, Biotype status and genetic polymorphism of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Greece: mitochondrial DNA and microsatellites. *Bull Entomol Res* **97**:29–40 (2007).
- Guirao P, Beitia F and Cenis JL, Biotype determination of Spanish populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Bull Entomol Res* **87**:587–593 (1997).
- Rua P, Simon B, Cifuentes D, Martinez-Mora C and Cenis JL, New insights into the mitochondrial phylogeny of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in the Mediterranean Basin. *J Zool Syst Evol Res* **44**:25–33 (2006).
- Reynaud P, *Bemisia tabaci* in France (in French). *Phytoma* **527**:18–21 (2000).
- Simon B, Moriones E, Soria C, Beitia F, Bosco D and Cenis JL, Variacion genetica de poblaciones de *Bemisia tabaci* Gennadius en la cuenca del Mediterraneo occidental [Genetic variation in populations of *Bemisia tabaci* Gennadius in the western Mediterranean basin]. In *Resúmenes del Congreso Nacional de Entomología Aplicada, VII Jornadas Científicas de la Sociedad Española de Entomología Aplicada, Aguadulce, Spain, November 8–12, 1999*, Consejería de Agricultura y Pesca, Junta de Andalucía, España, p. 20 (1999).
- Chu D, Wan F-H, Tao Y-L, Liu G-X, Fan Z-X and Bi Y-P, Genetic differentiation of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) biotype Q based on mitochondrial DNA markers. *Insect Sci* **15**:115–123 (2008).
- Hsieh C-H, Wang C-H and Ko C-C, Evidence from molecular markers and population genetic analyses suggests recent invasions of the Western North Pacific region by biotypes B and Q of *Bemisia tabaci* (Gennadius). *Environ Entomol* **36**:952–961 (2007).
- Alon M, Benting J, Lueke B, Ponge T, Alon F and Morin S, Multiple origins of pyrethroid resistance in sympatric biotypes of *Bemisia*

- tabaci* (Hemiptera: Aleyrodidae). *Insect Biochem Mol Biol* **36**:71–79 (2006).
- 39 Gauthier N, Clouet C, Perrakis A, Kapantaidaki D, Peterschmitt M and Tsagkarakou A, Genetic structure of *Bemisia tabaci* Med populations from home range countries inferred by nuclear and cytoplasmic markers: impact on the distribution of the insecticide resistance genes. *Pest Manag Sci* DOI: 10.1002/ps.3733.
  - 40 Henri H, Terraz G, Gnankiné O, Fleury F and Mouton L, Molecular characterisation of genetic diversity within the Africa/Middle East/Asia Minor and Sub-Saharan African groups of the *Bemisia tabaci* species complex. *Int J Pest Manag* **59**:329–338 (2013).
  - 41 Tsagkarakou A, Nikou D, Roditakis E, Sharvit M, Morin S and Vontas J, Molecular diagnostics for detecting pyrethroid and organophosphate resistance mutations in the Q biotype of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Pestic Biochem Physiol* **94**:49–54 (2009).
  - 42 Belkhir K, Borsari P, Chikhi L, Raufaste N and Bonhomme F, GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France (1996–2004).
  - 43 Weir BS and Cockerham CC, Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**:1358–1370 (1984).
  - 44 Raymond M and Rousset F, Genepop (Version 1.2) – population-genetics software for exact tests and ecumenism. *J Hered* **86**:248–249 (1995).
  - 45 Pritchard JK, Stephens M and Donnelly P, Inference of population structure using multilocus genotype data. *Genetics* **155**:945–959 (2000).
  - 46 R: A Language and Environment for Statistical Computing. [Online]. R Development Core Team (2008). Available: <http://www.r-project.org> [25 April 2014].
  - 47 Jombart T, *adeigenet*: an R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**:1403–1405 (2008).
  - 48 Goudet J, *hierfstat*, a package for R to compute and test hierarchical *F*-statistics. *Mol Ecol Notes* **2**:184–186 (2005).
  - 49 Tsagkarakou A, Mouton L, Kristoffersen JB, Dokianakis E, Grispu M and Bourtzis K, Population genetic structure and secondary endosymbionts of *Q Bemisia tabaci* (Hemiptera: Aleyrodidae) from Greece. *Bull Entomol Res* **1–13** (2012).
  - 50 Gauthier N, Dalleau-Clouet C and Bouvret M-E, Twelve new polymorphic microsatellite loci and PCR multiplexing in the whitefly, *Bemisia tabaci*. *Mol Ecol Resour* **8**:1004–1007 (2008).
  - 51 Moya A, Guirao P, Cifuentes D, Beitia F and Cenis JL, Genetic diversity of Iberian populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae) based on random amplified polymorphic DNA-polymerase chain reaction. *Mol Ecol* **10**:891–897 (2001).
  - 52 Parrella G, Nappo AG, Manco E, Greco B and Giorgini M, Invasion of the Q2 mitochondrial variant of Mediterranean *Bemisia tabaci* in southern Italy: possible role of bacterial endosymbionts. *Pest Manag Sci* DOI:10.1002/ps.3686 (2013).
  - 53 Saleh D, Laarif A, Clouet C and Gauthier N, Spatial and host-plant partitioning between coexisting *Bemisia tabaci* cryptic species in Tunisia. *Popul Ecol* **54**:261–274 (2012).
  - 54 Tahiri A, Halkett F, Granier M, Gueguen G and Peterschmitt M, Evidence of gene flow between sympatric populations of the Middle East–Asia Minor 1 and Mediterranean putative species of *Bemisia tabaci*. *Ecol Evol* **3**:2619–2633 (2013).
  - 55 Darby A and Douglas AE, Elucidation of the transmission patterns of an insect-borne bacterium. *Appl Environ Microbiol* **69**:4403–4407 (2003).
  - 56 Haine ER, Pickup NJ and Cook JM, Horizontal transmission of *Wolbachia* in a *Drosophila* community. *Ecol Entomol* **30**:464–472 (2005).
  - 57 Gonella E, Crotti E, Rizzi A, Mandrioli M, Favia G, Daffonchio D *et al.*, Horizontal transmission of the symbiotic bacterium *Asaia* sp. in the leafhopper *Scaphoideus titanus* Ball (Hemiptera: Cicadellidae). *BMC Microbiol* **12**(Suppl. 1):S4 (2012).
  - 58 Moran NA and Dunbar HE, Sexual acquisition of beneficial symbionts in aphids. *Proc Natl Acad Sci USA* **103**:12 803–12 806 (2006).
  - 59 Vavre F, Fleury F, Lepetit D, Fouillet P and Boulétreau M, Phylogenetic evidence for horizontal transmission of *Wolbachia* in host–parasitoid associations. *Mol Biol Evol* **16**:1711–1723 (1999).
  - 60 Chiel E, Zchori-Fein E, Inbar M, Gottlieb Y, Adachi-Hagimori T, Kelly SE *et al.*, Almost there: transmission routes of bacterial symbionts between trophic levels. *PLoS ONE* **4**:e4767 (2009).
  - 61 Mouton L, Thierry M, Henri H, Baudin R, Gnankiné O, Reynaud B *et al.*, Evidence of diversity and recombination in *Arsenophonus* symbionts of the *Bemisia tabaci* species complex. *BMC Microbiol* **12**(Suppl. 1):S10 (2012).
  - 62 Caspi-Fluger A, Inbar M, Mozes-Daube N, Katzir N, Portnoy V, Belasov E *et al.*, Horizontal transmission of the insect symbiont *Rickettsia* is plant-mediated. *Proc Biol Sci* **279**:1791–1796 (2012).
  - 63 Goto S, Anbutsu H and Fukatsu T, Asymmetrical interactions between *Wolbachia* and *Spiroplasma* endosymbionts coexisting in the same insect host. *Appl Environ Microbiol* **72**:4805–4810 (2006).
  - 64 Oliver KM, Moran NA and Hunter MS, Costs and benefits of a superinfection of facultative symbionts in aphids. *Proc Biol Sci* **273**:1273–1280 (2006).
  - 65 Fleury F, Vavre F, Ris N, Fouillet P and Boulétreau M, Physiological cost induced by the maternally-transmitted endosymbiont *Wolbachia* in the *Drosophila* parasitoid *Leptopilina heterotoma*. *Parasitology* **121**:493–500 (2000).
  - 66 Ferrari J, Darby AC, Daniell TJ, Godfray HCJ and Douglas AE, Linking the bacterial community in pea aphids with host-plant use and natural enemy resistance. *Ecol Entomol* **29**:60–65 (2004).
  - 67 Chen DQ, Montllor CB and Purcell AH, 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 (2000).
  - 68 Montllor C, Maxmen A and Purcell AH, Facultative bacterial endosymbionts benefit pea aphids *Acyrtosiphon pisum* under heat stress. *Ecol Entomol* **189–195** (2002).
  - 69 Xue X, Li S-J, Ahmed MZ, De Barro PJ, Ren S-X and Qiu B-L, Inactivation of *Wolbachia* reveals its biological roles in whitefly host. *PLoS ONE* **7**:e48148 (2012).
  - 70 Khasdan V, Levin I, Rosner A, Morin S, Kontsedalov S, Maslenin L *et al.*, DNA markers for identifying biotypes B and Q of *Bemisia tabaci* (Hemiptera: Aleyrodidae) and studying population dynamics. *Bull Entomol Res* **95**:605–613 (2005).
  - 71 Horowitz AR, Denholm I, Gorman KJ, Cenis JL, Kontsedalov S and Ishaaya I, Biotype Q of *Bemisia tabaci* identified in Israel. *Phytoparasitica* **31**:94–98 (2003).
  - 72 Byrne DN, Rathman RJ, Orum TV and Palumbo JC, Localised migration and dispersal by the sweet potato whitefly, *Bemisia tabaci*. *Oecologia* **105**:320–328 (1996).
  - 73 Gottlieb Y, Ghanim M, Chiel E, Gerling D, Portnoy V, Steinberg S *et al.*, Identification and localisation of a *Rickettsia* sp. in *Bemisia tabaci* (Homoptera: Aleyrodidae). *Appl Environ Microbiol* **72**:3646–3652 (2006).
  - 74 Weeks AR, Velten R and Stouthamer R, Incidence of a new sex-ratio-distorting endosymbiotic bacterium among arthropods. *Proc Biol Sci* **270**:1857–1865 (2003).
  - 75 Braig HR, Zhou W, Dobson SL and O'Neill SL, Cloning and characterisation of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. *J Bacteriol* **180**:2373–2378 (1998).
  - 76 Thao ML and Baumann P, Evidence for multiple acquisition of *Arsenophonus* by whitefly species (Sternorrhyncha: Aleyrodidae). *Curr Microbiol* **48**:140–144 (2004).
  - 77 Thao ML, Baumann L, Hess JM, Falk BW, Ng JCK, Gullan PJ *et al.*, Phylogenetic evidence for two new insect-associated Chlamydia of the family *Simkaniaceae*. *Curr Microbiol* **47**:46–50 (2003).
  - 78 Delatte H, Reynaud B, Granier M, Thornary L, Lett JM, Goldbach R *et al.*, A new silverleaf-inducing biotype Ms of *Bemisia tabaci* (Hemiptera: Aleyrodidae) indigenous to the islands of the south-west Indian Ocean. *Bull Entomol Res* **95**:29–35 (2007).
  - 79 Tsagkarakou A and Roditakis N, Isolation and characterisation of microsatellite loci in *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Mol Ecol Notes* **3**:196–198 (2003).







## Distribution of *Bemisia tabaci* (Homoptera: Aleyrodidae) biotypes and their associated symbiotic bacteria on host plants in West Africa

OLIVIER GNANKINÉ,<sup>1\*</sup> LAURENCE MOUTON,<sup>2\*</sup> HÉLÈNE HENRI,<sup>2</sup>  
GABRIEL TERRAZ,<sup>2</sup> THOMAS HOUNDETÉ,<sup>3</sup> THIBAUD MARTIN,<sup>4</sup>  
FABRICE VAVRE<sup>2</sup> and FRÉDÉRIC FLEURY<sup>2</sup>

<sup>1</sup>Unité de Formation et de Recherche en Sciences de la Vie et de la Terre (UFR-SVT), Laboratoire d'Entomologie Fondamentale et Appliquée, Université de Ouagadougou, Ouagadougou, Burkina Faso, <sup>2</sup>Laboratoire de Biométrie et Biologie Evolutive, UMR CNRS 5558, Université de Lyon, Université Lyon1, Villeurbanne Cedex, France, <sup>3</sup>Institut National des Recherches Agricoles du Bénin, Cotonou, Benin and <sup>4</sup>UR Hortsys, Centre de Coopération Internationale en Recherche Agronomique pour le développement (CIRAD), Montpellier, France

**Abstract.** 1. The whitefly *Bemisia tabaci* is a pest of many agricultural and ornamental crops worldwide and particularly in Africa. *B. tabaci* is a complex of more than 20 biotypes. Effective control of *B. tabaci* calls for a greater knowledge of the local biological diversity in terms of biotypes or putative species. Information is available about biotype distribution in Northern, Eastern, and Southern Africa, but data for Western Africa remain very scarce. At the time of this study, data were available for only three sampling sites in Burkina Faso, where three biotypes have been detected, the native Sub-Saharan Africa non-Silver Leafing (AnSL), the Sub-Saharan Africa Silverleafing (ASL), and the Mediterranean Q biotypes, but no information is available about their respective distributions on host plant species (Gueguen *et al.*, 2010).

2. Our study describes the biotypes and symbiotic bacterial communities of *B. tabaci* sampled in three West African countries, Burkina Faso, Benin, and Togo. A total of 527 individuals were collected from seven cultivated host plants.

3. In the 20 localities studied, we found the same three biotypes AnSL, ASL, and Q previously detected in Burkina Faso. These biotypes display a specific pattern of geographical distribution influenced by the host plant species. In Benin and Togo, the ASL and AnSL biotypes were predominant, while in Burkina Faso, the Q biotype was dominant, with two sub-groups, Q1 and Q3 (recorded to date only in this country), and ASL individuals found in sympatry with Q1 individuals in some localities. As previously reported, each biotype and each genetic group harbours a specific community of symbiotic bacteria.

**Key words.** *Bemisia tabaci*, biotypes, host plant, secondary symbionts, West Africa.

Correspondence: Olivier Gnankiné, Unité de Formation et de Recherche en Sciences de la Vie et de la Terre (UFR-SVT), Laboratoire d'Entomologie Fondamentale et Appliquée, Université de Ouagadougou, 03 BP 7021 Ouagadougou 03, Burkina Faso. E-mail: olgnankine@hotmail.com

\*These authors contributed equally.

### Introduction

The whitefly *Bemisia tabaci* (Gennadius) is a significant pest that damages many agricultural crops worldwide. *B. tabaci* is in fact a complex of cryptic species composed of 11 clearly differentiated, high-level groups containing at least 24 distinct

species composed of several low-level groups previously designated as 'biotypes' (De Barro *et al.*, 2010; Dinsdale *et al.*, 2010). Many aspects of this species complex remain unknown, such as the origins of the different species and the process of radiation, the borders between the species and their degree of genetic isolation, the geographical distribution, the biological characteristics and the within-species genetic diversity. *B. tabaci* biotypes can be distinguished using biochemical or molecular markers (allozymes, mitochondrial *Cytochrome Oxidase CO1*, nuclear *Internal Transcribed Spacer ITS*), but their biological significance and taxonomic status are still unclear. These biotypes can differ with regard to the induction of plant silverleafing disorders, host range, fecundity, insecticide resistance, ability to transmit plant viruses (Perring, 2001; De Barro *et al.*, 2010), and in the symbiotic bacterial community they harbour (Gueguen *et al.*, 2010). The situation is in fact even more complex, as significant genetic differentiation is also observed within biotypes. For example, within the Q biotype of the Mediterranean species, the genetic sub-groups designated Q1, Q2, and Q3 have been recognised both by *CO1* sequencing and from the composition of their endosymbiotic bacteria (McKenzie *et al.*, 2009; Gueguen *et al.*, 2010).

In tropical and sub-tropical countries, *B. tabaci* is a pest of primary importance, especially on crops of cassava, cotton, sweet potatoes, tobacco, and tomato. In West Africa, population outbreaks were observed in 1998 in cotton fields in Burkina Faso, Mali, and Ivory Coast inducing severe crop damages (Otoïdoba *et al.*, 2002, 2003). The resulting losses had a severe impact on the economic activity of these countries, as cotton production and trading is one of the main financial resources, especially in Burkina Faso. As a consequence, West African countries make wide use of many pesticides to control *B. tabaci* populations despite their harmful impact on potential natural enemies and on the environment. This practice also leads to high levels of resistance, which have been observed particularly in Burkina Faso (Houndété *et al.*, 2010). As *B. tabaci* biotypes have differing levels of insecticide resistance (Horowitz *et al.*, 2005; Wang *et al.*, 2010) and do not always share the same resistance mutation to a particular chemical class of insecticide (Alon *et al.*, 2006, 2008), pest management programs may inadvertently select for one of these biotypes locally. For instance, the polyphagous B biotype is known to be more susceptible to several chemical compounds than the Q biotype (Horowitz *et al.*, 2005), which explains why the Q biotype sometime displaces the B biotype (Horowitz *et al.*, 2005; Chu *et al.*, 2010; but see Crowder *et al.*, 2010). Another consequence is that the effectiveness of treatments may vary depending on the local composition of biotypes and the chemical classes of insecticides used. Identification of the biotypes present and information about their diversity are thus clearly required for effective control strategies of this pest.

Although major studies on *B. tabaci* have been performed in East Africa, little is known about the distribution and frequency of the various biotypes in West Africa. So far, three biotypes have been reported in this region: the Sub-Saharan African non-Silver Leafing (AnSL) biotype, the Sub-Saharan African

Silverleafing (ASL) biotype, and more recently, the Q biotypes (Berry *et al.*, 2004; Brown & Idris, 2005; Sseruwagi *et al.*, 2005; De la Rua *et al.*, 2006; Gueguen *et al.*, 2010). The AnSL biotype belongs to the Sub-Saharan African clade, and is considered to be the basal group of the species complex, and probably originated in Africa. AnSL has been reported throughout Africa, except in the northern countries near the Mediterranean basin (Berry *et al.*, 2004; Boykin *et al.*, 2007; Gueguen *et al.*, 2010). The closely related ASL and Q biotypes are genetically distant from AnSL. They both belong to the Mediterranean species within the Africa/Middle East/Asia Minor group (Boykin *et al.*, 2007; De Barro *et al.*, 2010). ASL has been reported in Ghana, Ivory Coast, Nigeria, Benin, Togo, and Burkina Faso (Brown & Idris, 2005; De la Rua *et al.*, 2006; Gueguen *et al.*, 2010). Recently, we reported that the Q biotype was present in Burkina Faso, where we detected two (Q1 and Q3) of the three genetic subgroups previously described (Gueguen *et al.*, 2010). However, there is a paucity of data for West Africa, and the precise geographical distribution, the frequency of the biotypes, and their genetic subdivisions, together with their possible association with symbiotic bacteria and host plant species remain largely unknown.

The present study set out to describe the distribution of the *B. tabaci* biotypes across West African countries, and its relationship with the host plant species. A total of 527 individuals collected from 20 localities in three countries (Burkina Faso, Benin and Togo), and from seven host plant species, were sampled during the rainy season in 2007 and 2009. In addition to identifying the biotype and genetic group, we determined the composition of the symbiotic bacteria community they harboured. Indeed, *B. tabaci*, like most phloem-feeding insects, is host to an obligatory primary endosymbiont, the bacterium *Portiera aleyrodidarum*, which is required to provide essential nutrients for its host's survival and development. *B. tabaci* is also infected by several facultative, vertically transmitted, symbiotic bacteria known as secondary endosymbionts (Zchori-Fein & Brown, 2002). Interestingly, each biotype or genetic group is infected by a specific symbiotic community (Chiel *et al.*, 2007; Gueguen *et al.*, 2010). This raises a question about the role of these symbiotic microorganisms in the biology of *B. tabaci*, and in the process of species radiation. Their phenotypic effects remain largely unknown in *B. tabaci*, but in other insect species, bacteria are known to be able to manipulate host reproduction (Stouthamer *et al.*, 1999; Weeks *et al.*, 2003), and to be involved in resistance to natural enemies (Oliver *et al.*, 2003) or host-plant specialisation (Tsuchida *et al.*, 2004). In *B. tabaci*, *Rickettsia* bacteria enhance susceptibility to insecticides (Kontsedalov *et al.*, 2008), but can also enhance whiteflies performance (Himler *et al.*, 2011). Moreover, symbiotic bacteria are thought to play a major role in the virus transmission capacities of the whitefly (Morin *et al.*, 1999; Gottlieb *et al.*, 2010).

The main objective of this study was to provide the first detailed description of the diversity of *B. tabaci* biotypes and their symbiotic bacteria communities in West Africa on several host plant species. This provides a good basis for answering questions about biotype dynamics in relation to insecticide resistance, as well as providing valuable information for developing integrated pest management programmes.

## Materials and methods

### Origin of the whiteflies

Sampling was performed in 2007 and 2009. Live adult *B. tabaci* (males and females) were collected and individually placed in 1.5-ml tubes containing 95% ethanol and kept at  $-20^{\circ}\text{C}$  until they underwent DNA extraction. The samples came from three neighbouring countries in West Africa: Burkina Faso, Benin, and Togo (Fig. 1). A total of 527 individuals were collected at 20 localities and from seven cultivated host plants (cotton, tomato, okra, tobacco, marrow, *Lantana camara*, cassava) (Table 1). At least eight individuals were analysed per locality and per plant. The cotton, tomato, marrow, and okra plantations had all been treated with insecticides (with the exception of the cotton from Kandi Sina in Benin) belonging to several classes: organophosphates, pyrethroids, and neonicotinoids. No treatment had been applied to the cassava and *Lantana camara* (Table 1).

### DNA extraction

Total DNA was extracted from each individual in 26  $\mu\text{l}$  of an extraction buffer containing 50 mM KCl, 10 mM Tris-base pH 8, 0.45% Nonidet P-40, 0.45% Tween 20 and 50  $\text{mg ml}^{-1}$  Proteinase K. Extraction buffer was added to the crude extract, incubated at  $65^{\circ}\text{C}$  for 3 h and then incubated at  $100^{\circ}\text{C}$  for 15 min. A volume of 35  $\mu\text{l}$  of pure water was added to this extract, which was then stored at  $-20^{\circ}\text{C}$  until use.

### Identification of *B. tabaci* biotypes and genetic subgroup

The biotypes were identified on the basis of the mitochondrial marker gene *COI* (*Cytochrome Oxidase I*) after a PCR reaction using the universal *COI* primers C1-J-2195 (5'-TTGATTT TTTGGTCATCCAGAAGT-3') and L2-N-3014 (5'-TCCAA TGCATAATCTGCCATATTA-3') (Khasdan *et al.*, 2005). PCR reactions were performed in 25- $\mu\text{l}$  volumes containing 200  $\mu\text{M}$  dNTP, 200 nM primers, 0.5 IU *Taq* DNA polymerase (Eurobio) and 2  $\mu\text{l}$  of DNA template. Cycling conditions consisted of an initial denaturing step at  $95^{\circ}\text{C}$  for 5 min, followed by 30 cycles of 1 min at  $95^{\circ}\text{C}$  for denaturing, 1 min at  $50^{\circ}\text{C}$  for annealing, 1 min at  $72^{\circ}\text{C}$  for extension, and a final extension of  $72^{\circ}\text{C}$  for 10 min (Tetrad®, Biorad, France). For some individuals (from 2 to 20 individuals per population), the PCR products were then sequenced to determine the biotype. For the others, we used the PCR-RFLP (PCR-Random Fragment Length Polymorphism) test developed and described by H. Henri, G. Terraz, O. Gnankiné, F. Fleury & L. Mouton (unpubl. data). Briefly, digestions of the *mtCOI* amplifications by the *XapI* and *BfmI* restriction enzymes result in different patterns of fragment migration depending on the *B. tabaci* genetic group. Products are visualised in 2% agarose gels containing ethidium bromide ( $0.5 \mu\text{g ml}^{-1}$ ) under UV illumination after a 1-h migration at 100 mV. This PCR-RFLP test can be used to identify all known genetic groups from Africa/Middle East/Asia Minor, and

major clades, including the Q1, Q2, and Q3 groups of the Mediterranean Q biotype.

### Detection and molecular identification of the secondary endosymbiotic bacteria (SS)

Individuals were screened for symbiotic bacterial infection using specific PCR primers that amplify the *16S rRNA* gene for *Hamiltonella*, *Cardinium*, and *Rickettsia*, the *23S rRNA* gene for *Arsenophonus* and *Fritschea* and the *wsp* gene for *Wolbachia* (Table 2). The presence of the obligatory symbiont *Portiera* was also checked to confirm the quality of DNA extraction. DNA was amplified in a final volume of 25  $\mu\text{l}$ , containing 1.5 mM  $\text{MgCl}_2$ , 200  $\mu\text{M}$  dNTPs, 200 nM of each primer, 0.5 IU *Taq* DNA polymerase (Eurobio) and 2  $\mu\text{l}$  of DNA. PCR products were visualised in 1% agarose gels stained with ethidium bromide ( $0.5 \mu\text{g ml}^{-1}$ ) under UV illumination to check for the presence/absence of the various bacteria.

### Phylogenetic analyses

Phylogenetic analyses of *B. tabaci* biotypes were performed on mitochondrial *COI* sequences. The accession numbers of the sequences obtained in the present study and used to construct the tree are listed in Table 3. Sequences of *B. tabaci* from Genbank were also included (Fig. 2). Multiple sequence alignment was carried out using MUSCLE software (Edgar, 2004). The appropriate model of evolution was estimated with jModeltest v 0.1.1 (Posada, 2008). The best likelihood score was evaluated with a corrected version of the Akaike information criterion for small samples (AICc). The model selected was HKY + G. Using this model, the phylogenetic tree was constructed by the maximum likelihood method with PhyML (version 3.0.1) (Guindon & Gascuel, 2003) and edited using Treedyn (Chevenet *et al.*, 2006). The robustness of the each node was assessed from 100 bootstrap replicates.

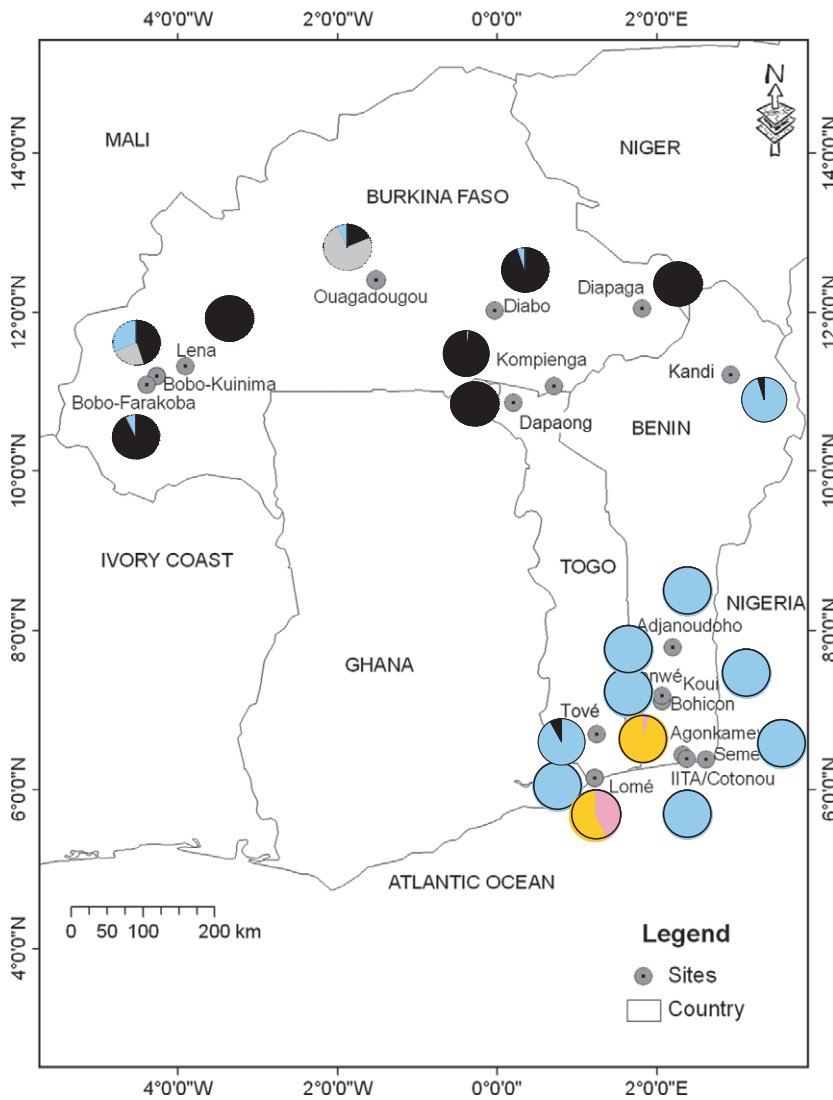
### Statistical analyses

The data collected were subjected to Fisher's exact test with simulated p-values based on 2000 replicates using R statistical software (<http://www.r-project.org/>).

## Results

### Geographical distribution of biotypes

Among the 527 individuals sampled from 20 localities in Burkina Faso, Togo, and Benin, three genetic groups recognised as biotypes were identified (Table 1): Q, ASL and the basal group AnSL. The findings revealed a clear distinction between the different countries (Fig. 1). In Togo and Benin, the African AnSL and the ASL biotypes of *B. tabaci* were predominant with two clearly differentiated genetic groups within the AnSL



**Fig. 1.** Geographical distributions of the samplings sites and of *Bemisia tabaci* biotypes.

**Biotype**

- Q1    ■ Q3    ■ ASL    ■ AnSL1    ■ AnSL2

biotype, designated AnSL1 and AnSL2 (Fig. 2), which correspond to the Sub Saharan Africa 1 and 2 genetic groups, respectively, described in Dinsdale *et al.* (2010) and Berry *et al.* (2004). These two sub-groups were found on cassava only and occurred together in sympatry (Table 1). However, ASL was the dominant biotype in this area (100% at 8 out of 11 sites). In 2009, the Q biotype was also recorded, but only in the northern parts of these countries. In Tové (Togo) and Kandi (Benin), this biotype was always observed in sympatry with ASL and at a low frequency, whereas at Dapaong (Togo), all individuals belonged to this biotype (Table 1; Fig. 1). The situation was quite different in Burkina Faso, where the Q biotype dominated at most of the sites sampled. The ASL biotype occurred at low frequency and always in sympatry with Q1 (Table 1; Fig. 1), and the AnSL biotype was never detected. However, it should be pointed out that no sample was available from cassava. Sequence analyses revealed the presence of two genetic

sub-groups of the Q biotype: Q1 and the newly described Q3 sub-group recently detected in West Africa (Gueguen *et al.*, 2010). Q1 and Q3 showed different distribution patterns in Burkina Faso in terms both of locality and host plant species. Q3 was only observed in the Western part of the country, and only on *L. camara* and tobacco, whereas Q1 was found at most sampling sites and on most host plants (Fig. 2). Finally, ASL was the only biotype recovered from all three countries, albeit at different frequencies.

*Distribution of biotypes on host plants*

As the samples were not always taken from the same host plants in all three countries, the plants sampled and the locality were not independent. This made it impossible to carry out a global statistical test to find out whether the host plant influences

**Table 1.** Frequencies of *Bemisia tabaci* biotypes related to host plants and localities.

Country	Acronym*	Locality	Host plant	Agricultural practice†	n‡	Year collected	Biotypes present (%)§
Burkina Faso	BF-Cott1a	Kompienga	Cotton	Treated	15	2007	Q1 (100)
	BF-Cott1b		Cotton	Treated	21	2009	Q1 (96), ASL (4)
	BF-Cott2a	Diapaga	Cotton	Treated	16	2007	Q1 (100)
	BF-Cott3a	Lena	Cotton	Treated	18	2007	Q1 (100)
	BF-Cott4a	Diabo	Cotton	Treated	21	2007	Q1 (95), ASL (5)
	BF-Cott4b		Cotton	Treated	20	2009	Q1 (96), ASL (4)
	BF-Tom5a	Bobo/Kuinima	Tomato	Treated	33	2007	Q1 (42), ASL (58)
	BF-Mar5a		Marrow	Treated	31	2007	Q1 (65), ASL (35)
	BF-Okr5a		Okra	Treated	15	2007	Q1 (87), ASL (13)
	BF-Tob5a		Tobacco	Untreated	30	2007	Q3 (86), Q1 (7), ASL (7)
	BF-Cott6a	Bobo/Farako-Bâ	Cotton	Treated	15	2007	Q1 (93), ASL (7)
	BF-Cam7a	Ouagadougou	<i>L.camara</i>	Untreated	44	2007	Q3 (100)
	BF-Cam7b		<i>L.camara</i>	Untreated	20	2009	Q3 (100)
	BF-Tom7a		Tomato	Treated	15	2007	Q1 (73), ASL (27)
Benin	Be-Cott1a	Tanwé	Cotton	Treated	10	2007	ASL (100)
	Be-Cott2a	Koui	Cotton	Treated	18	2007	ASL (100)
	Be-Cott3a	Adjanoudoho	Cotton	Treated	9	2007	ASL (100)
	Be-Tom4a	IITA	Tomato	Treated	10	2007	ASL (100)
	Be-Cott5a	CRA-CF/Bohicon	Cotton	Treated	10	2007	ASL (100)
	Be-Cott5b		Cotton	Treated	11	2009	ASL (100)
	Be-Tom6a	Sémé	Tomato	Treated	16	2007	ASL (100)
	Be-Cas7a	Agonkanmey	Cassava	Untreated	20	2007	AnSL1 (4), AnSL2 (96)
	Be-Cott8b	Kandi Sina	Cotton	Untreated	21	2009	Q1 (4.8), ASL (95.2)
Be-Cott9b	Kandi/Gamassosso	Cotton	Treated	18	2009	Q1 (5.5), ASL (94.5)	
Togo	To-Tom1a	Lome Ferme/UL	Tomato	Treated	13	2007	ASL (100)
	To-Cas2a	Lome Harbour	Cassava	Untreated	20	2007	AnSL1 (42), AnSL2 (58)
	To-Cott3b	Dapaong	Cotton	Treated	29	2009	Q1 (100)
	To-Cott4b	Tové	Cotton	Treated	8	2009	Q1 (8), ASL (92)

\*The first two letters of the acronyms correspond to the country (BF, Burkina Faso; Be, Benin; To, Togo), the second part indicates the host plant and the number refers to a specific locality. The a and b samplings were carried out in 2007 and 2009 respectively.

†This column indicates whether pesticides are used or not.

‡Number of individuals in the sample.

§Percentage of each biotype found at the collecting site.

**Table 2.** PCR primers and conditions used in the study.

Symbiont	Targeted gene	Primers	Primer sequences	Annealing temperature (°C)/ product size (bp)	References
<i>Portiera</i>	<i>ADNr16S</i>	28F1098 R	5'-AGAGTTTGATMTGGTCAG-3' 5'-AAAGTTCCCGCCTTATGCGT-3'	58 / ~1000	Zchori-Fein and Brown (2002)
<i>Rickettsia</i>	<i>ADNr16S</i>	Rb-FRb-R	5'-GCTCAGAACGAACGCTATC-3' 5'-GAAGGAAAAGCATCTCTGC-3'	58 / ~900	Gottlieb <i>et al.</i> (2006)
<i>Hamiltonella</i>	<i>ADNr16S</i>	Hb-FHb-R	5'-TGAGTAAAGTCTGGAATCTGG-3' 5'-AGTTCAAGACCGCAACCTC-3'	58 / ~700	Zchori-Fein and Brown (2002)
<i>Cardinium</i>	<i>ADNr16S</i>	CFB-FCFB-R	5'-GCGGTGTAATAAGTGAAGAAAC-3' 5'-ACCTMTTCTTAACTCAAGCCT-3'	56 / ~400	Weeks <i>et al.</i> (2003)
<i>Wolbachia</i>	<i>wsp</i>	81F415R	5'-TGGTCCAATAAGTGAAGAAAC-3' 5'-AAAAATTAACGCTACTCCA-3'	56 / ~700	Braig <i>et al.</i> (1998), Vautrin (2008)
<i>Arsenophonus</i>	<i>ADNr23S</i>	Ars-23S-1 Ars-23S-2	5'-CGTTTGATGAATTCATAGTCAAA-3' 5'-GGTCTCCAGTTAGTGTTACCCAAC-3'	60 / ~600	Thao and Baumann (2004)
<i>Fritschea bemisiae</i>	<i>ADNr23S</i>	Frit-FFrit-R	5'-TGGTCCAATAAGTGAAGAAAC-3' 5'-GCTCGGTACCCTTTAAATGGCG-3'	60 / ~600	Thao <i>et al.</i> (2003)

© 2012 The Authors

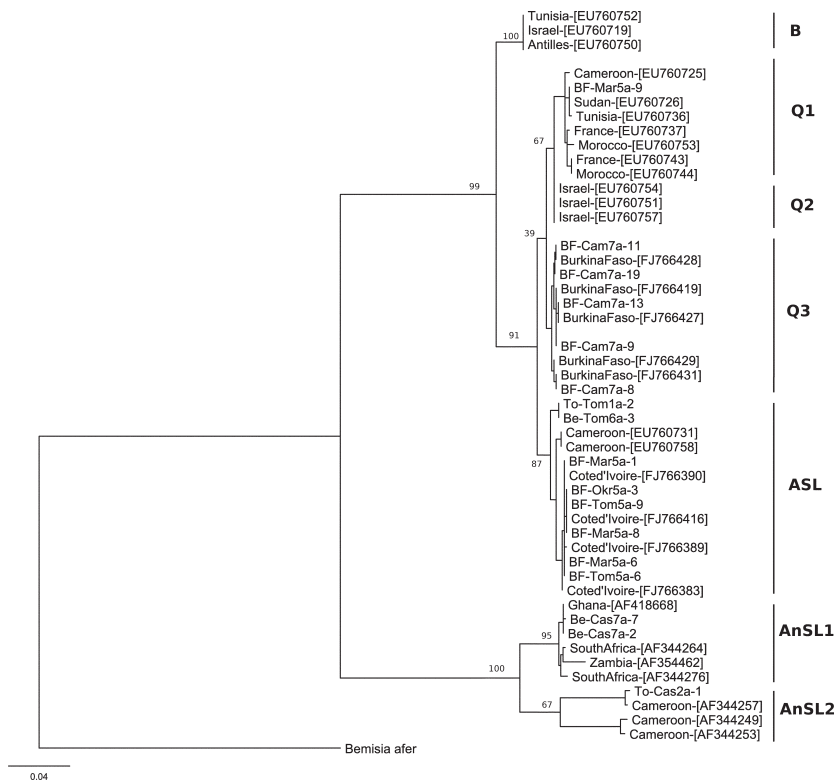
Insect Conservation and Diversity © 2012 The Royal Entomological Society, *Insect Conservation and Diversity*, 6, 411–421



**Table 3.** Names and accession numbers of samples used for the phylogenetic tree.

Sample	Locality	Country	Host plant	Biotype	Accession number
BF-Tom5a-1	Bobo/Kuinima	Burkina Faso	Tomato	ASL	HQ908636
BF-Tom5a-9	Bobo/Kuinima	Burkina Faso	Tomato	ASL	HQ908637
BF-Mar5a-1	Bobo/Kuinima	Burkina Faso	Marrow	ASL	HQ908638
BF-Mar5a-6	Bobo/Kuinima	Burkina Faso	Marrow	ASL	HQ908639
BF-Mar5a-8	Bobo/Kuinima	Burkina Faso	Marrow	ASL	HQ908640
BF-Okr5a-3	Bobo/Kuinima	Burkina Faso	Okra	ASL	HQ908641
BF-Cam7a-8	Ouagadougou	Burkina Faso	<i>L. camara</i>	Q3	HQ908642
BF-Cam7a-9	Ouagadougou	Burkina Faso	<i>L. camara</i>	Q3	HQ908643
BF-Cam7a-11	Ouagadougou	Burkina Faso	<i>L. camara</i>	Q3	HQ908644
BF-Cam7a-19	Ouagadougou	Burkina Faso	<i>L. camara</i>	Q3	HQ908645
BF-Cam7a-13	Ouagadougou	Burkina Faso	<i>L. camara</i>	Q3	HQ908646
Be-Tom6a-3	Sémé	Benin	Tomato	ASL	HQ908647
Be-Cas7a-2	Agonkanmey	Benin	Cassava	AnSL1	HQ908648
Be-Cas7a-7	Agonkanmey	Benin	Cassava	AnSL1	HQ908649
To-Tom1a-2	Ferme/UL	Togo	Tomato	ASL	HQ908650
To-Cas2a-1	Harbour zone	Togo	Cassava	AnSL2	HQ908651

AnSL, Sub-Saharan Africa non-Silver Leafing; ASL, Sub-Saharan Africa Silverleafing; *Lantana camara*, *L. camara*.



**Fig. 2.** *COI*-gene-based phylogeny of *Bemisia tabaci*.

the distribution of the biotypes. However, it did look as though there was a strict association between the AnSL biotype and cassava, no other biotype having ever been found on this plant, even when ASL was present in the same locality (Table 1). AnSL was never found on cotton crops, regardless of whether they had undergone insecticide treatment or not. Similarly, only the Q3 group was found on *L. camara*. Although Q3 individuals were also observed on tobacco, it was only on *L. camara* that

this biotype was not found in sympatry with other biotypes. Importantly, these two plants were the only ones that had not been treated with insecticides in our Burkina Faso samples, raising the possibility that the distribution of Q3 is mainly determined by insecticide use. In Bobo/Kuinima (Burkina Faso), where four plant species had been sampled in 2007, a significant relationship was found between the plant and the biotype distribution (Fisher's Exact Test,  $P = 0.0005$ ). In particular, Q3 was



only found on tobacco, and was strongly predominant on this plant. Observation of biotype distribution among the host plants also suggests an association between the Q1 biotype and cotton, since when it is present Q1 reaches very high frequencies of up to 100% on this plant. Once again, it is possible that this pattern was mediated by insecticide use.

#### *Prevalence, diversity, and distribution of endosymbiotic bacteria*

The presence of the seven known symbiotic bacteria infecting *B. tabaci* (the primary symbiont *Portiera* and the six secondary symbionts: *Hamiltonella*, *Arsenophonus*, *Rickettsia*, *Cardinium*, *Wolbachia*, *Fritschea*) was checked by specific diagnostic PCR. As expected, all individuals were found to be infected with *Portiera*, which demonstrated the high quality of the DNA extractions. Figure 3 shows the infection status of individuals from all 20 localities and from different host plants with regard to all the secondary endosymbionts except *Fritschea*, which was never detected. As previously reported (Chiel *et al.*, 2007; Gueguen *et al.*, 2010), all genetic groups (AnSL1, AnSL2, ASL, Q1, Q3) were found to host a specific bacterial assemblage (Fisher's Exact Test,  $P = 0.0005$ ): the most common bacteria being *Hamiltonella* for the Q1 biotype, *Arsenophonus* for ASL and AnSL biotypes, and the association *Arsenophonus/Rickettsia* for the Q3 group. However, the infection status of individuals belonging to the ASL biotype is variable, with a strong contrast being observed between Burkina Faso and Togo/Benin (Fisher's Exact Test,  $P = 5 \times 10^{-15}$ ). Indeed, while in Burkina Faso most of the individuals are infected by *Arsenophonus* (80% on average), the level of infection with this secondary symbiont was low in Togo and Benin (less than 12%), and it was not detected at all on some populations in Benin (Fig. 3). Here, we report for the first time the infection status of the basal clade AnSL, which appears to be very similar to that of the ASL biotype, having *Arsenophonus* as the main symbiotic bacterium and a high frequency of uninfected individuals.

In contrast to the findings of a previous study (Gueguen *et al.*, 2010), multiple infections with several secondary symbionts were uncommon in the areas we sampled (13.7%; Fig. 3). In most cases, they involved infection with two species (95.8%) and were mainly observed within the Q3 biotype co-infected by *Rickettsia* and *Arsenophonus*.

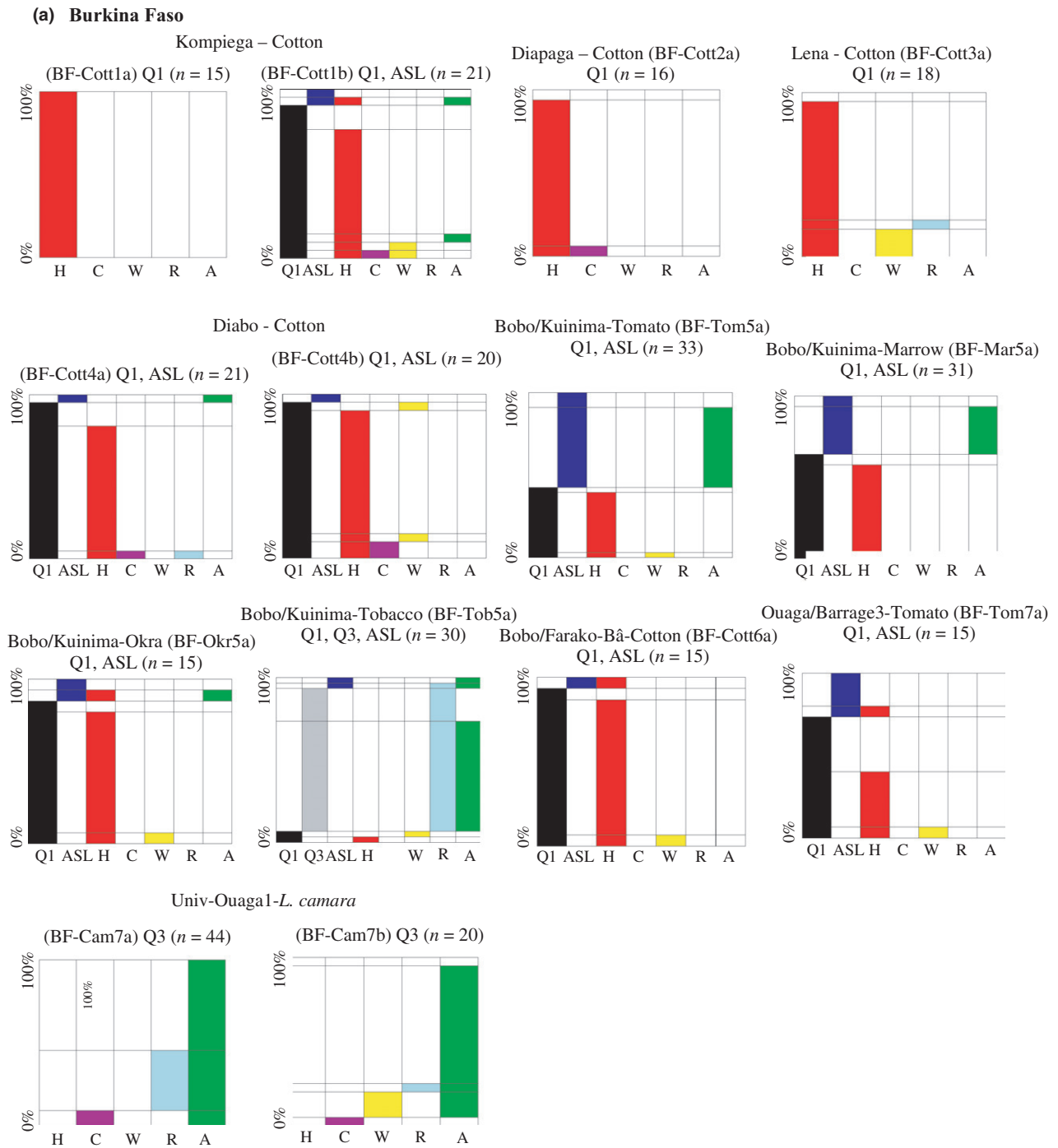
#### Discussion

This first detailed analysis of *B. tabaci* biotypes in West African countries provides valuable information about the level of regional biodiversity, including cytoplasmic biodiversity due to infection by symbiotic bacteria related to different host plant species and insecticide treatments. It also extends our knowledge about the geographical range of major biotypes that have spread throughout the world, such as the Q biotype. We identified the same biotypes as had previously been reported in Burkina Faso by Gueguen *et al.* (2010), i.e. the ancestral AnSL (Sub-Saharan Africa non-Silver Leafing), the ASL (SubSaha-

ran Africa Silver Leafing) and the Q biotypes. The frequencies of these biotypes showed marked differences depending on the site sampled. One central finding is that the Q biotype was present at a high frequency in Burkina Faso in 2007 and 2009, whereas it was only detected in the northern parts of Togo and Benin in 2009, which could be explained by a recent introduction into these latter countries of agricultural practices that facilitate its ability to spread. In addition to biotype diversity, the *COI* mitochondrial marker revealed the existence of sub-groups within the AnSL and Q biotypes (designated AnSL1, AnSL2 and Q1, Q3 respectively). The detection of diversity at such a small geographical scale raises numerous questions about the evolution of biotypes, the coexistence and distribution of sub-groups, and the consequences of this diversity for *B. tabaci* management.

These three biotypes (ASL, AnSL, and Q) have also been detected in East Africa, including five sub-groups of the AnSL biotype, but only one, the Q1 sub-group, of biotype Q (Berry *et al.*, 2004; Brown & Idris, 2005; Sseruwagi *et al.*, 2005; De la Rua *et al.*, 2006; Gueguen *et al.*, 2010). Sub-group Q3 had only been detected in Burkina Faso, but this could have been due to a sampling bias, as Q3 appears to be restricted to certain plants (tobacco and *L. camara*). In both East and West Africa, the Q biotype has been found at higher latitudes than the ASL and AnSL biotypes. In northern African countries, such as Morocco and Tunisia, the Q1 biotype is predominant (Gueguen *et al.*, 2010), and a similar trend can also be seen in the present study (Fig. 1). In East Africa, the Q biotype has only been detected in Sudan, which is in the North/Centre region of Africa, and in the Uganda, which borders it to the south. In countries further south (Tanzania, South Africa, Mozambique), AnSL is widespread. However, Cameroon, Benin, and Togo, where we almost exclusively find the AnSL biotype, are at a higher latitude than Uganda, so latitude alone cannot entirely explain the pattern of distribution, and longitude must also be taken into account.

Interestingly, in Burkina Faso, ASL was always found in sympatry with the Q biotype and on the same host plant, particularly the Q1 sub-group, suggesting the possible coexistence of the ASL and Q biotypes at least on a short time scale. However, on all these plants, Q1 was predominant. On cotton, ASL was rarely observed and when present, it never reached frequencies of more than 7%. This situation could indicate competitive interactions between ASL and Q in Burkina Faso, with the outcome of competition depending on environmental conditions, such as host plant species or agricultural practice. Indeed, resistance to neonicotinoids has been shown in this country to be correlated with the presence of the Q1 biotype, with ASL individuals being most susceptible (O. Gnankiné, L. Mouton, A. Savadogo, T. Martin, A. Sanon, F. Vavre, R. Dabire & F. Fleury, unpubl. data). More generally, the Q biotype is known to be highly resistant to several insecticides (Horowitz *et al.*, 2005), which could give it a competitive advantage when insecticides are used. This situation had already been observed among the B and Q biotypes, the latter excluding the former in the context of insecticide treatment (Chu *et al.*, 2010; Crowder *et al.*, 2010). The same phenomenon may occur in Western Africa among resistant Q and susceptible ASL biotypes, explaining



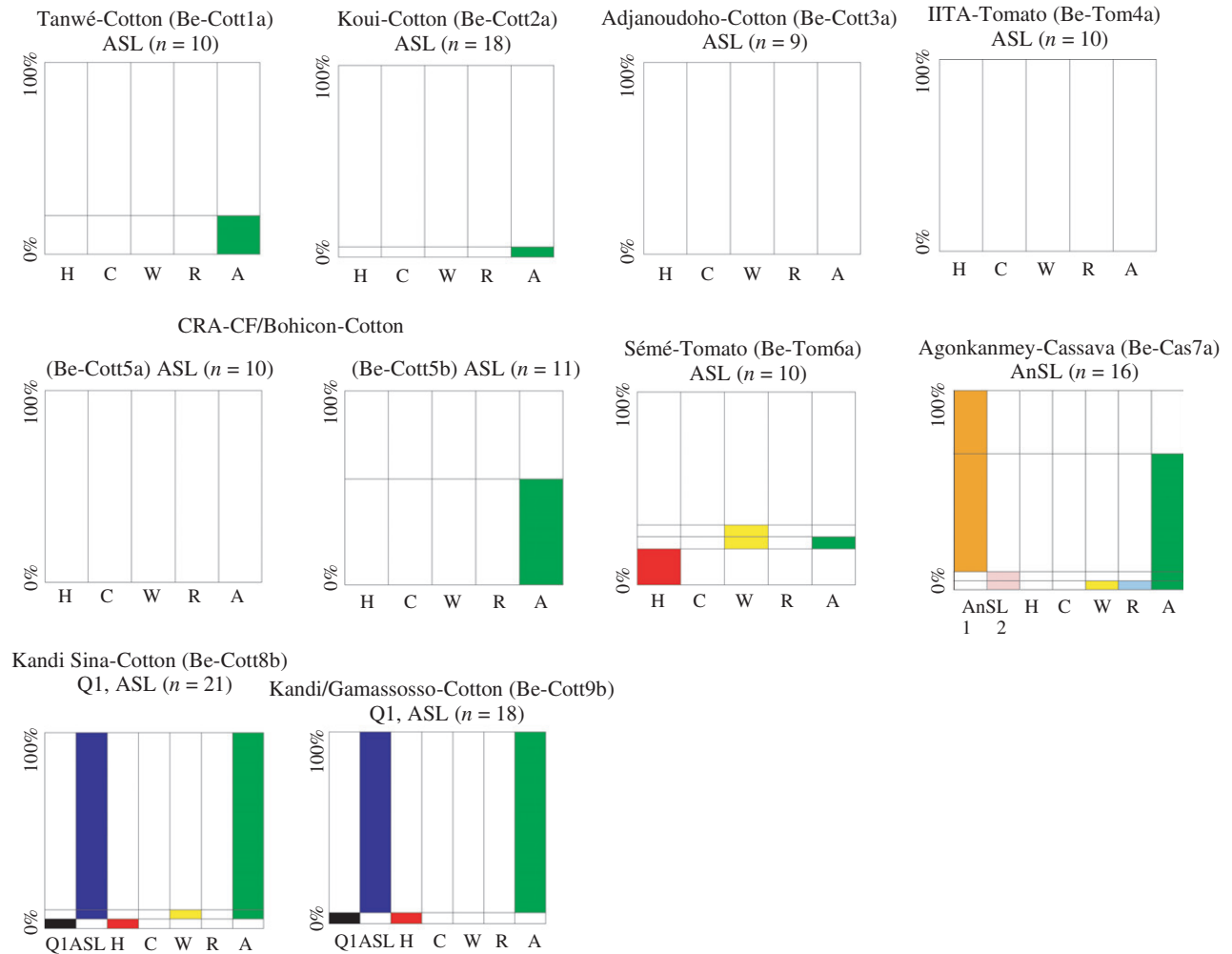
**Fig. 3.** Infection status of *Bemisia tabaci* individuals collected in 16 localities of three Western Africa countries.

why the sensitivity of adult populations of *B. tabaci* can vary greatly from one place and 1 year to another in Burkina Faso (Gnankiné *et al.*, 2007). However, in the localities where we sampled twice, in 2007 and in 2009, the prevalence of these biotypes was similar. This time lapse may have been too short, and other records on the spatial distribution of biotypes in Burkina Faso and in bordering countries spanning several years are

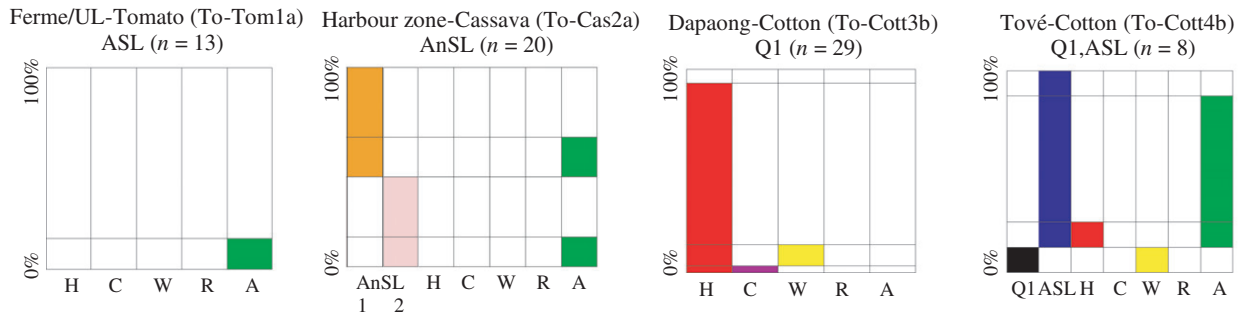
required before we can determine whether the Q biotype is actually spreading, and whether the sympatric situation of ASL and Q is stable or whether one biotype will eventually exclude the other, which could have major repercussions for pest management and agricultural production in this region.

The stable co-existence of biotypes could involve host-plant specialisation; for example, AnSL was the only biotype found

**(b) Benin**



**(c) Togo**



**Fig. 3.** (Continued).

on cassava. An earlier survey in sub-Saharan Africa (Abdullahi *et al.*, 2003) had already highlighted the fact that the populations that develop on cassava seem to be restricted to cassava only, whereas populations from other plants are polyphagous, but do not colonise cassava. Such niche specialisation of the AnSL, with cassava exercising a repulsive action on other biotypes,

could allow it to persist even if other biotypes are present locally. Unfortunately, all our samples collected on cassava originated from Togo and Benin, where ASL was the predominant biotype. Sampling on cassava in Burkina Faso could reveal whether AnSL is also found in this country despite the presence of Q1 and Q3. Q3 also seems to develop specifically in one host plant,

*L. camara*, where no other biotype has been detected, which suggests that even within the Q biotypes, there can be significant ecological and biological differences between sub-groups. However, this effect could also be due to the fact that this plant did not undergo insecticide treatment, something also suggested by the presence of Q3 on untreated tobacco. Further investigations of the ability of Q1, ASL and Q3 to develop on these different plants and on their resistance to insecticides are required to clarify and allow us to predict the dynamics of these biotypes and sub-groups.

Our study also confirms the initial observations that there is a link between symbiotic bacterial communities and the biotypes or genetic groups of *B. tabaci* (Chiel *et al.*, 2007; Gueguen *et al.*, 2010). As previously observed, within the Q biotype, most Q1 individuals harbour *Hamiltonella* and sometimes low frequencies of *Wolbachia* or/and *Cardinium*, whereas Q3 individuals were mostly found to harbour *Arsenophonus* with a high level of co-infection with *Rickettsia*. In the ASL and AnSL biotypes, *Arsenophonus* is the most common bacterium with higher frequency in the AnSL than the ASL biotype. In very few cases, other bacteria were reported such as *Wolbachia*, *Rickettsia*, or *Hamiltonella*. The reasons for these biotype-symbiont specific associations remain unknown, and the first step towards understanding the mechanisms involved will be to determine the phenotypic effects of these bacteria in *B. tabaci*. In aphids, some secondary symbionts are known to be involved in ecological specialisation, such as host plant utilisation (Tsuchida *et al.*, 2004), and so the symbiotic community could play a major role in biotype adaptation and competition.

One noteworthy finding is the fact that the prevalence of secondary symbionts varied considerably between the different localities and genetic groups. Most individuals from Burkina Faso belonged to the ASL and Q biotypes, and displayed high levels of infection, while in Benin and Togo, where the Q biotype is rare, the prevalence of secondary symbionts in ASL, consisting mainly *Arsenophonus*, was very low. Moreover, in Benin, some populations did not harbour any of the 6 secondary symbionts we checked. This is the first description of *B. tabaci* populations in which no secondary symbionts have been recorded. In addition, only a small number of individuals were infected by more than one secondary symbiont, with the exception of individuals belonging to the Q3 biotype, which contrasts with a previous study in which a high level of multiple infections was found (Gueguen *et al.*, 2010). However, it is impossible to rule out the possibility that other bacterial symbionts were in fact present in some *B. tabaci* individuals, and further studies involving more detailed investigations of the symbiotic content are called for. However, despite these limitations, the data reported raise several questions about the dynamics of infection by secondary symbionts and the influence of the environment on symbiotic composition.

## Acknowledgments

The authors thank Agence Universitaire de la Francophonie (AUF) for the Post-Doctoral fellowship of Olivier Gnankiné and the collaboration established between Ouagadougou Uni-

versity (Burkina Faso) and Lyon1 University (France). This research was supported by Agence Nationale de la Recherche (ANR).

## References

- Abdullahi, I., Winter, S., Atiri, G.I. & Thottappilly, G. (2003) Molecular characterization of whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae) populations infesting cassava. *Bulletin of Entomological Research*, **93**, 97–106.
- Alon, M., Alon, F., Nauen, R. & Morin, S. (2008) Organophosphates' resistance in the B-biotype of *Bemisia tabaci* (Hemiptera: Aleyrodidae) is associated with a point mutation in an acetylcholinesterase and overexpression of carboxylesterase. *Insect Biochemistry and Molecular Biology*, **38**, 940–949.
- Alon, M., Benting, J., Lueke, B., Ponge, T., Alon, F. & Morin, S. (2006) Multiple origins of pyrethroid resistance in sympatric biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Insect Biochemistry and Molecular Biology*, **36**, 71–79.
- Berry, S.D., Fondong, V.N., Rey, C., Rogan, D., Fauquet, C.M. & Brown, J.K. (2004a) Molecular evidence for five distinct *Bemisia tabaci* (Homoptera: Aleyrodidae) geographic haplotypes associated with cassava plants in sub-Saharan Africa. *Annals of Entomological Society of America*, **97**, 852–859.
- Boykin, L.M., Shatters, R.G., Rosell, R.C., McKenzie, C.L., Bagnall, R.A., De Barro, P. & Frohlich, D.R. (2007) Global relationships of *Bemisia tabaci* (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Molecular Phylogenetics and Evolution*, **44**, 1306–1319.
- Braig, H.R., Zhou, W., Dobson, S.L. & O'Neill, S.L. (1998) Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. *Journal of Bacteriology*, **180**, 2373–2378.
- Brown, J.K. & Idris, A.M. (2005) Genetic differentiation of whitefly *Bemisia tabaci* mitochondrial cytochrome oxidase I, and phylogeographic concordance with the coat protein of the plant virus genus Begomovirus. *Annals of Entomological Society of America*, **98**, 827–837.
- Chevenet, F., Brun, C., Bañuls, A.L., Jacq, B. & Christen, R. (2006) TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics*, **7**, 439.
- Chiel, E., Gottlieb, Y., Zchori-Fein, E., Mozes-Daube, N., Katzir, N., Inbar, M. & Ghanim, M. (2007) Biotype-dependent secondary symbiont communities in sympatric populations of *Bemisia tabaci*. *Bulletin of Entomological Research*, **97**, 407–413.
- Chu, D., Wan, F.H., Zhang, Y.J. & Brown, J.K. (2010) Change in the biotype composition of *Bemisia tabaci* in Shandong Province of China from 2005 to 2008. *Environmental Entomology*, **39**, 1028–1036.
- Crowder, D.W., Horowitz, A.R., De Barro, P.J., Liu, S.S., Showalter, A.M., Kontsedalov, S., Khasdan, V., Shargal, A., Liu, J. & Carrière, Y. (2010) Mating behaviour, life history and adaptation to insecticides determine species exclusion between whiteflies. *Journal of Animal Ecology*, **79**, 563–570.
- De Barro, P.J., Liu, S.-S., Boykin, L.M. & Dinsdale, B. (2010) *Bemisia tabaci*: a statement of species status. *Annual Review of Entomology*, **56**, 1–19.
- De la Rua, P., Simon, B., Cifuentes, D., Martinez-Mora, C. & Cenis, J.L. (2006) New insights into the mitochondrial phylogeny of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in the Mediterranean Basin. *Journal of Zoological Systematics and Evolutionary Research*, **44**, 25–33.



- Dinsdale, A., Cook, L., Riginos, C., Buckley, Y.M. & De Barro, P. (2010) Refined global analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodidae: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. *Annals of the Entomological Society of America*, **103**, 196–208.
- Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acid Research*, **32**, 1792–1797.
- Gnankiné, O., Traoré, D., Sanon, A., Traoré, N.S. & Ouedraogo, A.P. (2007) Traitements insecticides et dynamique des populations de *Bemisia tabaci* en culture cotonnière au Burkina Faso. *Cahiers Agricultures*, **16**, 101–109.
- Gottlieb, Y., Ghanim, M., Chiel, E., Gerling, D., Portnoy, V. & Steinberg, S. (2006) Identification and localization of a *Rickettsia* sp. in *Bemisia tabaci* (Homoptera: Aleyrodidae). *Applied and Environmental Microbiology*, **72**, 3646–3652.
- Gottlieb, Y., Zchori-Fein, E., Mozes-Daube, N., Kontsedalov, S., Skaljic, M., Brumin, M., Sobol, I., Czosnek, H., Vavre, F., Fleury, F. & Ghanim, M. (2010) The transmission efficiency of tomato yellow leaf curl virus by the whitefly *Bemisia tabaci* is correlated with the presence of a specific symbiotic bacterium species. *Journal of Virology*, **84**, 9310–9317.
- Gueguen, G., Vavre, F., Gnankine, O., Peterschmitt, M., Charif, D., Chiel, E., Gottlieb, Y., Ghanim, M., Zchori-Fein, E. & Fleury, F. (2010) Endosymbiont metacommunities, mtDNA diversity and the evolution of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. *Molecular Ecology*, **19**, 4365–4378.
- Guindon, S. & Gascuel, O. (2003) A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.
- Himler, A.G., Adachi-Hagimori, T., Bergen, J.E., Kozuch, A., Kelly, S.E., Tabashnik, B.E., Chiel, E., Duckworth, V.E., Dennehy, T.J., Zchori-Fein, E. & Hunter, M.S. (2011) Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. *Science*, **332**, 254–256.
- Horowitz, A.R., Kontsedalov, S., Khasdan, V. & Ishaaya, I. (2005) Biotypes B and Q of *Bemisia tabaci* and their relevance to neonicotinoid and pyriproxyfen resistance. *Archives of Insect Biochemistry and Physiology*, **58**, 216–225.
- Houndété, T.A., Kétoh, G.K., Hema, O.S.A., Brévaukt, T., Glitho, I.A. & Martin, T. (2010) Insecticide resistance in field populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in West Africa. *Pest Management Science*, **11**, 1181–1185.
- Khasdan, V., Levin, I., Rosner, A., Morin, S., Kontsedalov, S., Maslenin, L. & Horowitz, A.R. (2005) DNA markers for identifying biotypes B and Q of *Bemisia tabaci* (Hemiptera: Aleyrodidae) and studying population dynamics. *Bulletin of Entomological Research*, **95**, 605–613.
- Kontsedalov, S., Zchori-Fein, E., Chiel, E., Gottlieb, Y., Inbar, M. & Ghanim, M. (2008) The presence of *Rickettsia* is associated with increased susceptibility of *Bemisia tabaci* (Homoptera: Aleyrodidae) to insecticides. *Pest Management Science*, **68**, 789–792.
- McKenzie, C.L., Hodges, G., Osborne, L.S., Byrne, F.J. & Shatters, R.G. Jr (2009) Distribution of *Bemisia tabaci* (Hemiptera: Aleyrodidae) biotypes in Florida—investigating the Q invasion. *Journal of Economic Entomology*, **102**, 670–676.
- Morin, S., Ghanim, M., Zeidan, M., Czosnek, H., Verbeek, M. & van den Heuvel, J.F. (1999) A GroEL homologue from endosymbiotic bacteria of the whitefly *Bemisia tabaci* is implicated in the circulative transmission of tomato yellow leaf curl virus. *Virology*, **256**, 75–84.
- Oliver, K.M., Russell, J.A., Moran, A.N. & Hunter, M.S. (2003) Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the National Academy of Sciences USA*, **100**, 1803–1807.
- Otoïdobia, L.C., Vincent, C. & Stewart, R.K. (2002) Susceptibility of field populations of adult *Bemisia tabaci* (Homoptera: Aleyrodidae) and *Eretmocerus* sp. (Hymenoptera: Aphelinidae) on the population dynamics of the pest in Burkina Faso (West Africa). *Pest Management Science*, **59**, 97–106.
- Otoïdobia, L.C., Vincent, C. & Stewart, R.K. (2003) Field efficacy and baseline toxicities of pyriproxyfen, acetamiprid and diafenthiuron *Bemisia tabaci* Gennadius (Homoptera: Aleyrodidae) in Burkina Faso (West Africa). *Journal of Environmental Science and Health B*, **38**, 757–769.
- Perrin, T.M. (2001) The *Bemisia tabaci* species complex. *Crop Protection*, **20**, 725–737.
- Posada, D. (2008) JmodelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.
- Sseruwagi, P., Legg, J.P., Maruthi, M.N., Colvin, J., Rey, M.E.C. & Brown, J.K. (2005) Genetic diversity of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) populations and presence of the B biotype and a non-B biotype that can induce silverleaf symptoms in squash, in Uganda. *Annals of Applied Biology*, **147**, 253–265.
- Stouthamer, R., Breeuwer, J.A.J. & Hurst, G.D.D. (1999) *Wolbachia pipiensis*: Microbial manipulator of Arthropod reproduction. *Annual Review of Microbiology*, **53**, 71–102.
- Thao, M.L. & Baumann, P. (2004) Evidence for multiple acquisition of *Arsenophonus* by whitefly species (Sternorrhyncha: Aleyrodidae). *Current Microbiology*, **48**, 140–144.
- Thao, M.L., Baumann, L., Hess, J.M., Falk, B.W., Ng, J.C., Gullan, P.J. & Baumann, P. (2003) Phylogenetic evidence for two new insect-associated chlamydia of the family Simkaniaceae. *Current Microbiology*, **47**, 46–50.
- Tsuchida, T., Koga, R. & Fukatsu, T. (2004) Host plant specialization governed by facultative symbiont. *Science*, **303**, 1989.
- Vautrin, E. (2008) Les infections multiples dans les symbioses insectes/Bactéries: évolution intra-Hôte et conséquences populationnelles. PhD thesis, Université de Lyon 1, France.
- Wang, Z., Yan, H., Yang, Y. & Wu, Y. (2010) Biotype and insecticide resistance status of the whitefly *Bemisia tabaci* from China. *Pest Management Science*, **66**, 1360–1366.
- Weeks, A.R., Velten, R. & Stouthamer, R. (2003) Incidence of a new sex-ratio-distorting endosymbiotic bacterium among arthropods. *Proceedings of the Royal Society Series B*, **270**, 1857–1865.
- Zchori-Fein, E. & Brown, J.K. (2002) Diversity of prokaryotes associated with *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae). *Annals of the Entomological Society of America*, **95**, 711–718.

Accepted 2 March 2012

First published online 08 May 2012

Editor: Koos Boomsma

Associate editor: Martijn Bezemer





## Molecular characterization of genetic diversity within the Africa/Middle East/Asia Minor and Sub-Saharan African groups of the *Bemisia tabaci* species complex

Hélène Henri<sup>a</sup>, Gabriel Terraz<sup>a</sup>, Olivier Gnankiné<sup>b</sup>, Frédéric Fleury<sup>a</sup> and Laurence Mouton<sup>a\*</sup>

<sup>a</sup>Laboratoire de Biométrie et Biologie Evolutive, Université Lyon 1, Université de Lyon, UMR CNRS 5558, Villeurbanne Cedex, France;

<sup>b</sup>Laboratoire d'Entomologie Fondamentale et Appliquée, Unité de Formation et de Recherche en Sciences de la Vie et de la Terre (UFR-SVT), Université de Ouagadougou, Ouagadougou, Burkina Faso

(Received 16 July 2013; final version received 21 November 2013)

*Bemisia tabaci* Gennadius is a major plant pest in many agricultural systems worldwide. It is a complex of cryptic species that differ in many ecological respects, including damage-causing potential. Monitoring the genetic composition of *B. tabaci* populations in the field is, therefore, essential for achieving effective control. Sequencing is costly and time consuming; thus, efficient diagnostic tools must be developed to perform epidemiological studies involving hundreds of individuals. Here, we describe a polymerase chain reaction (PCR) restriction fragment length polymorphism (RFLP) method for identifying all the putative species comprising the Africa/Middle East/Asia Minor group, including those that are the most devastating crop pests. Moreover, intra-specific diversity associated with specific symbiotic bacteria known to manipulate insect host phenotypes can also be detected. In addition, this method discriminates between the Africa/Middle East/Asia Minor and the Sub-Saharan African groups, which are sympatric in some areas. This simple, reliable and cost-effective diagnostic tool is ideal for the rapid analysis of a high number of individuals and, thus, has potential applications in field contexts, where it could provide valuable baseline information for pest management programs.

**Keywords:** PCR-RFLP; *Bemisia tabaci*; *mtCOI* gene; molecular diagnostic; Sub-Saharan African group; MEAM1 species; Mediterranean species; pest management program

### 1. Introduction

The whitefly *Bemisia tabaci* Gennadius is one of the most important pests worldwide, causing significant damage to food and ornamental crops directly by feeding on sap and, to an even greater extent, indirectly by transmitting many plant-pathogenic viruses (Jones 2003). *B. tabaci* displays considerable genetic and biological variability despite its morphological homogeneity, which has led to the conclusion that it is in fact a complex of morphologically indistinguishable “cryptic” species (Perring 2001; De Barro et al. 2011). The genetic complexity of *B. tabaci* has mainly been investigated using mitochondrial cytochrome oxidase gene (*mtCOI*) sequence data (Frohlich et al. 1999; Boykin et al. 2007). This has led to the identification of several genetic groups known as biotypes, but the use of the term “biotype” is now subjected to debate. The taxonomy and systematics of *B. tabaci* have not yet been completely elucidated, but recently it was proposed that a value of 3.5% *mtCOI* sequence divergence should be used to distinguish 28 putative cryptic species within the *B. tabaci* complex (Dinsdale et al. 2010; Hu et al. 2011; see De Barro et al. 2011 and Liu et al. 2012 for review). Those species fall into 11 major groups (>11% divergence), which include those previously described as biotypes. In this study, we have adopted the nomenclature proposed by Dinsdale et al. (2010), but we have also included the biotype designation so as to make it easier to relate our study to the previously published literature.

Among the 28 putative species, two stand out as being the most predominant and devastating worldwide, the Middle East-Asia Minor species 1 (commonly known as biotype B; referred to hereafter as MEAM1) and the Mediterranean MED species (which includes the Q, J, L and ASL biotypes). These two species belong to the same Africa/Middle East/Asia Minor major group of *B. tabaci* identified by Dinsdale et al. (2010), which also includes MEAM2 and Indian Ocean IO species (formerly known as the Ms biotype). The MED and MEAM1 species have colonized large areas of most continents as a result of introduction/invasion processes (Bedford et al. 1994; Liu et al. 2007; De Barro and Ahmed 2011). Their invasive ability is mainly attributable to their having a broad spectrum of host plants, better reproductive and competitive potential and greater insecticide resistance. In the Mediterranean basin, the MED species predominates. In some areas, it coexists locally with MEAM1. The MED species is common in North Africa and has spread, probably recently, into West Central Africa and South Africa, where it coexists with the Sub-Saharan African indigenous species (also known as the AnSL biotypes; referred to hereafter as SSAF species) (Gueguen et al. 2010; Esterhuizen et al. 2013).

The MED species appears to be genetically diverse not only because it includes the Q, J, L and Sub-Saharan African Silverleaf ASL biotypes (De Barro et al. 2011), but also because high genetic variability has been observed within the Q group, leading to the recognition of

\*Corresponding author. Email: [laurence.mouton@univ-lyon1.fr](mailto:laurence.mouton@univ-lyon1.fr)

several cytotypes on the basis of *mtCOI* gene sequencing and symbiotic bacterial infection (Gueguen et al. 2010; Chu et al. 2012). Chu et al. (2012) recognized five subclades, designated as Q1–Q5, four of which correspond to the previously known Q1, Q2, Q3 and ASL genetic groups (Gueguen et al. 2010). These genetic groups display some particular features. First, they have different geographical and host plant ranges. For example, Q1 and Q2 have been sampled at a large geographical scale on a wide spectrum of host plants, whereas Q3 has only been observed in Western Africa on *Lantana camara* L. and tobacco (Gnankiné et al. 2012). The Q subclades also differ with regard to the composition of the bacterial symbiotic communities they harbor (Gueguen et al. 2010; Gnankiné et al. 2012). Some of these bacteria are known to induce drastic phenotypic change in *B. tabaci*, conferring features such as a fitness benefit (Kontseladov et al. 2008; Himler et al. 2011). Finally, they also differ as regards the frequencies of alleles resistant to pyrethroids and organophosphates (Mouton et al. unpublished data).

The spread of the MED and MEAM1 species worldwide has had a considerable impact on agriculture (for review, see Oetting and Buntin 1996), and many studies have focused on identifying the mechanisms underlying their invasiveness and the best methods of control. This requires monitoring *B. tabaci* populations to track the evolution of the species composition in the invaded areas. Gene sequencing is costly and time consuming; therefore, other effective diagnostic tools need to be developed in order to perform epidemiological studies involving hundreds of individuals. One of the various polymerase chain reaction (PCR) based methods, PCR-RFLP, consists of digestion with restriction enzymes after a PCR, and it provides a simple, fast and effective way to detect some of the variations in DNA sequences without sequencing. This technique has already been developed for use in *B. tabaci*, mainly to distinguish MEAM1, MED and non-MEAM1, non-MED species (Bosco et al. 2006; Tsagkarakou et al. 2007; Vassiliou et al. 2008; Ma et al. 2009; Shatters et al. 2009). Moreover, other PCR-RFLP protocols have been proposed to detect genetic variability in the Mediterranean basin, but these do not identify all the genetic diversity known to exist in the Africa/Middle East/Asia Minor group (Sartor et al. 2008; Chu et al. 2012). In this study, we describe a PCR-RFLP method that can be used to identify all the putative species that belong to the major groups (MED, MEAM1, MEAM2 and IO species) and also to recognize within-species variations. We also considered the Sub-Saharan African group, which is composed of several Sub-Saharan African species (SSAF1–5 species), which are sympatric with the MED species in some localities (Berry et al. 2004; Gueguen et al. 2010; Gnankiné et al. 2012; Esterhuizen et al. 2013).

## 2. Materials and methods

### 2.1. Phylogenetic analyses

We performed a phylogenetic analysis to determine the genetic diversity within *B. tabaci* populations in the two

major groups under consideration in this study, i.e. the Africa/Middle East/Asia Minor and the Sub-Saharan African clades. All the *mtCOI* sequences of *B. tabaci* available in Genbank for these two groups were analyzed, and only the sequences of around 657 bp with no undefined nucleotides, and no gaps or indels, were used. This gave us 143 sequences originating from samples collected worldwide (Table 1). Multiple sequence alignment was carried out using MUSCLE software (Edgar 2004) implemented in CLC DNA Workbench (CLC bio). Only non-redundant sequences were retained to construct the tree; this gave us 25 haplotypes (for each haplotype, the sequence used for the analysis is shown in bold type in Table 1). Phylogenetic analyses were performed using maximum likelihood (ML) and Bayesian inferences. The appropriate model of evolution selected with jModelTest v0.1.1 (Posada 2008) was GTR + I + G for both methods. ML analyses were performed with PhyML v3.0 (100 bootstrap replicates) (Guindon et al. 2010). Bayesian analyses were done using MrBayes v3.1.2 with 500,000 generations (Ronquist and Huelsenbeck 2003). Results were plotted using FigTree v1.3.1.

### 2.2. In silico selection of restriction enzymes

Restriction enzymes were selected by *in silico* analyses of the 25 haplotypes using the CLC DNA Workbench 6.0 (CLC Bio) program. CLC sequence viewer was used to predict restriction sites on a 657-bp fragment of the *mtCOI* gene sequence with 50 restriction enzymes.

### 2.3. PCR-RFLP on field samples

The PCR-RFLP tool developed in this study was tested on almost 1100 *B. tabaci* individuals originating from France, Spain, Israel, Greece, Togo, Benin and Burkina Faso, and collected from various host plants (Table 2). DNA was extracted using the procedure of Gueguen et al. (2010). Briefly, individual insects were ground in 25  $\mu$ L of extraction buffer containing 50 mM KCl, 10 mM Tris-base pH 8.0, 0.45% Nonidet P-40, 0.45% Tween 20 and 500 mg/mL proteinase K. Samples were incubated at 65 °C for 3 h and then at 100 °C for 15 min. A 35- $\mu$ L aliquot of pure water was added to this extract, which was then stored at –20 °C until use. An 819-bp fragment of *mtCOI* was amplified with the primers C1-J-2195 and L2-N-3014 (Simon et al. 1994). PCR amplifications were performed in a final volume of 25  $\mu$ L containing 200  $\mu$ M dNTPs, 200 nM of each primer, 1.5 mM of MgCl<sub>2</sub>, 0.5 U of *Taq* polymerase (EuroblueTaq, Eurobio, France) and 2  $\mu$ L of DNA template. The cycling profile consisted of an initial denaturing phase at 95 °C for 2 min, followed by 35 cycles consisting of: 94 °C for 1 min (denaturing), 50 °C for 1 min (annealing) and 72 °C for 1 min (extension) followed by a final extension phase at 72 °C for 10 min. PCR products (10  $\mu$ L) were digested with *Xap*I, and another 10  $\mu$ L was digested with *Bfm*I (10 UI) at 37 °C for 3 h. The products obtained were then separated by electrophoresis on a 2% agarose gel at 100 V for 1 h and visualized

Table 1. Geographic origin and types of restriction profiles of *Bemisia tabaci* strains retrieved from Genbank for the *in silico* analysis. Numbers in bold indicate the sequences used for the phylogenetic analysis.

<i>B. tabaci</i> major group	<i>B. tabaci</i> species	Biotype	Geographic origin	Type of restriction profile					Genbank accession number
				<i>XapI</i>	<i>BfmI</i>	Profile number	<i>XapI</i>	<i>BfmI</i>	
Africa/Middle East/Asia Minor	Middle East-Asia Minor1	B	India	2	2	Profile 1		AF321927, AJ748368	
			Morocco					AM176570, AJ517768	
			Israel					AY766369, AY766373, DQ174536	
			USA					EF080824, AY057123	
			Australia					DQ174535	
			Guadeloupe					AM180064	
			France					AJ550170	
			Argentina					AF340215	
			South Africa					AY057140	
			Japan					AB204578, AB204580, AB204581	
			Pakistan					AJ510071, AJ510074, AJ510075, AJ510076, AJ510079, AJ510081	
			Indian Ocean					AJ550173, AJ550174, AJ877260	
			China					AY686062, AY686063, AY686073	
			United Arab Emirates					DQ133382	
			Italy					GU086341	
			Taiwan					GU086342	
			Kuwait					GU086346	
			Jordan					GU086347	
			Spain					GU086348	
Dominican Republic					GU086349				
Kuwait			10	2	Profile 2	GU086354, GU086355			
Iran						GU086350, GU086352, GU086353			
Saudi Arabia						GU086357, GU086358, GU086344, GU086345			
Yemen						GU086343, GU086359, GU086360			
South Korea			2	9	Profile 3	DQ174538			
Reunion			1	5	Profile 4	AJ550177			
Algeria			1	1	Profile 5	AM176575			
Morocco						EU760753, AY0571387			
France						EU760737			
Burkina Faso						FJ766425			
Cameroon						EU760722			
China						HM802268			
Japan						AB204586, AB204587, AB204588			
USA						EF080823			
Turkey			8	1	Profile 6	AF342776			
Israel						EU760756			
USA						FJ188686, EF080821			
Burkina Faso			8	2	Profile 7	FJ766384, FJ766429			
Cameroon			5	4	Profile 8	AF344258			

(continued)

Table 1. (Continued)

<i>B. tabaci</i> major group	<i>B. tabaci</i> species	Biotype	Geographic origin	Type of restriction profile			Genbank accession number
				<i>XapI</i>	<i>BfmI</i>	Profile number	
Sub-Saharan Africa	Zimbabwe						AF344285
	Ivory Coast						AY057136
	Ghana			8	4	Profile 9	AY827579, AY827580, AY827587
	Ghana			5	6	Profile 10	AY827581
	Ghana	L		1	1	Profile 5	AY827588, AY827590
	Sudan			1	2	Profile 11	AY827612, AY827614, AY827615
	Zimbabwe	ASL		5	4	Profile 8	AY827613
	Cameroon						AF344286
	Burkina Faso						EU760758
	Uganda						FJ766410, FJ766435
	Côte d'Ivoire						AY903571
	Ghana						FJ766383
	Croatia			1	1	Profile 5	AY827589
	Reunion	Ms		9	8	Profile 12	GU086334, GU086335
	Uganda			6	3	Profile 13	EU760759
	South Africa	AnSL1					AY903539
	Ghana						AF344267, AF344268, AF344264
Zambia						AF418668, AY827591	
Swaziland						AF344281	
Tanzania						AF344276, AF344270	
Uganda						AF418667	
Uganda			4	3	Profile 14	AY057162	
						AY057168, AY057149, AY057151, AY057178,	
						AY057179, AY057180, AY057182, AY057183,	
						AY057185, AY057210, AY057181	
			7	3	Profile 15	AY057169	
Uganda	AnSL2		4	3	Profile 14	AY057141, AY057208, AY827611, AF418669,	
Uganda						AY057143, AY057146, AY057173, AY057194,	
						AY827605	
			4	7	Profile 16	AY827604	
Mali						AY827607	
Nigeria			4	3	Profile 14	AF344257	
Cameroon	AnSL3		3	3	Profile 17	AF344245, AF344246, AF344249, AF344251,	
Cameroon	AnSL4					AF344252, AF344254, AF344255, AF344247	
South Africa	AnSL5		3	3	Profile 17	JN104711, JN104712, JN104713, JN104714,	
						JN104715, JN104716, JN104717, JN104718,	
						JN104719, JN104720	

Table 2. Samples of *Bemisia tabaci* used in this study.

<i>B. tabaci</i> major group	<i>B. tabaci</i> species	Biotype	Geographic origin	Host plant	PCR-RFLP	Sequencing	
Sub-Saharan Africa	Sub-Saharan Africa 1	AnSL1	Togo	Cassava	11	2	
			Benin	Cassava	15	1	
	Sub-Saharan Africa 2	AnSL2	Togo	Cassava	7	1	
			Benin	Cassava	1	3	
Africa/Middle East/ Asia Minor	Mediterranean	ASL	Burkina Faso	Cotton	6	0	
				Tomato	74	0	
				Eggplant	2	0	
				Marrow	8	11	
				Tobacco	119	0	
				Benin	Cotton	10	6
			Tomato	26	6		
			Q1	Burkina Faso	Cotton	175	21
					Tomato	91	8
					Eggplant	9	4
					Marrow	26	3
					Tobacco	2	1
		Cucumber			29	0	
		Togo		Cotton	29	2	
		Benin		Cotton	2	2	
		Spain		Melon	5	1	
		Tomato		28	1		
		Cucumber		38	1		
		France		Hibiscus	15	1	
		Eggplant	15	1			
		Greece	Cucumber	20	1		
		Q2	Greece	Eggplant	10	1	
				Spain	Cucumber	7	1
				Melon	26	1	
			Sweet pepper	38	1		
			Tomato	42	1		
			France	Hibiscus	47	1	
Eggplant	15	1					
Q3	Israel	Cotton	20	1			
		Tomato	4	0			
		Tobacco	16	9			
Middle East-Asia Minor1	B	Spain	<i>Lantana camara</i>	83	21		
			Cotton	7	4		
			Israel	Cotton	1	1	

Note: Numbers indicated in the "PCR-RFLP" and "sequencing" columns correspond to the number of individuals tested by PCR-RFLP and the number of samples sequenced, respectively.

by ethidium bromide staining. In parallel, PCR products were directly sequenced from 120 individuals.

### 3. Results

#### 3.1. Phylogenetic structure

The phylogenetic tree obtained with the 25 selected haplotypes is consistent with the most recent *B. tabaci* phylogenies available in the literature (Dinsdale et al. 2010; De Barro et al. 2011). Since the ML and the Bayesian analyses led to substantially the same topology, we present only the ML analysis (Figure 1). Genetic variability was observed within the major groups and species recognized by De Barro et al. (2011). Fifteen haplotypes were identified in the Africa/Middle East/Asia Minor group, and seven in the Sub-Saharan African group. Genetic diversity was also found within species, with up to nine haplotypes in the MED species and four in MEAM1. Polymorphism was lower in the other species analyzed (five haplotypes

for SSAF5, three for SSAF1, two for SSAF2 and one for MEAM2, IO, SSAF3 and SSAF4), but this may have been due to the limited number of sequences available in Genbank for these species.

#### 3.2. In silico PCR-RFLP

Two restriction enzymes were selected after *in silico* analysis on the basis of their capacity to discriminate between the Africa/Middle East/Asia Minor and Sub-Saharan African major groups, and their ability to discriminate within these groups between the putative species described by Dinsdale et al. (2010) and De Barro et al. (2011). Ten and nine profile types were found for *XapI* and *Bfml*, respectively (Figure 2). Both enzymes were reliably able to distinguish between the two major groups. For *XapI*, we found profile types 1, 2, 5, 8, 9 and 10 for the Africa/Middle East/Asia Minor group, and profile types 3, 4, 6 and 7 for the Sub-Saharan African major group. For *Bfml*,



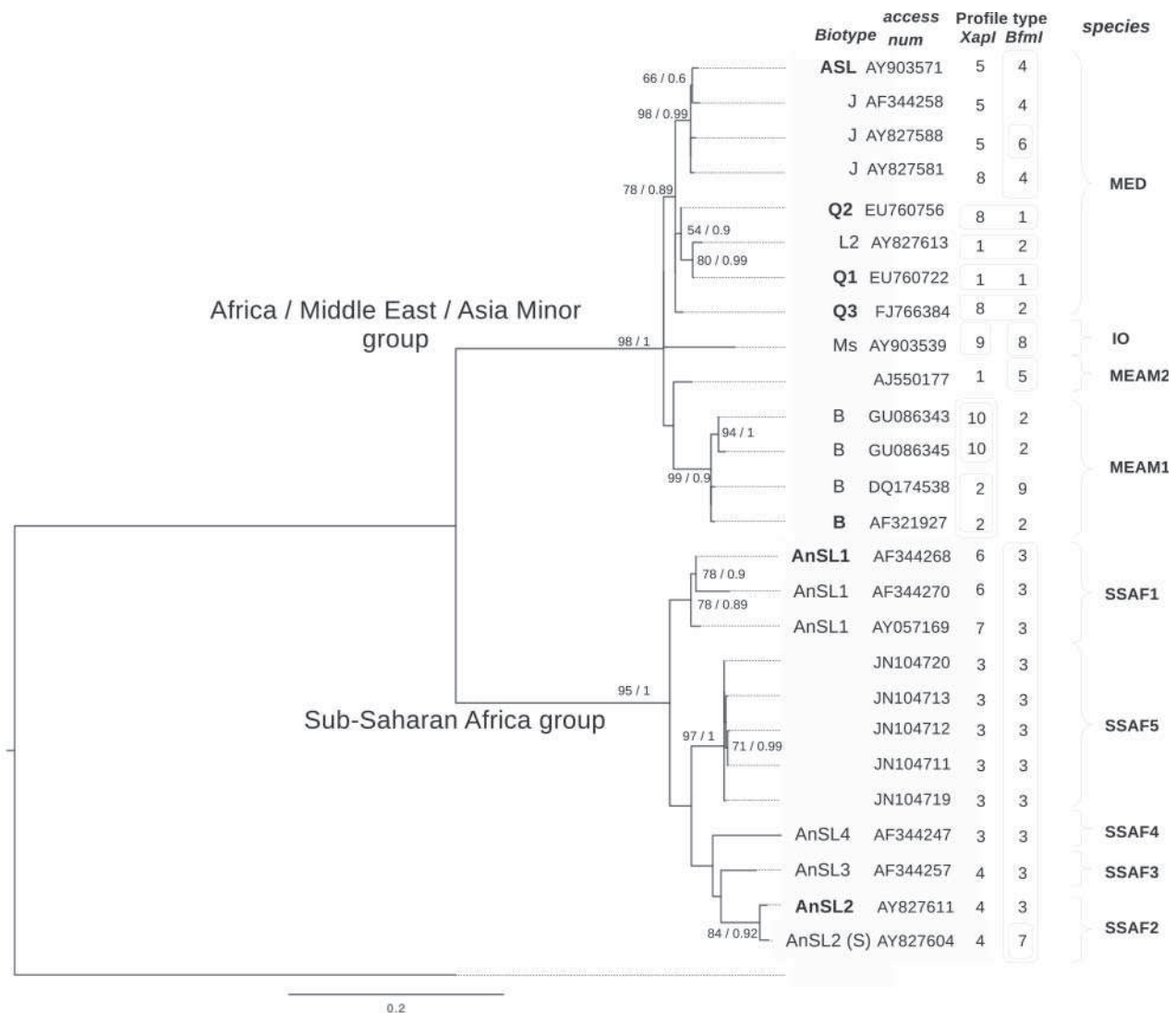


Figure 1. *mtCOI* gene-based phylogeny of the Africa/Middle East/Asia Minor and Sub-Saharan African major groups of *B. tabaci* using maximum likelihood (ML) analyses. Bootstrap values are shown at the nodes for ML analysis, followed by the corresponding posterior probabilities obtained by Bayesian inferences. Sequences were retrieved from Genbank. *Bemisia afer* was used as the outgroup. The analysis was based on a 657-bp sequence. Biotypes and access numbers are indicated at each branch, as is the type of profile obtained with each enzyme (*XapI*/*BfmI*). Frames indicate the specific profiles that can be used to distinguish between the sequences (for example, profiles 10 and 2 obtained with *XapI* are specific to the MEAM1 species, whereas restriction with both enzymes is required to identify Q1, Q2, Q3 and L2).

profiles 1, 2, 4, 5, 6, 8, 9 correspond to the Africa/Middle East/Asia Minor group, while profiles 3 and 7 characterized the Sub-Saharan African major group (Figure 1 and Table 1). Seventeen profiles were obtained when simultaneous DNA digestion by both enzymes was performed: 12 for the Africa/Middle East/Asia Minor group and 7 for the Sub-Saharan African group (Table 1). Within the Sub-Saharan African group, we were unable to distinguish all the SSAF species. In contrast, within the Africa/Middle East/Asia Minor group, PCR-RFLP profiles distinguished all the putative species: profiles 1–3 were particular to the MEAM1 species; profiles 5–11 were specific to the MED species and profiles 4 and 12 characterized the MEAM2 and IO species, respectively. In summary, *in silico* analysis suggests that *XapI* and *BfmI* acting simultaneously offer good potential identification of all four putative

species in the Africa/Middle East/Asia Minor group, including the ability to detect within-species diversity, since 12 profiles were obtained for 15 known haplotypes.

### 3.3. PCR-RFLP on field samples

The *in silico* predictions were checked on field samples belonging to the Africa/Middle East/Asia Minor (MEAM1 and MED species) and the Sub-Saharan African (SSAF1 and SSAF2 species) groups. We performed both PCR-RFLP and *mtCOI* gene sequencing on 120 *B. tabaci* individuals collected worldwide from several plants (Table 2). For another 959 individuals, the identification was done using the PCR-RFLP method only, in order to check the reliability of the method. Digestion with *XapI* and *BfmI* generated six and four distinct restriction



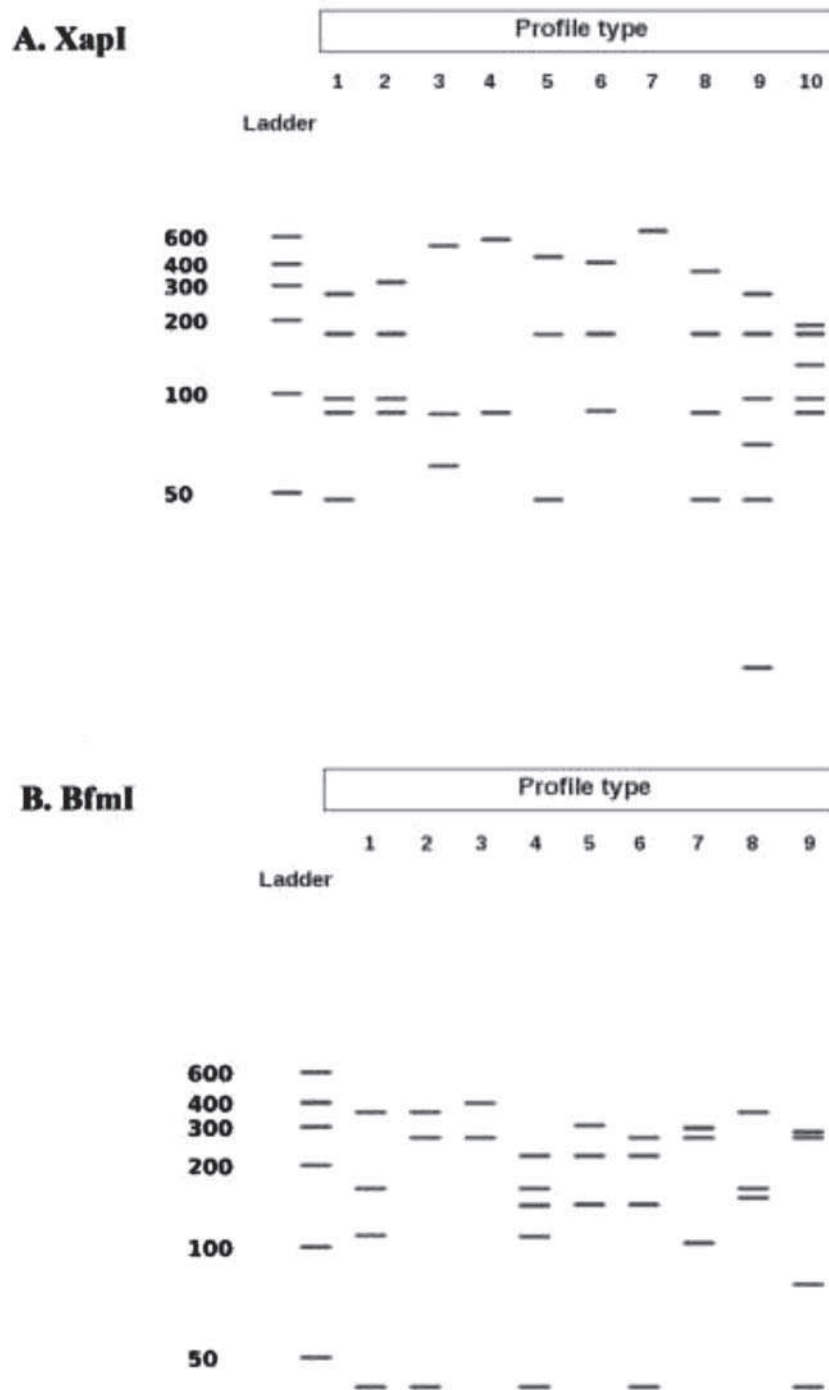


Figure 2. *In silico* restriction profiles obtained with *XapI* (A) and *Bfml* (B). The sizes of the bands corresponding to the ladder are indicated (bp).

patterns in our samples, respectively, differing by the number and/or size of the fragments, making it possible to identify the MEAM1, MED and SSAF species (Figure 3) as predicted by the *in silico* analysis (Table 1). Moreover, restriction profiles differentiated haplotypes within the MED species, making it possible to recognize the commonly known ASL, Q1, Q2 and Q3 genetic groups previously described as different biotypes or cytotypes, which harbor different bacterial endosymbionts. Fragments of less than 50 bp were not detected, because they were too

small to be visualized by electrophoresis in a routine agarose gel assay. Moreover, the size of the bands observed for some restriction profiles deviated slightly from predictions; this was because the *in silico* analysis was performed on 657 bp, which is only a part of the 867-bp amplicon obtained with the primers used to amplify the *mtCOI* gene in *B. tabaci* (C1-J-2195 and L2-N-3014; Simon et al. 1994).

For all 120 individuals in which biotype identification was carried out using both standard *mtCOI* gene

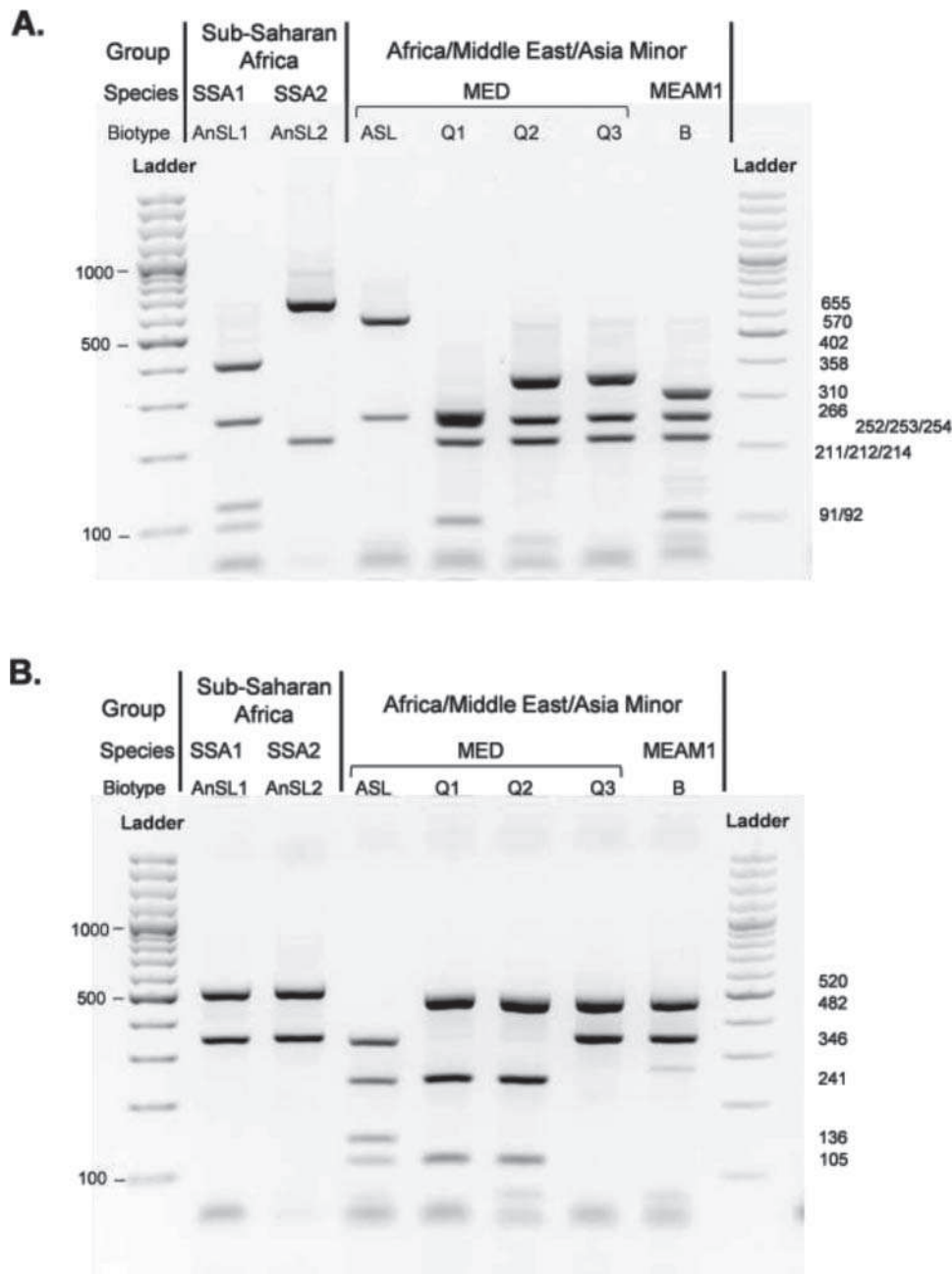


Figure 3. PCR-RFLP assays with *XapI* (A) and *BfmI* (B). The sizes of the different bands obtained and the sizes (bp) of the bands corresponding to the ladder are indicated on the left and the right of the figure, respectively.

sequencing and PCR-RFLP assay, the same results were obtained by both methods (Table 2). Moreover, we did not detect any unexpected restriction profiles for the 959 other individuals screened by PCR-RFLP only. This implies there was no other sequence variation at the *XapI* and *BfmI* restriction sites than those detected by the *in silico* analysis.

#### 4. Discussion

In this paper, a simple diagnostic tool based on a PCR-RFLP method on the *mtCOI* gene was used:

- (1) to reliably distinguish between the two major groups of *B. tabaci* that coexist in sympatry in

some places in the Mediterranean basin and Africa (Gueguen et al. 2010; Gnankiné et al. 2012; Esterhuizen et al. 2013): the Africa/Middle East/Asia Minor group, which includes the two most invasive members (B and Q biotypes), and the Sub-Saharan African group, which is regarded as the basal clade of *B. tabaci*;

- (2) to detect all the putative species defined in the Africa/Middle East/Asia Minor group by De Barro et al. (2011) and Dinsdale et al. (2010). These species include MEAM1 (formerly referred to as biotype B) and MED (biotype Q), which are widespread and considered to be the ones that cause the most damage (Oetting and Buntin

1996). In many places, displacement of MEAM1 by MED has been observed (Horowitz et al. 2005; Crowder et al. 2010), and the mechanism(s) underlying this phenomenon have been investigated for many years; this requires monitoring *B. tabaci* populations to track the evolution of the species composition in these areas. One explanation for this displacement could be the fact that these two species differ in their susceptibility towards various insecticides. Indeed, it has been demonstrated that the use of neonicotinoids or pyriproxyfen positively selects for the MED species, which exhibits greater resistance to these insecticides than the MEAM1 species (Horowitz et al. 2005).

Genetic diversity at the mitochondrial level has also been detected at a finer scale, i.e., within species. This PCR-RFLP protocol makes it possible to identify entities previously recognized as the Q1, Q2, Q3 and ASL genetic groups on the basis of *mtCOI* gene sequencing. While it is still not completely clear whether these groups are of any biological significance, some arguments suggest that this distinction should be taken into account. First, these haplotypes harbor particular symbiont communities that can potentially modify a number of biological traits in their host (Chiel et al. 2007; Gueguen et al. 2010; Gnankiné et al. 2012). Second, all the genetic groups of MED species are structured geographically (Gueguen et al. 2010; Gnankiné et al. 2012), so that tracking them provides a way of identifying the invasion routes of *B. tabaci*. Finally, they are associated with biological features, such as host plant use (Gnankiné et al. 2012) and the frequency of insecticide-resistant alleles (Mouton et al. unpublished data). This further suggests that the mitochondrial diversity could also be associated with nuclear variation and possible genetic isolation. This last point is important because species definition by Dinsdale et al. (2010) is based on *mtCOI* divergence, which means that some so far unidentified, genetically isolated entities may exist at a finer phylogenetic scale. In a recent study, Chu et al. (2012) found five haplotypes within the MED species by *mtCOI* sequencing. However, the PCR-RFLP method they developed for the *mtCOI* gene using the restriction enzyme *VspI* only distinguished one haplotype, which corresponds to the commonly known Q1 biotype, but the method was unable to detect the other four. The tool we developed here can be used to identify four out of the five *mtCOI* haplotypes described in the MED species. It can thus detect diversity at a very fine scale, which may make it very useful in population studies and may have major implications for pest management programs.

Levels of resistance and resistance mutation frequencies to insecticides differ between *B. tabaci* species (Horowitz et al. 2005; Alon et al. 2006, 2008; Wang et al. 2010) but also within species as has been observed within the MED species (Mouton pers.com) Therefore, to be effective, strategies developed to control this pest must differ according to the local *B. tabaci* composition and the

chemical classes of insecticides used. Monitoring the evolution of genetic diversity is, thus, clearly required for effective control strategies of this pest. The tool developed here has already been successfully used for describing the diversity of *B. tabaci* genetic groups in western Africa (Gnankiné et al. 2012; Gnankiné et al. 2013; Mouton, pers.com) and could be useful to monitor the dynamics of *B. tabaci* populations in relation to insecticide resistance in countries, like west African countries, where insecticides are widely used and lead to high levels of resistance (Houndété et al. 2010).

### Acknowledgements

This study was partially funded by CNRS (IFR41-UMR5558) and by the Agence Universitaire de la Francophonie (AUF). Gabriel Terraz is the recipient of a PhD studentship from the Rhône-Alpes region ("Program Cible" Grant). The authors would like to thank A. Tsagkarakou, J. Arno, G. Bout and E. Zchori-Fein for providing samples, and F. Vavre for helpful discussion and comments.

### References

- Alon M, Alon F, Nauen R, Morin S. 2008. Organophosphates' resistance in the B-biotype of *Bemisia tabaci* (Hemiptera: Aleyrodidae) is associated with a point mutation in an acetylcholinesterase and overexpression of carboxylesterase. *Insect Biochem Mol Biol.* 38:940–949.
- Alon M, Benting J, Lueke B, Ponge T, Alon F, Morin S. 2006. Multiple origins of pyrethroid resistance in sympatric biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Insect Biochem Mol Biol.* 36:71–79.
- Bedford ID, Briddon RW, Brown JK, Rosell RC, Markham PG. 1994. Geminivirus transmission and biological characterisation of *Bemisia tabaci* (Gennadius) biotypes from different geographic regions. *Ann Appl Biol.* 125:311–325.
- Berry SD, Fondong VN, Rey MEC, Rogan D, Fauquet C, Brown JK. 2004. Molecular evidence for five distinct *Bemisia tabaci* (Homoptera: Aleyrodidae) geographic haplotypes associated with cassava plants in Sub-Saharan Africa. *Ann Entomol Soc Am.* 97:852–859.
- Bosco D, Loria A, Sartor C, Cenis JL. 2006. PCR-RFLP identification of *Bemisia tabaci* biotypes in the Mediterranean basin. *Phytoparasitica.* 34:243–251.
- Boykin LM, Shatters RG, Rosell RC, McKenzie CL, Bagnall RA, De Barro P, Frohlich DR. 2007. Global relationships of *Bemisia tabaci* (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Mol Phylogenet Evol.* 44:1306–1319.
- Chiel E, Gottlieb Y, Zchori-Fein E, Mozes-Daube N, Katzir N, Inbar M, Ghanim M. 2007. Biotype-dependent secondary symbiont communities in sympatric populations of *Bemisia tabaci*. *Bull Entomol Res.* 97:407–413.
- Chu D, Hu X, Gao C, Zhao H, Nichols RL, Li X. 2012. Use of mitochondrial cytochrome oxidase I polymerase chain reaction-restriction fragment length polymorphism for identifying subclades of *Bemisia tabaci* Mediterranean group. *J Econ Entomol.* 105:242–251.
- Crowder DW, Horowitz R, De Barro PJ, Liu SS, Showalter AM, Kontsedalov S, Khasdan V, Shargal A, Liu J, Carriere Y. 2010. Mating behaviour, life history and adaptation to insecticides determine species exclusion between whiteflies. *J Anim Ecol.* 79:563–570.
- De Barro P, Ahmed M. 2011. Genetic networking of the *Bemisia tabaci* cryptic species complex reveals pattern of biological invasions. *PLoS One.* 6:e25579.

- De Barro PJ, Liu SS, Boykin LM, Dinsdale B. 2011. *Bemisia tabaci*: a statement of species status. *Ann Rev Entomol.* 56:1–19.
- Dinsdale A, Cook L, Riginos C, Buckley YM, De Barro P. 2010. Refined global analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodoidae: Aleyrodidae) mitochondrial cytochrome oxidase I to identify species level genetic boundaries. *Ann Entomol Soc Am.* 103:196–208.
- Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* 5:113.
- Esterhuizen LL, Mabasa KG, van Heerden SW, Czosnek H, Brown JK, van Heerden H, Rey MEC. 2013. Genetic identification of members of the *Bemisia tabaci* cryptic species complex from South Africa reveals native and introduced haplotypes. *J Appl Entomol.* 137:122–135.
- Frohlich DR, Torres-Jerez I, Bedford ID, Markham PG, Brown JK. 1999. A phylogeographical analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers. *Mol Ecol.* 8:1683–1691.
- Gnankiné O, Mouton L, Henri H, Terraz G, Houndété T, Martin T, Vavre F, Fleury F. 2012. Distribution of the *Bemisia tabaci* biotypes (Homoptera: Aleyrodidae) and their associated symbiotic bacteria on host plants in Western Africa. *Insect Conservation Divers.* 6:411–421.
- Gnankiné O, Mouton L, Savadogo A, Martin T, Sanon A, Vavre F, Fleury F. 2013. Biotype status and resistance to neonicotinoid and carbosulfan in *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Burkina Faso, West Africa. *Int J Pest Manag.* 59:95–102.
- Gueguen G, Vavre F, Gnankine O, Peterschmitt M, Charif D, Chiel E, Gottlieb Y, Ghanim M, Zchori-Fein E, Fleury F. 2010. Endosymbiont metacommunities, mtDNA diversity and the evolution of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. *Mol Ecol.* 19:4365–4378.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59:307–321.
- Himler AG, Adachi-Hagimori T, Bergen JE, Kozuch A, Kelly SE, Tabashnik BE, Chiel E, Duckworth VE, Dennehy TJ, Zchori-Fein E, Hunter MS. 2011. Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. *Science.* 332:254–256.
- Horowitz AR, Kontsedalov S, Khasdan V, Ishaaya I. 2005. Biotypes B and Q of *Bemisia tabaci* and their relevance to neonicotinoid and pyriproxyfen resistance. *Arch Insect Biochem Physiol.* 58:216–225.
- Houndété TA, Kétoh GK, Hema OSA, Brévaukt T, Glitho IA, Martin T. 2010. Insecticide resistance in field populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in West Africa. *Pest Manag Sci.* 11:1181–1185.
- Hu J, De Barro P, Zhao H, Wang J, Nardi F, Liu S-S. 2011. An extensive field survey combined with a phylogenetic analysis reveals rapid and widespread invasion of two alien whiteflies in China. *PLoS One.* 6:e16061.
- Jones DR. 2003. Plant viruses transmitted by whiteflies. *Eur J Plant Pathol.* 109:195–219.
- Kontsedalov S, Zchori-fein E, Chiel E, Gottlieb Y, Inbar M, Ghanim M. 2008. The presence of *Rickettsia* is associated with increased susceptibility of *Bemisia tabaci* (Homoptera: Aleyrodidae) to insecticides. *Pest Manag Sci.* 792:789–792.
- Liu SS, Colvin J, De Barro PJ. 2012. Species concepts as applied to the whitefly *Bemisia tabaci* systematics: how many species are there? *J Integr Agric.* 11:176–186.
- Liu SS, De Barro PJ, Xu J, Luan JB, Zang LS, Ruan YM, Wan FH. 2007. Asymmetric mating interactions drive widespread invasion and displacement in a whitefly. *Science.* 318:1769–1772.
- Ma W-H, Li X-C, Lei C-L, Wang M, Degain BA, Nichols RL. 2009. Utility of mtCOI polymerase chain reaction-restriction fragment length polymorphism in differentiating between Q and B whitefly *Bemisia tabaci* biotypes. *Insect Sci.* 16:107–114.
- Oetting RD, Buntin GD. 1996. *Bemisia* damage expression in commercial greenhouse production. In: Gerling D, Mayer RT, editors. *Bemisia: 1995 taxonomy, biology, damage, control and management.* Andover (UK): Intercept; p. 201–208.
- Perring TM. 2001. The *Bemisia tabaci* species complex. *Crop Prot.* 20:725–737.
- Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol.* 25:1253–1256.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 19:1572–1574.
- Sartor C, Demichelis S, Cenis JL, Coulibaly AK, Bosco D. 2008. Genetic variability of *Bemisia tabaci* in the Mediterranean and Sahel regions. *Bull Insect.* 61:161–162.
- Shatters RG, Powell C, Boykin LM, Liansheng H, McKenzie CL. 2009. Improved DNA barcoding method for *Bemisia tabaci* and related Aleyrodidae: development of universal and *Bemisia tabaci* biotype-specific mitochondrial cytochrome c oxidase I polymerase chain reaction primers. *J Econ Entomol.* 102:750–758.
- Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved PCR primers. *Ann Entomol Soc Am.* 87:651–701.
- Tsagkarakou A, Tsigenopoulos CS, Gorman K, Lagnel J, Bedford ID. 2007. Biotype status and genetic polymorphism of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Greece: mitochondrial DNA and microsatellites. *Bull Entomol Res.* 97:29–40.
- Vassiliou VA, Jagge C, Grispou M, Pietrantonio PV, Tsagkarakou A. 2008. Biotype status of *Bemisia tabaci* from various crops in Cyprus. *Phytoparasitica.* 36:400–404.
- Wang Z, Yan H, Yang Y, Wu Y. 2010. Biotype and insecticide resistance status of the whitefly *Bemisia tabaci* from China. *Pest Manag Sci.* 66:1360–1366.

# Detection of genetically isolated entities within the Mediterranean species of *Bemisia tabaci*: new insights into the systematics of this worldwide pest

Laurence Mouton,<sup>a\*</sup> Olivier Gnankiné,<sup>b</sup> H  l  ne Henri,<sup>a</sup> Gabriel Terraz,<sup>a</sup> Guillaume Ketoh,<sup>c</sup> Thibaud Martin,<sup>d</sup> Fr  d  ric Fleury<sup>a</sup> and Fabrice Vavre<sup>a</sup>



## Abstract

**BACKGROUND:** The taxonomy of the species complex *Bemisia tabaci*, a serious agricultural pest worldwide, is not well resolved yet, even though species delimitation is critical for designing effective control strategies. Based on a threshold of 3.5% mitochondrial (*mtCOI*) sequence divergence, recent studies have identified 28 putative species. Among them, mitochondrial variability associated with particular symbiotic compositions (=cytotypes) can be observed, as in MED, which raises the question of whether it is a single or a complex of biological species.

**RESULTS:** Using microsatellites, an investigation was made of the genetic relatedness of Q1 and ASL cytotypes that belong to MED. Samples of the two cytotypes were collected in West Africa where they live in sympatry on the same hosts. Genotyping revealed a high level of differentiation, without evidence of gene flow. Moreover, they differed highly in frequencies of resistance alleles to insecticides, which were much higher in Q1 than in ASL.

**CONCLUSION:** Q1 and ASL are sufficiently reproductively isolated for the introgression of neutral alleles to be prevented, suggesting that they are actually different species. This indicates that nuclear genetic differentiation must be investigated within groups with less than 3.5% *mtCOI* divergence in order to elucidate the taxonomy of *B. tabaci* at a finer level. Overall, these data provide important information for pest management.

   2014 Society of Chemical Industry

Supporting information may be found in the online version of this article.

**Keywords:** MED species; *B. tabaci* complex species; microsatellites; insecticide resistance; bacterial symbionts

## 1 INTRODUCTION

The whitefly *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) is one of the most devastating agricultural pests worldwide. Highly polyphagous, it causes damage in various agricultural crops either directly by sucking phloem sap or indirectly by excreting honeydew onto the surface of leaves and fruits.<sup>1</sup> Furthermore, *B. tabaci* is a major vector of several hundred plant viruses (Begomovirus, Crinivirus, Ipomovirus, Torradovirus) that are responsible for severe crop losses.<sup>2</sup>

*B. tabaci* is a complex of cryptic species whose delimitation is critical if effective control strategies are to be devised.<sup>3</sup> Recently, important progress has been made at the taxonomic level, with the definition of 28 genetic groups based on a threshold of 3.5% mitochondrial cytochrome oxidase I (*mtCOI*) sequence divergence that are regarded as cryptic species.<sup>4–7</sup> However, two recent studies from Boykin *et al.*<sup>3</sup> and Lee *et al.*<sup>8</sup> have indicated that this number is probably underestimated, as there is important variability within some species. This is notably the case within the Mediterranean (MED) species, with six mitochondrial haplotypes identified so far, generally associated with particular symbiotic

communities in which they are in linkage disequilibrium.<sup>4,9,10</sup> The biological significance of these mitochondria–endosymbiont assemblages (referred to hereafter as cytotypes) is still unclear, but they call into question whether they are also associated with nuclear differentiation; in other words, whether putative species, especially MED, are really a single or a complex of biological species. Clearly, additional data are needed, notably data

\* Correspondence to: Laurence Mouton, Universit   de Lyon, Universit   Lyon 1, Laboratoire de Biom  trie et Biologie Evolutive, UMR CNRS 5558, 43 Bd du 11 Novembre 1918, 69622 Villeurbanne Cedex, France. E-mail: laurence.mouton@univ-lyon1.fr

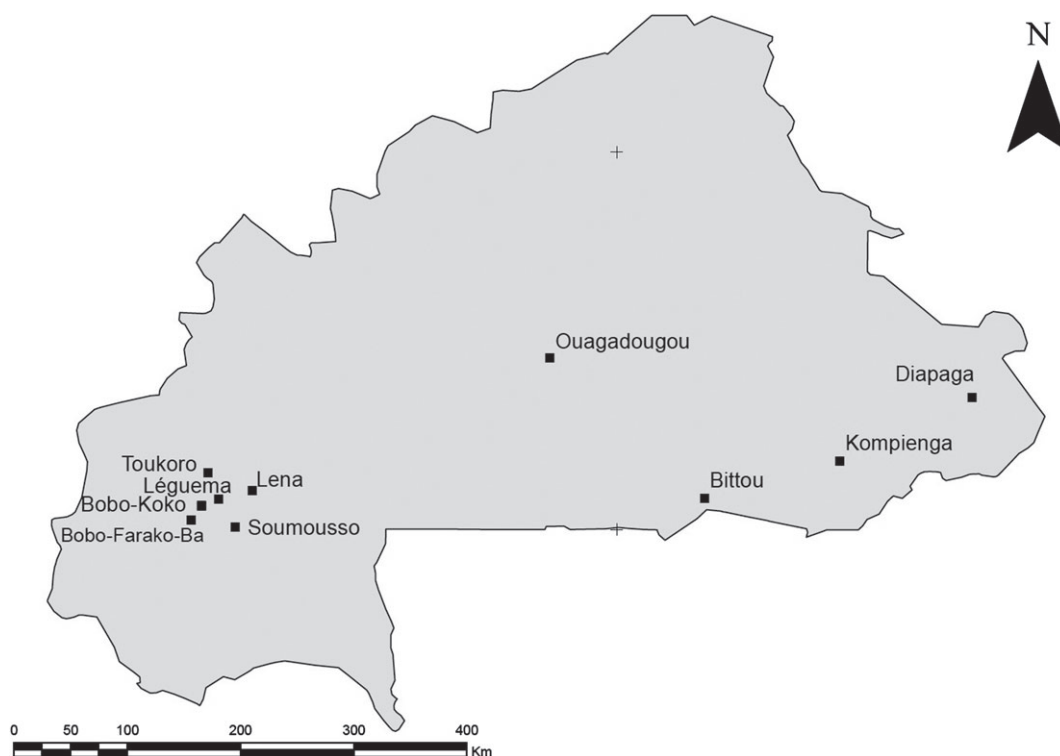
a Laboratoire de Biom  trie et Biologie Evolutive, Universit   de Lyon, Villeurbanne, France

b Laboratoire d'Entomologie Appliqu  e, Universit   de Ouagadougou, Burkina Faso

c Facult   des Sciences, Universit   de Lome, Lome, Togo

d UR Hortsys, CIRAD, Montpellier, France





**Figure 1.** Sampling locations in Burkina Faso.

indicating whether interbreeding occurs in the field between cytotypes, before the taxonomy and systematics of *B. tabaci* can be reliably defined.<sup>7</sup> However, field population studies are difficult to perform because many factors, such as host plant specialisation and geographical barriers, can limit genetic exchanges. The ideal way to avoid these confounding effects is to study cytotypes that live sympatrically, but this situation is uncommon in *B. tabaci*.<sup>11–13</sup>

In the present study, the authors took advantage of the situation recently observed in Burkina Faso (West Africa), where two cytotypes belonging to the putative MED species, Q1 and Africa silverleafing (ASL), live in sympatry on the same host plants,<sup>14</sup> to investigate their genetic relatedness at the nuclear level. These two cytotypes present less than 3.1% *mtCOI* sequence variation.<sup>15</sup> In spite of the occurrence of another cytotpe classified within the MED species, the cytotpe Q3, in this country, it was not included because it develops specifically on a restricted number of host plants, especially *Lantana camara*, an ornamental plant where no other cytotpe has been detected so far.<sup>14</sup>

For the present study, seven microsatellite loci already described in the literature<sup>16–19</sup> were used as neutral markers to explore the genetic relatedness of Q1 and ASL cytotypes. Moreover, as insecticides are intensively used by farmers on cotton and vegetables in West Africa,<sup>20</sup> the presence of alleles associated with insecticide resistance was also investigated. In *B. tabaci*, two mutations in the *para*-type voltage-gated sodium channel gene, L925I and T929V, and one mutation in the acetylcholinesterase enzyme *ace1* (F331W) confer resistance to pyrethroids and organophosphates respectively.<sup>21–24</sup> Previous studies have shown that the use of these insecticides results in the development of resistant *B. tabaci* populations,<sup>25,26</sup> which is undermining the current management of this pest.<sup>27</sup>

The data obtained clarify the degree of interbreeding between Q1 and ASL and thus provide important biological insights for

*B. tabaci* systematics. They also provide information about the frequencies and spatial distribution of the resistant mutation alleles in Burkina Faso.

## 2 EXPERIMENTAL METHODS

### 2.1 *Bemisia tabaci* samples

Sampling was performed in several locations in Burkina Faso (West Africa) (Fig. 1). The samples used for microsatellite genotyping were collected in 2007 (Table 1), and those used for investigating the presence of insecticide resistance alleles were collected in 2009 (Table 2), except those from Bobo/Kuinima (2007). Adult whiteflies were collected from cultivated vegetables (tomato, eggplant, marrow and cucumber) and industrial crops (cotton and tobacco), and stored in ethanol 95% at  $-20^{\circ}\text{C}$  until use. The origin of the samples (location and host plant) and the number of individuals are summarised in Tables 1 and 2.

### 2.2 DNA extraction

Total DNA was extracted from individual insects in 25  $\mu\text{L}$  of an extraction buffer containing 50 mM of KCl, 10 mM of Tris-base pH 8, 0.45% Nonidet P-40, 0.45% Tween 20 and 50 mg mL<sup>-1</sup> of proteinase K. After 3 h at 65  $^{\circ}\text{C}$ , samples were incubated at 100  $^{\circ}\text{C}$  for 15 min. Pure water (35  $\mu\text{L}$ ) was then added to the extract.

### 2.3 Identification of *B. tabaci* cytotypes

Cytotypes were identified using the polymerase chain reaction–random fragment length polymorphism (PCR-RFLP) diagnostic assay based on the mitochondrial marker cytochrome oxidase 1 gene sequence (*mtCOI*) described in Henri et al.,<sup>15</sup> which can be used to discriminate between all the known cytotypes detected in West Africa.<sup>14</sup>



**Table 1.** Samples of *B. tabaci* from Burkina Faso genotyped for genetic structure analyses and inbreeding coefficient ( $F_{IS}$ ) of each population

Acronym <sup>a</sup>	Locality	Host plant	Biotype	$N_{ind}^b$	$F_{IS}$
Cot1	Kompienga	Cotton	Q1	15	0.224
Cot2	Diapaga	Cotton	Q1	14	0.223
Cot3	Lena	Cotton	Q1	15	0.190
Cot6	Bobo/Farako-Bâ	Cotton	Q1	14	0.276
Egg5	Bobo/Kuinima	Eggplant	Q1	8	0.119
			ASL	1	
Tom5	Bobo/Kuinima	Tomato	Q1	9	0.041
			ASL	34	0.039
Mar5	Bobo/Kuinima	Marrow	Q1	11	0.068
			ASL	14	0.064

<sup>a</sup> Cot: cotton; Tom: tomato; Egg: eggplant; Mar: marrow; each number refers to a specific locality.  
<sup>b</sup>  $N_{ind}$ : the number of individual whiteflies genotyped, Cot: cotton, Tom: tomato, Egg: eggplant, Mar: marrow.

**Table 2.** Frequencies of sodium channel and *ace-1* resistance mutations in *B. tabaci* from West Africa per host plant and locality<sup>a</sup>

Acronym <sup>b</sup>	Locality	Host plant	Cytotype	Sodium channel							Ace								
				Male genotypes			Female genotypes				Allele frequencies	Male genotypes			Female genotypes				Allele frequencies
				n	r1	s	n	r1r1	r1s	ss		n	R	S	n	RR	RS	SS	
Cot4	Kompienga	Cotton	Q1	6	6	0	10	8	2	0	r1=0.92	6	5	1	10	8	2	0	R=0.88
Tom12	Bobo/Koko	Tomato	ASL	0	0	0	5	0	0	5	r1=0	0	0	0	5	0	2	3	R=0.20
			Q1	1	1	0	10	10	0	0	r1=1	1	1	0	10	10	0	0	R=1
Tom9	Léguéma	Tomato	ASL	2	0	2	12	0	0	12	r1=0	2	1	1	12	4	2	6	R=0.42
			Q1	3	3	0	12	12	0	0	r1=1	3	3	0	12	12	0	0	R=1
Mar9	Léguéma	Marrow	ASL	0	0	0	5	0	0	5	r1=0	0	0	0	5	1	1	3	R=0.3
			Q1	0	0	0	6	6	0	0	r1=1	0	0	0	6	5	1	0	R=0.92
Tom10	Toukoro	Tomato	Q1	2	2	0	13	13	0	0	r1=1	2	2	0	13	13	0	0	R=1
Cuc5	Bobo/Kuinima	Cucumber	Q1	0	0	0	29	29	0	0	r1=1	0	0	0	29	29	0	0	R=1
Cot11	Bittou	Cotton	Q1	4	4	0	19	19	0	0	r1=1	4	4	0	19	18	1	0	R=0.97
Tom8	Soumouso	Tomato	ASL	4	1	3	6	0	0	6	r1=0.06	4	4	0	6	1	3	2	R=0.56
			Q1	2	2	0	15	15	0	0	r1=1	2	2	0	15	15	0	0	R=1

<sup>a</sup> r1 and R refer to resistant alleles for the sodium channel and Ace genes respectively; s: susceptible allele.  
<sup>b</sup> Cot: cotton; Tom: tomato; Cuc: cucumber; Mar: marrow; each number refers to a specific locality.

## 2.4 Microsatellite genotyping

Genetic structure analysis using microsatellite markers was done on 135 individuals collected from five localities in Burkina Faso (Table 1). *B. tabaci* is a haplodiploid species, i.e. males hatch from unfertilised eggs and thus are haploids, and therefore genotyping was done on females only to estimate observed heterozygosity.

Genotyping was performed on seven microsatellite loci (supporting information Table S1). PCR reactions were performed on each primer pair separately in 12.5  $\mu$ L volumes containing 200  $\mu$ M dNTPs, 200 nM of primers (fluorescently labelled at the 5' end), 0.5 IU of *Taq* DNA polymerase (Eurobio, Courtaboeuf, France), and 1  $\mu$ L of DNA template. The cycling conditions were as follows: initial denaturing at 94 °C for 2 min, followed by 35 cycles of 30 s at 94 °C for denaturing, 30 s at 52 or 57 °C, depending on the primer set (supporting information Table S1), for annealing, and 1 min at 72 °C for elongation, with a final elongation phase at 72 °C for 10 min.

Fluorescent amplicons and a size standard (GS400Rox size standard; Applied Biosystems, Warrington, UK) were loaded onto an ABIPrism3100-Avant genetic analyser (Applied

Biosystems) automated sequencer. Allele sizes were calculated using GeneMapper software (Applied Biosystems). Null alleles were searched for using Micro-Checker (freely available at: <http://www.microchecker.hull.ac.uk/>)<sup>28</sup>.

## 2.5 Genetic data analysis

### 2.5.1 Population differentiation tests

Genepop<sup>29</sup> was used to assess the linkage disequilibrium between all pairs of loci and to test for Hardy–Weinberg equilibrium within populations and cytotypes. We considered as a population whiteflies collected in the same locality and on the same host plant. Bonferroni correction was applied for the chi-squared tests performed to analyse the linkage disequilibrium. The mean number of alleles per locus, observed and expected heterozygosities and *F*-statistic parameters of Weir and Cockerham<sup>30</sup> were calculated using GENETIX.<sup>31</sup>

### 2.5.2 Inferences about population structure

Two programs based on different assumptions were used to assess the level of population structure: Structure<sup>32</sup> and Instruct

(<http://cbsuapps.tc.cornell.edu/InStruct.aspx>). The first assumes Hardy–Weinberg equilibrium for all loci, while the second does not. As data were congruent, we focused mainly on Structure (see De Barro<sup>33</sup>). This free software uses a Bayesian clustering approach to determine the number of subpopulations ( $K$ ) present on the basis of the individual genotypes. The admixture model was used, with a burning period of 60 000 iterations followed by 600 000 iterations. Log-likelihood estimates were calculated for  $K = 1–10$ , with ten replicates of each. The modal value of  $\Delta K$ , a quantity based on the second-order rate of change in the log probability of data with respect to  $K$ , was used to determine the number of clusters.<sup>34</sup> Assignment tests were also performed with GeneClass, using the Bayesian method (available at <http://www1.montpellier.inra.fr/URLB/geneclass/geneclass.html><sup>35</sup>).

## 2.6 Identification of susceptible and resistant alleles of the sodium channel and *ace1* genes

Resistant and susceptible alleles in the *para*-type voltage-gated sodium channel and *ace1* genes were detected using the diagnostic assays developed by Tsagkarakou *et al.*<sup>36</sup> Briefly, *ace1* susceptible (F331) and *ace1* resistant (W331) alleles, as well as susceptible (L925) and resistant (I925) *para*-type voltage-gated sodium channel alleles, were detected using PCR-RFLP, and the T929V resistant mutation was detected by PCR using primer pairs that amplified only one of the two alleles (V929 or T929). Several PCR products were sequenced for each susceptible and resistant allele and each country.

## 3 RESULTS

### 3.1 Microsatellite variability

A total of 135 females from five sites and four host plants belonging to Q1 and ASL cytotypes were genotyped on the basis of seven microsatellite loci, with 6.3% missing data (Tables 1 and 3). ASL individuals were detected at a single location but from three different host plants; thus, three populations were considered for this cytotype. No null alleles were detected using Micro-Checker. The number of alleles ranged from 4 to 31, with an average of 12.1 per locus in the whole sample, and from 1 to 29 per cytotype, with averages of 11 for Q1 and 5 for ASL. One allele was fixed for locus 145 in ASL.

Heterozygote deficiency was detected for all loci and in all populations, leading to high  $F_{IS}$  values. This deficit persisted even when samples were grouped by cytotype, with a highly significant multilocus  $F_{IS}$  (0.188 for Q1 and 0.218 for ASL;  $P < 0.001$ ). No linkage disequilibrium was found between the loci, implying that the seven microsatellite loci carried independent genetic information.

### 3.2 Genotype-based inferences about population structure

The clustering analysis carried out using Structure and Instruct on all individuals of *B. tabaci* was performed from  $K = 1$  to  $K = 10$  subpopulations (Fig. 2; supporting information Table S2). The highest likelihood value was obtained for  $K = 2$ , and this was supported by using  $\Delta K$  as the predictor of the real number of clusters. The two clusters defined corresponded to the cytotypes Q1 and ASL, demonstrating that they correlated with genetic differentiation at the nuclear level (Fig. 2). Assignment tests using GeneClass gave similar results, as all the individuals were correctly assigned to their cytotype (86 Q1 and 49 ASL). The overall differentiation among

**Table 3.** Summary of genetic variation at the seven microsatellite loci<sup>a</sup>

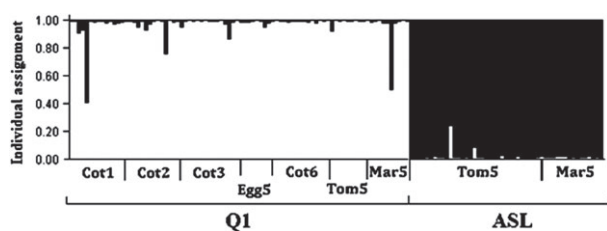
Locus		Q1 $N = 86$	ASL $N = 49$	All samples $N = 135$
BtIS2.3	$N_A$	12	4	14
	Allele range	102–126	105–109	102–126
	$H_E$	0.857	0.511	0.669
	$H_O$	0.726	0.488	0.533
	$f$	0.154	0.045	0.203
BtIS1.13	$N_A$	4	3	4
	Allele range	107–117	113–117	107–117
	$H_E$	0.524	0.279	0.431
	$H_O$	0.465	0.245	0.333
	$f$	0.112	0.124	0.228
BtIS1.1	$N_A$	10	4	11
	Allele range	239–275	241–249	239–275
	$H_E$	0.562	0.385	0.399
	$H_O$	0.482	0.121	0.304
	$f$	0.142	0.689	0.241
BtIS1.2	$N_A$	29	8	31
	Allele range	274–368	294–348	274–368
	$H_E$	0.901	0.629	0.622
	$H_O$	0.724	0.617	0.467
	$f$	0.198	0.019	0.251
Locus11	$N_A$	5	5	7
	Allele range	160–178	168–176	160–176
	$H_E$	0.609	0.512	0.431
	$H_O$	0.430	0.396	0.296
	$f$	0.295	0.229	0.314
BT83	$N_A$	12	10	17
	Allele range	132–174	138–164	132–164
	$H_E$	0.795	0.691	0.598
	$H_O$	0.679	0.488	0.482
	$f$	0.147	0.295	0.196
Locus145	$N_A$	5	1	6
	Allele range	167–183	173	167–183
	$H_E$	0.407	0	0.271
	$H_O$	0.305	0	0.185
	$f$	0.253	-	0.317
All	Mean $N_A$	11	5	12.1
	Mean $H_E$	0.665	0.429	0.670
	Mean $H_O$	0.544	0.336	0.509
	Multilocus $F_{IS}$	0.182	0.219	0.187

<sup>a</sup>  $N$ : number of samples;  $N_A$ : number of alleles;  $H_E$ : expected heterozygosity;  $H_O$ : observed heterozygosity;  $f$ : inbreeding coefficient.

all the samples in clusters was supported by a high  $F_{ST}$  of 0.183 ( $P < 0.001$ ).

Analyses were also done on each cytotype separately using Structure, from  $K = 1$  to  $K = 5$  (supporting information Table S2). The highest likelihood value was obtained for  $K = 1$ , indicating that there was no differentiation among populations within these cytotypes. Genetic differentiation within cytotypes was also analysed by computing  $F_{ST}$  estimates, and the results supported the previous findings, as values of  $F_{ST}$  were not significant for Q1 or ASL ( $P > 0.5$ ), with values of 0.009 and 0.003 respectively.

The present data do not comply with one of the criteria required by Structure, as there was a deficiency of heterozygotes. However, all the other analyses, which do not require Hardy–Weinberg



**Figure 2.** Clustering results for all samples obtained using Structure. Cot: cotton; Tom: tomato; Egg: eggplant; Mar: marrow; each number refers to a specific locality.

equilibrium, did support the validity of the results, as they were confirmed using Instruct, the assignment tests of GeneClass and *F*-values.

### 3.3 Detection of sodium channel and *ace1* resistant mutations and allele frequencies

Either PCR-RFLP or specific PCR, as described in Tsagkarakou *et al.*,<sup>36</sup> was used to look for resistant mutations in the *ace1* (F331W) and in the *para*-type voltage-gated sodium channel (L925I and T929V) genes in samples collected in Burkina Faso. The results are shown in Table 2. For both genes, the frequencies of resistant alleles were very variable, depending on the cytotypic. Overall, the frequencies of the F331W mutation in the *ace1* gene were 0.98 and 0.59 for Q1 and ASL respectively. For the *para*-type voltage-gated sodium channel gene, only one of the two resistant mutations described in *B. tabaci* was found – L925I; this corresponds to the allele designated r1 by Alon *et al.*<sup>23</sup> It was almost fixed in the Q1 cytotypic, with a frequency of 0.99, whereas it was rarely found in the ASL populations (frequency 0.02).

Within cytotypes, while frequencies of the L925I mutation were homogeneous in all the populations (Fisher's exact test,  $P > 0.05$  for the two cytotypes), frequencies of the *ace1* resistant allele were significantly different (Fisher's exact test,  $P < 0.05$ ) and varied depending on the host plant species ( $P = 0.01$  for Q1 and  $P < 0.001$  for ASL). The genotype frequencies did not differ significantly from those expected under Hardy–Weinberg equilibrium for each gene ( $P > 0.22$  for both Q1 and ASL). There was no linkage disequilibrium between the two genes in either of the two cytotypes (Fisher's exact test,  $P > 0.10$ ).

## 4 DISCUSSION

The genetic relatedness between the Q1 and ASL cytotypes classified in the MED putative species of *B. tabaci* by Dinsdale *et al.*<sup>4</sup> and De Barro *et al.*<sup>6</sup> have been characterised on the basis of microsatellite markers. Genetic data analysis revealed low  $F_{ST}$  values within the Q1 and ASL cytotypes between populations collected from locations 23 to 655 km apart, and from different host plant species (eggplant, tomato and marrow), indicating low levels of genetic differentiation. This suggests that, on this geographic scale, gene flows occur between populations found on different plants, as previously observed by De Barro.<sup>33</sup>

On the other hand, analyses of genetic relationships between cytotypes have revealed that ASL and Q1 individuals are genetically isolated. Indeed, clustering methods assigned the 135 females genotyped into two clusters strictly corresponding to the two cytotypes identified on the basis of the mitochondrial *mtCOI* marker. These data are corroborated by the high  $F_{ST}$  values between Q1 and ASL, further suggesting that Q1 and ASL do not

interbreed. On the basis of these results, it is proposed that MED is composed of at least two different species.

This reproductive isolation is not due to physical barriers, as ASL and Q1 live in sympatry on the same host plants. It is also unlikely that there is any cytoplasmic incompatibility (CI) due to bacteria, even if Q1 and ASL individuals do not harbour the same symbiotic community.<sup>9,14</sup> Each of these cytotypes harbours a distinct and highly prevalent bacterium (*Hamiltonella* in Q1, *Arsenophonus* in ASL<sup>14</sup>), but neither of them has been shown to induce CI in either of their hosts. Prevalences of the other bacteria they harbour are too low to lead to reproductive isolation due to CI (less than 10% each<sup>14</sup>). It is thus more likely that the lack of gene flow between Q1 and ASL involves either premating behavioural isolation, resulting in the absence of mating, or post-mating incompatibility, such as nuclear or nucleocytoplasmic incompatibilities, resulting in non-viable hybrids.

Four other cytotypes are classified in the MED putative species: Q2, Q3, J and L.<sup>6</sup> Recently, a study based on population genetics on field samples and laboratory crossing experiments has shown that Q1 and Q2 interbreed,<sup>37</sup> indicating that they belong to the same species. At the time, there are no data concerning the other cytotypes. Studies considering all cytotypes are required to clarify the systematics of the MED putative species, even though this is not an easy task because, except for Q1/Q2 and Q1/ASL, they do not live sympatrically. More generally, such studies are required on all the putative species where several cytotypes have been detected, such as in the New World.<sup>6</sup> This is all the more important, given that species delimitation is critical for optimising integrated pest management programmes.

The present study, taken together with previous data, calls into question the use of the *mtCOI* gene with a 3.5% threshold for *B. tabaci* taxonomy. Indeed, Lee *et al.*<sup>8</sup> highlighted that the variability within *B. tabaci* putative species has been underestimated for at least six of them.<sup>8</sup> Moreover, it has been shown that high *mtCOI* distances (up to 7%) are not inconsistent with possible hybridization between *B. tabaci* cytotypes.<sup>38</sup> Furthermore, there can be a bias to use mitochondrial markers because mitochondria are subjected to indirect selection caused by linkage disequilibrium with maternally inherited endosymbionts.<sup>39</sup> According to Galtier *et al.*,<sup>40</sup> mitochondrial DNA should not be used for population genetic studies and phylogenetic analysis but should be restricted to the description of biodiversity. Alternative approaches based on nuclear markers should thus be used for *B. tabaci* species identification. Among them, microsatellites are the most common markers that could help to distinguish true species, but recent methods based on new-generation sequencing (NGS) tools such as restriction-site-associated DNA markers (RADseq) could also be used. Indeed, RADseq markers make it possible to obtain genome-wide data without previous information and have been proven to be a powerful tool for phylogenetic inferences even for non-model organisms.<sup>41,42</sup>

The mutation that confers resistance to organophosphates was found in both Q1 and ASL, but, while this resistant allele was almost fixed in Q1 [mean 0.98, standard error (SE) 0.02], its frequency was intermediate in ASL (mean 0.59, SE 0.08). The high selection pressure due to the repetitive organophosphate treatments applied in western Africa, particularly on cotton and vegetables,<sup>20,27</sup> has probably driven the spread of this mutation. On the other hand, while the resistant mutation on the *para*-type voltage-gated sodium channel gene that confers resistance to pyrethroids was also almost fixed in Q1 populations, it was rarely detected in ASL.

ASL and Q1 have always been found in sympatry and thus have been exposed to the same insecticide compounds.<sup>14</sup> This raises the question of how ASL can be maintained. One hypothesis would be that ASL individuals use other mechanisms for insecticide resistance, such as metabolic resistance. Indeed, detoxifying enzymes such as esterases, glutathione S-transferases and cytochrome-P450-dependent monooxygenases are involved in resistance to numerous insecticide classes.<sup>23,43,44</sup> An alternative explanation would be that the coexistence of the two cytotypes is recent, and that either ASL has just arrived in West Africa and has not yet been counterselected or Q1 has just arrived and ASL has not gone to extinction yet because it also encounters some plant refuges. However, the present study revealed the presence of Q1 and ASL for at least 2007, i.e. 2 years before the survey of resistant alleles was performed. Monitoring the situation across years will help in determining whether this situation is at equilibrium or not.

## 5 CONCLUSIONS

While the taxonomy and systematics of *B. tabaci* have been recently clarified, species boundaries remain mainly based on divergence of mitochondrial sequences. This allows a clearer representation of the genetic diversity of this species complex on a large scale, and makes it possible to define major groups and putative species. However, genetic diversity at the mitochondrial level has also been detected on a finer scale, and the present results suggest that cytotypes within putative species could also be associated with nuclear variation, making it possible that some genetically isolated entities still exist on a phylogenetic scale finer than 3.5% *mtCO1* sequence divergence. This study demonstrates that reproductive compatibility must be backed up by observations of gene flows to support species delimitation. Moreover, the data indicate the presence of resistant alleles for organophosphates and pyrethroids in Burkina Faso where insecticides are widely used. However, while Q1 and ASL share the same host plants, they differ in the frequencies of resistant mutations. This last point is of particular importance for the choice of the strategy employed to control *B. tabaci*. Overall, these data provide new insights into the biological significance of *B. tabaci* cytotypes and are important for management of this pest.

## ACKNOWLEDGEMENTS

This work was funded by UNESCO (France) and CNRS (IFR41-UMR5558). These results have been achieved within the framework of the first call on Mediterranean agriculture carried out by ARIMNet, with funding from MOARD (Israel), ANR (France), INIA (Spain), NAGREF-DEMETER (Greece) and GDAR (Turkey). The authors are grateful to the INERA Institute in Burkina Faso, University of Lomé, for sample collection and technical assistance. G Terraz is the recipient of a PhD studentship from the Rhône-Alpes region ('Program Cible' Grant). The authors thank S Charlat for useful discussions.

## SUPPORTING INFORMATION

Supporting information may be found in the online version of this article.

## REFERENCES

- Oliveira MRV, Henneberry TJ and Anderson P, History, current status, and collaborative research projects for *Bemisia tabaci*. *Crop Prot* **20**:709–723 (2001).
- Brown JK and Czosnek H, Whitefly transmission of plant viruses. *Adv Bot Res* **36**:65–100 (2002).
- Boykin LM, Armstrong KF, Kubatko L and De Barro P, Species delimitation and global biosecurity. *Evol Bioinf* **8**:1–37 (2012).
- Dinsdale A, Cook L, Riginos C, Buckley YM and De Barro P, Refined global analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodidae: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. *Ann Entomol Soc Am* **103**:196–208 (2010).
- Hu J, De Barro P, Zhao H, Wang J, Nardi F and Liu SS, An extensive field survey combined with a phylogenetic analysis reveals rapid and widespread invasion of two alien whiteflies in China. *PLoS ONE* **6**:e16061 (2011).
- De Barro PJ, Liu SS, Boykin LM and Dinsdale B, *Bemisia tabaci*: a statement of species status. *Annu Rev Entomol* **56**:1–19 (2011).
- Shu-Sheng L, Colvin J and De Barro PJ, Species concepts as applied to the whitefly *Bemisia tabaci* systematics: how many species are there? *J Integr Agric* **11**:176–186 (2012).
- Lee W, Park J, Lee G-S, Lee S and Akimotot SI, Taxonomic status of the *Bemisia tabaci* complex (Hemiptera: Aleyrodidae) and reassessment of the number of its constituent species. *PLoS ONE* **8**:e63817 (2013).
- Gueguen G, Vavre F, Gnankiné O, Peterschmitt M, Charif D, Chiel E et al., Endosymbiont metacommunities, mtDNA diversity and the evolution of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. *Mol Ecol* **19**:4365–4378 (2010).
- Chu D, Hu X, Gao C, Zhao H, Nichols RL and Li X, Use of mitochondrial cytochrome oxidase I Polymerase Chain Reaction–Restriction Fragment Length Polymorphism for identifying subclades of *Bemisia tabaci* Mediterranean group. *J Econ Entomol* **105**:242–251 (2012).
- Moya A, Guirao P, Cifuentes D, Beitia F and Cenis JL, Genetic diversity of Iberian populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae) based on random amplified polymorphic DNA-polymerase chain reaction. *Mol Ecol* **10**:891–897 (2001).
- Khasdan V, Levin I, Rosner A, Morin S, Kotsedalov S, Maslenin L et al., DNA markers for identifying biotypes B and Q of *Bemisia tabaci* (Hemiptera: Aleyrodidae) and studying population dynamics. *Bull Entomol Res* **95**:605–613 (2005).
- Horowitz AR, Kotsedalov S, Khasdan V and Ishaaya I, Biotypes B and Q of *Bemisia tabaci* and their relevance to neonicotinoid and pyriproxyfen resistance. *Arch Insect Biochem Physiol* **58**:216–225 (2005).
- Gnankiné O, Mouton L, Henri H, Terraz G, Houndeté T, Martin T et al., Distribution of *Bemisia tabaci* biotypes (Homoptera: Aleyrodidae) and their associated symbiotic bacteria on host plants in western Africa. *Insect Conserv Diversity* **6**:411–421 (2013).
- Henri H, Terraz G, Gnankiné O, Fleury F and Mouton L, Molecular characterization of genetic diversity within the Africa/Middle East/Asia Minor and Sub-Saharan African groups of the *Bemisia tabaci* species complex. *Int J Pest Manag* **59**:329–338 (2013).
- Tsagkarakou A and Roditakis N, Isolation and characterization of microsatellite loci in *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Mol Ecol Notes* **3**:196–198 (2003).
- Delatte H, David P, Granier M, Lett JM, Goldbach R, Peterschmitt M et al., Microsatellites reveal extensive geographical, ecological and genetic contacts between invasive and indigenous whitefly biotypes in an insular environment. *Genet Res* **87**:109–124 (2006).
- Gauthier N, Dalleau-Clouet C and Bouvret ME, Twelve new polymorphic microsatellite loci and PCR multiplexing in the whitefly, *Bemisia tabaci*. *Mol Ecol Res* **8**:1004–1007 (2008).
- Dalmon S, Halkett F, Granier M, Delatte H and Peterschmitt M, Genetic structure of the invasive pest *Bemisia tabaci*: evidence of limited but persistent genetic differentiation in glasshouse populations. *Heredity* **100**:316–325 (2008).
- Ahouangninou C, Fayomi B and Martin T, Évaluation des risques sanitaires et environnementaux des pratiques phytosanitaires des producteurs maraîchers dans la commune rurale de Tori-Bossito (Sud-Bénin). *Cah Agric* **20**:216–222 (2011).
- Morin S, Williamson MS, Goodson SJ, Brown JK, Tabashnik BE and Dennehy TJ, Mutations in the *Bemisia tabaci* para-sodium channel gene associated with resistance to a pyrethroid plus organophosphate mixture. *Insect Biochem Mol Biol* **32**:1781–1791 (2002).
- Roditakis E, Tsagkarakou A and Vontas J, Identification of mutations in the para sodium channel of *Bemisia tabaci* from Crete, associated with resistance to pyrethroids. *Pestic Biochem Physiol* **85**:161–166 (2006).



- 23 Alon M, Alon F, Nauen R and Morin S, Organophosphates' resistance in B-biotype of *Bemisia tabaci* (Hemiptera: Aleyrodidae) is associated with a point mutation in an *ace1*-type acetylcholinesterase and over-expression of carboxylesterase. *Insect Biochem Mol Biol* **38**:940–949 (2008).
- 24 Alon M, Benting J, Lueke B, Ponge T, Alon F and Morin S, Multiple origins of pyrethroid resistance in sympatric biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Insect Biochem Mol Biol* **36**:71–79 (2006).
- 25 Palumbo JC, Horowitz AR and Prabhaker N, Insecticidal control and resistance management for *Bemisia tabaci*. *Crop Prot* **20**:739–765 (2001).
- 26 Houndete TA, Ketoh GK, Hema OSA, Brevault T, Glitho IA and Martin T, Insecticide resistance in field populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in West Africa. *Pest Manag Sci* **66**:1181–1185 (2010).
- 27 Gnankiné O, Traoré D, Sanon A, Traoré NS and Ouedraogo AP, Traitements insecticides et dynamique des populations de *Bemisia tabaci* en culture cotonnière au Burkina Faso. *Cah Agric* **16**:101–109 (2007).
- 28 Van Oosterhout C, Hutchinson WF, Wills DPM and Shipley O, Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* **4**:535–538 (2004).
- 29 Raymond M and Rousset F, Genepop version 2.0: population genetics software for exact tests and ecumenicism. *J Heredity* **86**:248–249 (1995).
- 30 Weir BS and Cockerham CC, Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**:1358–1370 (1984).
- 31 Belkhir K, Borsa P, Chikhi L, Raufaste N and Bonhomme F, *GENETIX 4.05, Logiciel sous Windows™ pour la Génétique des Populations*. Laboratoire Génome, Populations, Interactions, Université de Montpellier II, Montpellier, France (2001).
- 32 Pritchard JK, Stephens M and Donnelly P, Inference of population structure using multilocus genotype data. *Genetics* **155**:945–959 (2000).
- 33 De Barro PJ, Genetic structure of the whitefly *Bemisia tabaci* in the Asia-Pacific region revealed using microsatellite markers. *Mol Ecol* **14**:3695–3718 (2005).
- 34 Evanno G, Regnaut S and Goudet J, Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**:2611–2620 (2005).
- 35 Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L and Estoup A, GENECLASS 2: a software for genetic assignment and first generation migrant detection. *J Heredity* **95**:536–539 (2004).
- 36 Tsagkarakou A, Nikou D, Roditakis E, Sharvit M, Morin S and Vontas J, Molecular diagnostics for detecting pyrethroid and organophosphate resistance mutations in the Q biotype of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Pest Biochem Physiol* **94**:49–54 (2009).
- 37 Terraz G, Gueguen G, Fleury F and Mouton L, Nuclear and cytoplasmic differentiation among Mediterranean populations of *Bemisia tabaci*: testing the biological relevance of cytotypes. *Pest Manag Sci*. DOI: 10.1002/ps.3792.
- 38 Thierry M, Becker N, Hajri A, Reynaud B, Lett JM and Delatte H, Symbiont diversity and non-random hybridization among indigenous (Ms) and invasive (B) biotypes of *Bemisia tabaci*. *Mol Ecol* **20**:2172–2187 (2011).
- 39 Hurst GD and Jiggins FM, Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc Biol Sci*. **272**:1525–1534 (2005).
- 40 Galtier N, Nabholz B, Glémin S and Hurst GD, Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol Ecol* **18**:4541–4550 (2009).
- 41 Cariou M, Duret L and Charlat S, Is RAD-seq suitable for phylogenetic inference? An *in silico* assessment and optimization. *Ecol Evol* **3**:846–852 (2013).
- 42 Keller I, Wagner CE, Greuter L, Mwaiko S, Selz OM, Sivasundar A *et al.*, Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Mol Ecol* **22**:2848–2863 (2013).
- 43 Rauch N and Nauen R, Identification of biochemical markers linked to neonicotinoid cross resistance in *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Arch Insect Biochem Physiol* **54**:165–176 (2003).
- 44 Ma W, Li X, Dennehy TJ, Lei C, Wang M, Degain BA *et al.*, Pyriproxyfen resistance of *Bemisia tabaci* (Homoptera: Aleyrodidae) biotype B: metabolic mechanism. *J Econ Entomol* **103**:158–165 (2010).





# Optimization of multiplexed RADseq libraries using low-cost adaptors

Hélène Henri · Marie Cariou · Gabriel Terraz ·  
Sonia Martinez · Adil El Filali · Marine Veyssiere ·  
Laurent Duret · Sylvain Charlat

Received: 29 April 2014 / Accepted: 3 February 2015  
© Springer International Publishing Switzerland 2015

**Abstract** Reduced representation genomics approaches, of which RADseq is currently the most popular form, offer the possibility to produce genome wide data from potentially any species, without previous genomic information. The application of RADseq to highly multiplexed libraries (including numerous specimens, and potentially numerous different species) is however limited by technical constraints. First, the cost of synthesis of Illumina adaptors including molecular identifiers (MIDs) becomes excessive when numerous specimens are to be multiplexed. Second, the necessity to empirically adjust the ratio of adaptors to genomic DNA concentration impedes the high throughput application of RADseq to heterogeneous samples, of variable DNA concentration and quality. In an attempt to solve these problems, we propose here some adjustments regarding the adaptor synthesis. First, we show that the common and unique (MID) parts of adaptors can be synthesized separately and subsequently ligated, which drastically reduces the synthesis cost, and thus allows multiplexing hundreds of specimens. Second, we show that self-ligation of adaptors, which makes the adaptor concentration so critical, can be simply prevented by using unphosphorylated adaptors, which significantly improves the ligation and sequencing yield.

**Keywords** Reduced representation genomics · RADseq · Protocol · Multiplexing

## Introduction

Reduced representation genomics, as opposed to full genome sequencing or random metagenomics, allow the sequencing of numerous homologous loci from numerous genomic extracts. In particular, methods involving the sequencing of regions flanking restriction sites (RADseq and affiliated approaches) attract growing interest, for they can produce powerful datasets in any organism to address any question related to genetic variability, from genetic mapping to deep phylogenies (Davey and Blaxter 2011; Cariou et al. 2013).

In this article, we propose some adjustments of standard RADseq protocols (Etter et al. 2011) to facilitate the application of this method to large numbers of heterogeneous genomic extracts in a single sequencing reaction. Our first objective is to reduce the cost of synthesizing large numbers of adaptors, which becomes excessive when several hundreds of genomic extracts are to be multiplexed. Our second objective is to design a method that would be more robust to variation and uncertainty in the initial DNA concentration and genome size.

RADseq library preparation starts with a digestion of genomic extracts, with one restriction enzyme, targeting rare or common restriction sites, depending on the required density of markers (Baird et al. 2008) (in the so-called “double digest” protocol, marker density can also be adjusted by combining two restriction enzymes; Peterson et al. 2012). Following digestion, adaptors (denoted P1) are specifically ligated to the free sticky ends. P1 adaptors include regions for PCR amplification and

**Electronic supplementary material** The online version of this article (doi:10.1007/s10709-015-9828-3) contains supplementary material, which is available to authorized users.

H. Henri (✉) · M. Cariou · G. Terraz · S. Martinez ·  
A. El Filali · M. Veyssiere · L. Duret · S. Charlat  
Laboratoire de Biométrie et Biologie Evolutive, CNRS, UMR  
5558, Université Lyon 1, Université de Lyon, 43 boulevard du  
11 novembre 1918, 69622 Villeurbanne, France  
e-mail: helene.henri@univ-lyon1.fr

Illumina sequencing, as well as a Molecular Identifier (MID) that should be at least 8 bp long if several hundreds of genomic extracts are to be sequenced together. The full adaptors contain two strands of respectively 72 and 68 nucleotides, bringing the cost of synthesis to approximately 100 Euros (~120\$) per adaptor, that is, 50,000 Euros (~62,000\$) for an experiment where 500 genomic extracts would be multiplexed. It would obviously be possible to reduce the cost of such an experiment by pooling fewer specimens in several independent libraries and sequencing reactions, with the positive outcome of increasing the average sequencing depth per locus. But if a large sequencing depth is not required, or if money is limiting, increasing the number of libraries and sequencing reactions is not an optimal approach. Alternatively, paired end sequencing can be used, and a combination of different MIDs in the forward and reverse adaptors then allows one to envisage high levels of multiplexing with a limited number of distinct MIDs. However, this approach is also not economical since it requires the preparation of independent libraries (one for each combination of forward and reverse tag), that can only be pooled at a late stage (that is, after the P2 adaptors have been ligated). We propose instead an additional step in the preparation of P1 adaptors that drastically reduces the synthesis cost: the long region that is common to all adaptors is synthesized only once, while the short region containing the MIDs are synthesized independently. The common and unique parts are then ligated to produce the complete adaptors. With these modifications, the cost of synthesizing 500 distinct adaptors drops from 50,000 to 5600 Euros.

The second modification we introduce aims at making library preparation more robust to variation in the initial amount of genomic DNA. Indeed, current protocols indicate that a deficit but also an excess of P1 adaptor reduces ligation efficiency, so that P1 concentration must be appropriately adjusted to genomic DNA concentration (Peterson et al. 2012), which cannot be reasonably achieved for a large number of heterogeneous genomic extracts. We hypothesized that this problem stemmed from the formation of P1–P1 dimers, and could thus be solved by using P1 adaptors that are devoid of a 5' phosphate and thus cannot ligate to themselves.

We assessed these protocol modifications on a test RADseq experiment including a diverse set of genomic extracts from 11 species, and several replicates of *Drosophila melanogaster* DNA used as positive controls, that is, samples where a reference genome could be used to assess the quality of the data. We show that our cheap P1 adaptors are efficient and that the use of unphosphorylated adaptors significantly improves the yield and robustness of RAD library preparation.

## Adaptors design and synthesis

8 bp MIDs were designed using the barcrawl program (Frank 2009), which provided 755 potential MIDs with the following characteristics and provided in the supplementary material S1: at least three differences between all MIDs, no homopolymers of more than two nucleotides, no hairpins or heteroduplexes of more than four nucleotides, no possible convergence of MIDs to identical sequences through 1 bp deletion. For the test experiment described here, a combination of 100 MIDs (listed in supplementary material S2) was selected to ensure an approximately balanced base composition at each position (even ratios of A, C, G and T). Indeed, cluster detection during Illumina sequencing is facilitated by heterogeneity of the fragments to be sequenced, especially in the five first nucleotides.

Figure 1 provides a schematic view of the preparation of our P1 adaptors, containing a region common to all adaptors and a variable region unique to each adaptor. The common region includes the sequence required for PCR and Illumina sequencing. The variable region contains the MID and 2 sticky ends on 5' and 3'. The 5' end corresponds to an overhang of the bottom strand that is specific of the restriction enzyme (and will thus ligate to the genomic DNA) and the 3' end to an overhang of the top strand that is cohesive to the 5' end of the common part.

A full, step by step protocol for the preparation of adaptors is provided in supplementary material 3. In brief, the common and unique parts of the adaptors (respectively in blue and red in Fig. 1) are synthesized and hybridized separately and finally ligated to produce 30 µL of complete adaptors at 0.5 µM. The 5' end of these full adaptors is not phosphorylated so that ligation to the genomic DNA relies on the 5' phosphate carried by digested genomic DNA. The resulting nick is filled in with Bst 2.0 polymerase (NEB).

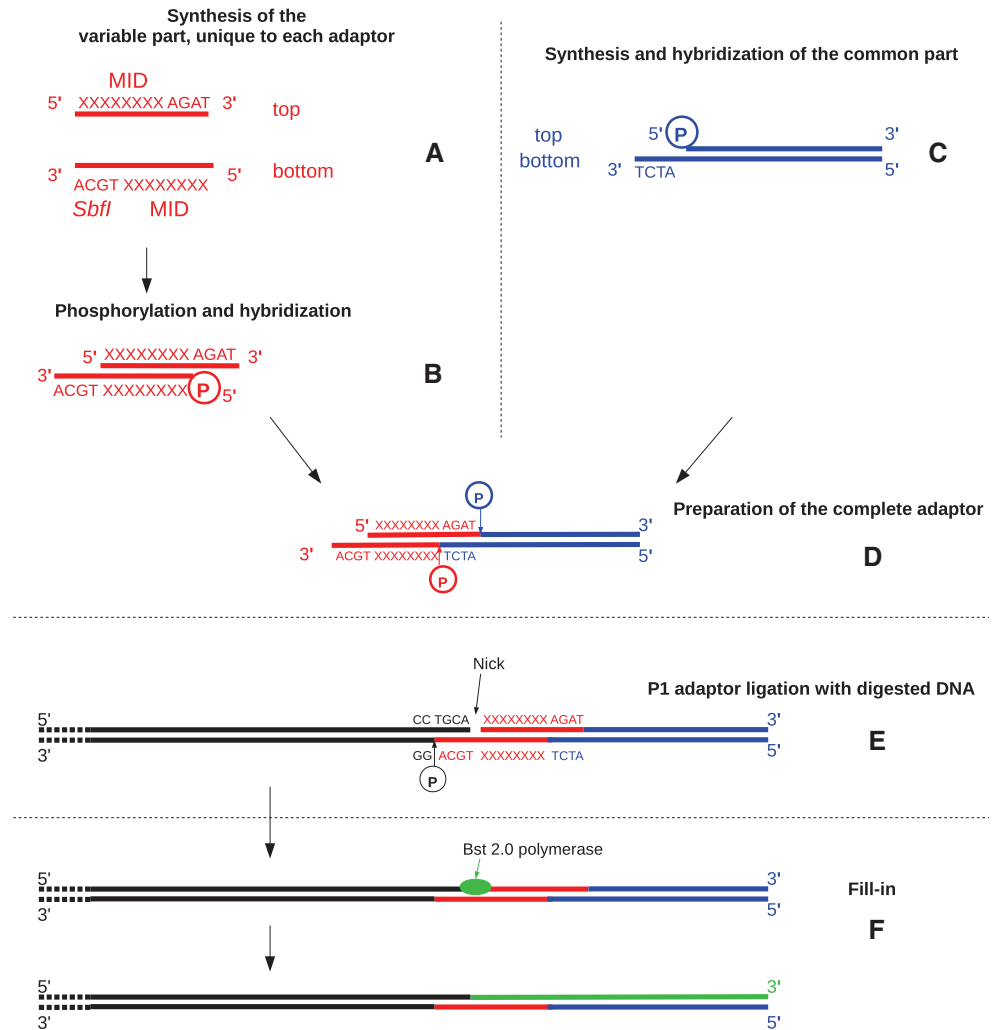
To verify that ligation of P1 adaptors to genomic DNA was effective, we randomly chose an SbfI cut site in the *D. melanogaster* genome, and designed PCR primers on the two flanking regions (mel1-CGA-CCA-GCA-GAC-CAATAA and mel2-GCT-CCA-CTA-CCA-GCT-ATA-A). Two PCRs were performed on each *D. melanogaster* sample, involving a forward primer targeting P1 (AAT-GAT-ACG-GCG-ACC-ACC-GA), and the mel1 and mel2 two reverse primers targeting the regions flanking the SbfI site. The amplification of these two fragments confirmed that P1 adaptors were properly ligated.

## Design of the test experiment

Genomic DNA of 77 specimens from 11 species (see table in supplementary material 4 for details) was extracted either with an affinity column (DNeasy kit, Qiagen) or

**Fig. 1** Step by step synthesis of the P1 adaptors (*a–f*) and ligation to genomic DNA.

*a* Synthesis of the two strands of the “MID + restriction site” part (*red*). *b* To reduce the cost of oligonucleotides synthesis, we order unphosphorylated oligos and add a phosphate to the 5′ end of the bottom strand only. The two strands are then hybridized. *c* Synthesis and hybridization of the common part (*blue*). The top strand is ordered with 5′ phosphorylation. *d* The two parts of the adaptors are ligated with T4 DNA ligase to produce complete adaptors. *e* Ligation of the complete P1 adaptor with DNA digested with *Sbf*I (*black*). Since the complete adaptor does not carry any 5′ phosphate, the ligation to the *Sbf*I-digested DNA relies only on a single phosphate provided by the genomic DNA. *f* The nick on the top strand is filled-in with *Bst*2.0 polymerase through its 5′-3′ DNA polymerase and strand displacement activities. The newly synthesized region is shown in *green*. (Color figure online)



through phenol–chloroform precipitation (Sambrook and Russel 2001). Six specimens of *D. melanogaster* were included in this experiment to be used as positive controls, based on knowledge of their full genome. DNA concentration, as measured by fluorometry using Picogreen (Invitrogen) widely varied among specimens, ranging from 0.3 to 118 ng/μL, allowing us to assess the robustness of our protocol with regard to DNA concentration. Digestion and P1 ligation were performed as described in the step by step protocol provided in the supplementary material 3. Notably, the P1 adaptor was always in large excess (from 40 to 5000 X) relative to its genomic targets. Although some authors recommend a lower ratio of adaptor to genomic DNA (Peterson et al. 2012), presumably to avoid the excessive formation of P1–P1 dimers, we intentionally used a larger ratio to test the robustness of our approach to variation in genomic DNA concentration. To assess the efficiency of the two types of adaptors (classic, phosphorylated, versus new, unphosphorylated) we tested each of the *D. melanogaster* templates with the two types of

adaptors (as well as some specimens of other species, see supplementary 4 for details). To assess the repeatability of the experiment, two *D. melanogaster* samples were also replicated, that is, tagged with two distinct MIDs. The experiment thus includes a total of 16 *D. melanogaster* samples: 2 adaptor types × (2 replicated templates + 4 non-replicated templates).

### Library preparation and sequencing

Library preparation, starting from P1-ligated DNA, and library sequencing, were performed by the ProfileXpert facility in Lyon. The TruSeq library preparation was slightly modified from standard Illumina protocols. Namely, (1) samples were first pooled by species, purified using AmpureXP beads and DNA concentration was estimated. (2) One of our objectives was to test if the sequencing yield was robust to variation in the initial DNA concentration; we thus only partially standardized

concentrations in the pooled library, which resulted in a 55-fold variation between the most and least abundant genomic extract in the final pool. (3) Fragmentation by sonication, repair ends, size selection by gel excision, 3'-dA overhang addition and ligation to P2 adapters were performed following Illumina's recommendations. Finally (4), ligated fragments were amplified by PCR for 14 cycles from 10  $\mu$ L of template and a single purification using AmpureXP beads was performed. 5 % of phiX DNA was added to the final library to facilitate cluster detection during the sequencing reaction.

### Data processing

We used the Stacks software pipeline (version 1.11) to process the data (Catchen et al. 2013). The *process\_radtags* program was used to assign reads to specimens (allowing no more than 1 mismatch in the MID) and to eliminate poor quality reads as well as reads devoid of the expected SbfI cut site after the MID sequence (options `-barcode_dist 2 -q -e sbfI`). For the *D. melanogaster* specimens, we used *Ustack* to produce consensus sequences of RAD tags (options `-m 3` and `-M 4`). The consensus of each RAD tag (each stack) was then mapped to the *D. melanogaster* reference genome (dm3, Apr. 2006 assembly obtained from <http://genome.ucsc.edu/>) using BWA version 0.6.2 with default parameters (Li and Durbin 2009).

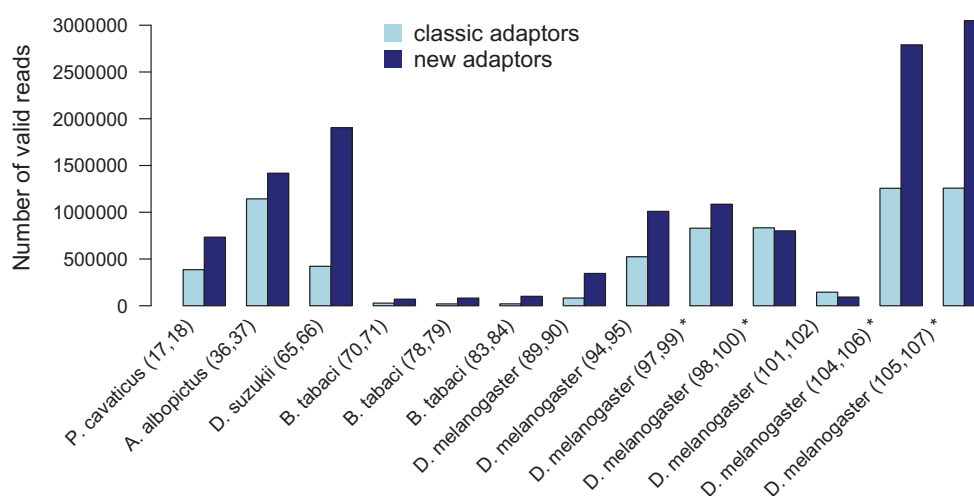
### Results and discussion

One lane of HiSeq 2000 Illumina sequencing produced 192 million of single reads of 51 bp. 76 % of the reads passed the

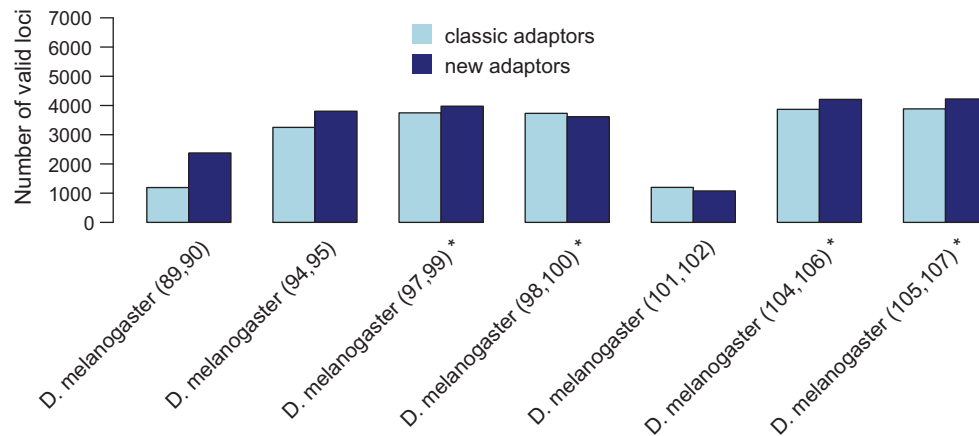
default quality filters and contained an identifiable MID (with no more than one mismatch). 58 % of those started with the expected SbfI overhang and were thus considered as valid for further analysis (see below for a solution to increase the proportion of valid reads). The *D. melanogaster* samples provide us with a mean to control the quality of the data. 92 % of the valid reads from these specimens mapped to RAD loci expected from the reference genome. Difference in the number of valid reads between two replicates varied from 0.1 to 26 %. Considering all samples, the average number of valid reads per specimen was 855,000, but varied from 20,120 to 13,930,998. As expected, a large part of this variation (48 %) was explained by the initial DNA concentration of each sample (ANOVA model: number of reads  $\sim$  concentration + type of adaptor, with square root of log transformation). Residual variation can be explained by variation among genomes in the density of cut sites. In particular, slight variation in GC content can have large impact on the density of restriction sites. For example a shift from 40 to 45 % GC induces a twofold increase in cut site density.

To compare the yield of our new adaptors (devoid of a 5' phosphate and thus supposed to prevent the formation of P1 dimers) to that of classic adaptors, we used 11 DNA templates that were tagged with the two types of adaptors (two of which were replicated), making 13 possible comparisons between classic and new adaptors. In average, the templates ligated to the new adaptors produced 2.4 times more valid reads. Figure 2 provides the detailed data for the 13 comparisons, which clearly indicates that the avoidance of P1 dimers improves the yield of the reaction ( $p = 0.001$ , paired Wilcoxon signed rank test).

To further assess the benefit of using unphosphorylated P1 adaptors, we used data from the *D. melanogaster* controls to count the number of RAD loci that mapped to the



**Fig. 2** Comparison of the classic and new adaptors, in number of valid reads per replicate. Sample ids are given in parenthesis after species name. \*Experimental replicates



**Fig. 3** Comparison of the classic and new adaptors, in number of loci per *D. melanogaster* replicate. Here only loci that mapped to the *D. melanogaster* reference genome were included. Sample ids are given in parenthesis after species name. \*Experimental replicates

reference *D. melanogaster* genome (Fig. 3). The number of valid loci was significantly larger with unphosphorylated adaptors ( $p = 0.04$ , paired Wilcoxon signed rank test, 12 % more loci in average). (see table in supplementary material 5 for details). Notably, we did not observe P1–P1 dimers in our sequencing results, suggesting that such dimers are either eliminated by purification or do not get sequenced. In other words, the benefit of using unphosphorylated P1 adaptors appears to stem mainly from a more efficient ligation.

In contrast, we found that P1–P2 dimers represented a substantial proportion (52 %) of the “non valid reads” (devoid of the expected SbfI overhang), suggesting that free P1s remained in the solution and ligated non specifically to P2 adaptors. In an attempt to solve this problem, we added a second AMPure purification step after PCR amplification in a subsequent library preparation. This extra step increased the proportion of usable reads to much larger values (82 % on average, details not shown).

In conclusion, our experiment demonstrated that numerous distinct P1 adaptors can be obtained at reasonable cost if one synthesizes separately the constant and unique part of the adaptors, and that such “home made” adaptors are fully efficient. In addition, we showed that unphosphorylated P1 adaptors are more efficient than classic adaptors, providing more reads and more RAD tags. This likely results from the absence of P1–P1 dimers, making the ligation step more efficient, and also more robust to variation in the P1 to genomic DNA ratio. The modifications we propose allow one to envisage the preparation of highly multiplexed RADseq libraries, which will prove more and more useful as sequencing technologies increase the number of reads obtained in a single sequencing reaction.

**Acknowledgments** This work was funded by the Centre National de la Recherche Scientifique (ATIP Grant to SC) and the Agence Nationale de la Recherche (Grant ClimEvol). GT is the recipient of a Ph.D. studentship from the Rhône-Alpes region (“Program Cible” Grant). We would like to thank the two anonymous reviewers for their critical assessment of our work.

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Baird N, Etter P, Atwood T et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376. doi:10.1371/journal.pone.0003376
- Cariou M, Duret L, Charlat S (2013) Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecol Evol* 3(4):846–852. doi:10.1002/ece3.512
- Catchen J, Hohenlohe PA, Bassham S et al (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* 22:3124–3140. doi:10.1111/mec.12354
- Davey JW, Blaxter ML (2011) RADSeq: next generation population genetics. *Brief Funct Genomics* 9:416–423. doi:10.1093/bfpg/eq031
- Etter PD, Bassham S, Hohenlohe PA et al (2011) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In: Orgogozo V, Rockman MV (eds) *Molecular methods for evolutionary genetics*. Humana Press, Totowa, pp 157–178
- Frank DN (2009) BARCRAWL and BARTAB: software tools for the design and implementation of barcoded primers for highly multiplexed DNA sequencing. *BMC Bioinformatics* 10:362. doi:10.1186/1471-2105-10-362
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760. doi:10.1093/bioinformatics/btp324
- Peterson BK, Weber JN, Kay EH et al (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7:e37135. doi:10.1371/journal.pone.0037135
- Sambrook J, Russel D (2001) Commonly used techniques in molecular cloning. In: Sambrook J, Russel D (eds) *Appendix 8, in molecular cloning, vol 3, 3rd edn*. Cold Spring Harbor Laboratory Press, New York





# Annexes

A	Plantes hôtes pour <i>B. tabaci</i>	221
B	Dynamique des clades à l'échelle du bassin méditerranéen	223
C	Comparaison de traits d'histoire de vie des cytotypes Q1CH et Q1RH face à des traitements insecticides et thermiques	225
D	Protocole d'extraction d'ADN pour la préparation d'échantillons <i>RADSeq</i>	227
E	Chevauchement des <i>read</i>	229
F	Analyse des témoins drosophiles	231
G	Résoudre le problème d'augmentation de la densité de SNP en fin de locus	233
H	Analyse en composante principale des données de <i>RADSeq</i>	235



## Plantes hôtes pour *B. tabaci*

Liste non exhaustive de plantes hôtes pour *B. tabaci*. Seuls les genres sont indiqués, regroupés par famille, elles-mêmes triées par ordre alphabétique. Données issues de *Genbank*, février 2016.

**Acanthaceae** : *Dicliptera, Hypoestes, Ruellia*

**Aizoaceae** : *Trianthema*

**Amaranthaceae** : *Achyranthes, Chenopodium, Spinacia*

**Apiaceae** : *Apium*

**Apocynaceae** : *Adenium, Mandevilla*

**Araceae** : *Anthurium, Xanthosoma*

**Asparagaceae** : *Asparagus*

**Asteraceae** : *Acanthospermum, Ageratina, Ageratum, Aspilia, Bellis, Bidens, Chromolaena, Chrysanthemum, Conyza, Crepidiastrum, Echinacea, Eclipta, Emilia, Erigeron, Eupatorium, Gaillardia, Gerbera, Helianthus, Ixeris, Lactuca, Laggera, Paneroa, Siegesbeckia, Sonchus, Synedrella, Tithonia, Wedelia, Xanthium, Youngia, Zinnia*

**Begoniaceae** : *Begonia*

**Boraginaceae** : *Heliotropium*

**Brassicaceae** : *Brassica, Raphanus*

**Cannabaceae** : *Humulus, Trema*

**Caprifoliaceae** : *Lonicera*

**Cleomaceae** : *Cleome*

**Commelinaceae** : *Commelina*

**Convolvulaceae** : *Argyreia, Calystegia, Convolvulus, Ipomoea*

**Cucurbitaceae** : *Benincasa, Citrullus, Cucumis, Cucurbita, Lagenaria, Luffa, Momordica, Seschium, Trichosanthes*

**Ericaceae** : *Calluna, Rhododendron*

**Euphorbiaceae** : *Cassava, Codiaenum, Euphorbia, Jatropha, Manihot, Poinsettia, Ricinus*

**Fabaceae** : *Arachis, Bauhinia, Cyamopsis, Erythrina, Glycine, Lablab, Leucaena, Medicago, Mucuna, Phaseolus, Sesbania, Trigonella, Vicia, Vigna*

**Gentianaceae** : *Eustoma*

**Hydrangeaceae** : *Hydrangea*

**Lamiaceae** : *Ajuga, Caryopteris, Lamium, Leonotis, Leonurus, Melissa, Mentha, Mesona, Ocimum, Perilla, Phlomis, Platostoma, Salvia, Thymus*

**Malvaceae** : *Abelmoschus, Alcea, Althaea, Corchorus, Gossypium, Hibiscus, Malva, Malvatum, Pavonia, Sida, Urena*

**Moraceae** : *Broussonetia, Ficus, Morus*

**Myricaceae** : *Myrica*

**Oleaceae** : *Jasminum*

**Oxalidaceae** : *Oxalis*

**Passifloraceae** : *Passiflora*

**Phyllanthaceae** : *Sauropus*

**Plantaginaceae** : *Veronica*

**Portulacaceae** : *Portulaca*

**Rosaceae** : *Rosa, Rubus*

**Rubiaceae** : *Pentas*

**Rutaceae** : *Citrus*

**Solanaceae** : *Capsicum, Datura, Nicotiana, Physalis, Solanum*

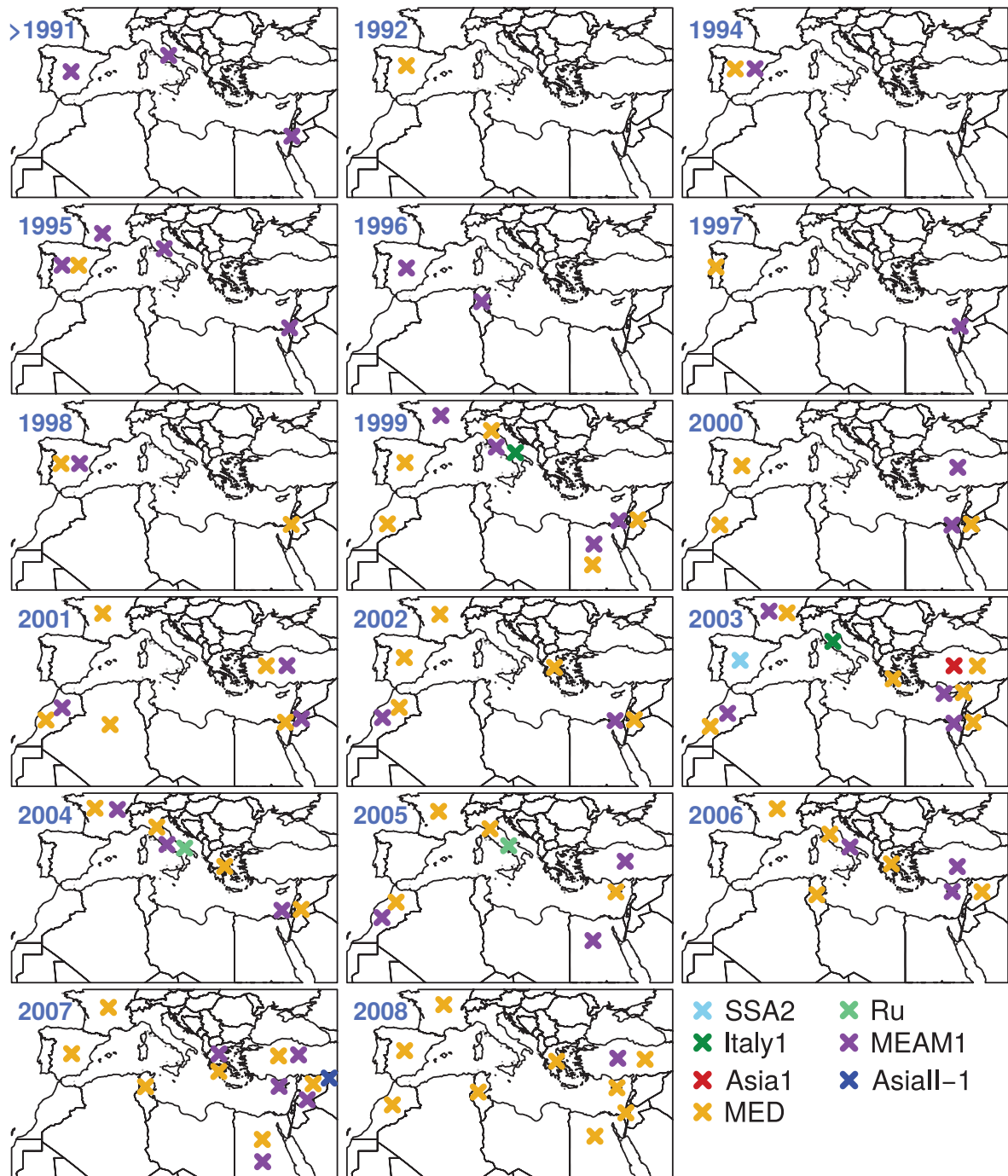
**Tropaeolaceae** : *Tropaeolum*

**Urticaceae** : *Boehmeria, Urtica*

**Verbenaceae** : *Duranta, Lantana, Verbena*

**Vitaceae** : *Vitis*

## Dynamique des clades à l'échelle du bassin méditerranéen

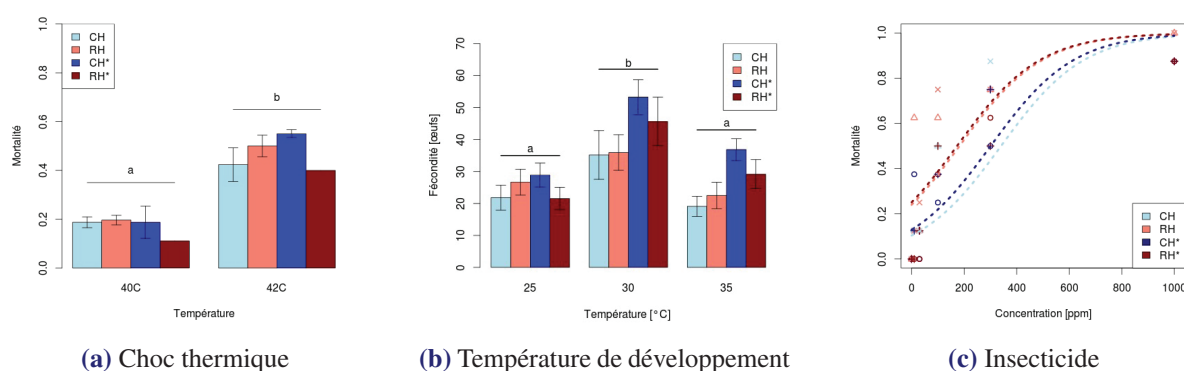






## Comparaison de traits d’histoire de vie des cytotypes Q1CH et Q1RH face à des traitements insecticides et thermiques

Nous avons testé ces trois facteurs en utilisant des F1 (pour éviter le facteur confondant du fond nucléaire) issues de croisement d’une lignée Q1CH et d’une lignée Q1RH. i) La survie a été mesurée après exposition à un choc thermique de 40 et 42°C ii) La fécondité a été mesurée en faisant varier la température de développement (25, 30 et 35°C) iii) La survie a été mesurée après exposition à la formulation commerciale d’un insecticide de la famille des pyréthriinoïdes, le Decis. Les résultats sont présentés sur la figure A.1.



**FIGURE A.1** – Traits mesurés sur les F1 des croisement Q1CH \* Q1RH. À gauche, survie mesurée après le choc thermique ; au centre : fécondité mesurée à une température de 25, 30 et 35 °C ; à droite : mortalité mesurée 48 h après l’exposition à l’insecticide. Les hybrides sont indiqués avec une astérisque, les autres sont les lignées parentales témoins.

Les résultats montrent que les deux lignées parentales répondent de la même façon aux effets thermiques, que ce soit un choc ou le développement à différentes températures. Par contre nous voyons que les deux types d’hybrides sont plus résistants dans le cas du choc thermique, probablement grâce à un effet hétérosis. Ces résultats pourraient donc suggérer que la température n’influence pas la compétition inter-hôte dans le cas d’une population mixte Q1RH-Q1CH.

Par contre un effet du cytotype a été mis en évidence sur la capacité à résister au Decis. Les hybrides de cytotypes RH sont plus sensibles que les hybrides CH ( $p < 0,5$ ). Le même résultat est visible sur les lignées parentales ( $p < 0,5$ ). Les individus infectés par *Rickettsia* et *Hamiltonella* sont plus sensibles que les individus infectés par *Cardinium* et *Hamiltonella*.

La sensibilité à l’insecticide varie aussi entre les hybrides ( $glm, p = 0,01$ ), ainsi les hybrides RH sont plus sensibles à l’insecticide que les hybrides CH. Les doses létales 50 sont respectivement de  $252 \pm 24$  et  $380 \pm 75$ . La même tendance est visible pour les lignées parentales : RH ( $glm, p = 0,01$ ), est plus sensibles que la lignée CH.

Même si on ne peut démêler l’effet de *Cardinium* et de *Rickettsia* (est ce que l’un augmente la susceptibilité, où bien l’autre la diminue ?), on peut dire que l’association prise comme un tout est plus ou moins sensible, en l’occurrence le cytotype Q1RH paraît plus sensible.



# Protocole d'extraction d'ADN pour la préparation d'échantillons *RADSeq*

## Extraction Phénol-Chloroforme + kit *NucleoSpin*<sup>®</sup> *Tissue XS* (MACHEREY-NAGEL)

Pour le broyage à sec, mettre 10 mn au congélateur -80 °C, puis les bêtes sont broyées dans la barrette avec une bille de 1,4 mm de diamètre, dans le TissueLyser (1 mn, 30 Hz).

- Ajouter 100  $\mu$ l de tampon de broyage à l'individu broyé.
- Ajouter 2  $\mu$ l de protéinase K et incubé 2 heures à 55 °C (bain-marie).
- Ajouter 0,5  $\mu$ l de RNase A (Fermentas, 10 mg/ml) par échantillon et laisser incubé 30 mn à 37 °C (thermomixer). Placer à 4 °C (glace).
- Ajouter 12  $\mu$ l de KO Acétate 5 M.
- Placer à 4 °C pendant 15 mn.
- Centrifuger à 4 °C et à 12 000 rcf pendant 15 mn. Récupérer le surnageant.
- Ajouter 100  $\mu$ l de phénol (utiliser sous la hotte, pipeter au fond du flacon), vortexer chaque tube pendant 15 s puis centrifuger à 4 °C à 12 000 rcf pendant 5 mn.
- Récupérer le surnageant et ajouter 100  $\mu$ l de phénol (solution à température ambiante), vortexer pendant 15 s puis centrifuger à 4 °C à 12 000 rcf pendant 5 mn.
- Récupérer le surnageant, puis passer au kit tissue XS à partir de l'étape 4 « Adjust DNA binding conditions » avec 100  $\mu$ l d'éthanol (Voir page suivante).

---

**4 Adjust DNA binding conditions**

Add **80 µL ethanol (96–100 %)** to the lysate and mix by vortexing 2 x 5 s.

Spin down briefly to clear the lid.



**+ 80 µL ethanol**

---

**5 Bind DNA**

For each sample, place one **NucleoSpin® Tissue XS Column** into a **Collection Tube (2 mL)**. Apply the sample to the column. Centrifuge for **1 min** at **11,000 x g**. Discard the flow-through and place the column into a new Collection Tube (2 mL).

*If the sample is not drawn completely through the matrix, repeat the centrifugation step at 11,000 x g.*



**Load lysate**



**11,000 x g,  
1 min**

---

**6 Wash silica membrane****1<sup>st</sup> wash**

Add **50 µL Buffer B5** to NucleoSpin® Tissue XS Column. Centrifuge for **1 min** at **11,000 x g**. It is not necessary to discard the flow-through. Reuse the Collection Tube.



**+ 50 µL B5**



**11,000 x g,  
1 min**

**2<sup>nd</sup> wash**

Add **50 µL Buffer B5** to the NucleoSpin® Tissue XS Column. Centrifuge for **2 min** at **11,000 x g**. Discard Collection Tube with flow-through.



**+ 50 µL B5**



**11,000 x g,  
2 min**

---

**7 Elute DNA**

Place the NucleoSpin® Tissue XS Column in a new 1.5 mL microcentrifuge tube (not provided) and apply **20 µL Buffer BE** directly onto the center of the silica membrane of the column. Centrifuge for **1 min** at **11,000 x g**.

*Elution volume may be varied from approximately 5–30 µL. For a correlation of elution volume, DNA concentration and DNA amount eluted from the column see section 2.4–2.5.*



**+ 20 µL BE**



**11,000 x g,  
1 min**

---

**8 Optional: Remove residual ethanol**

Incubate elution fraction with open lid for **8 min** at **90 °C**.

*See section 2.5 for further comments and alternative incubation times and temperatures for a removal of residual ethanol.*



**Optional:**

**8 min,  
90 °C**

---

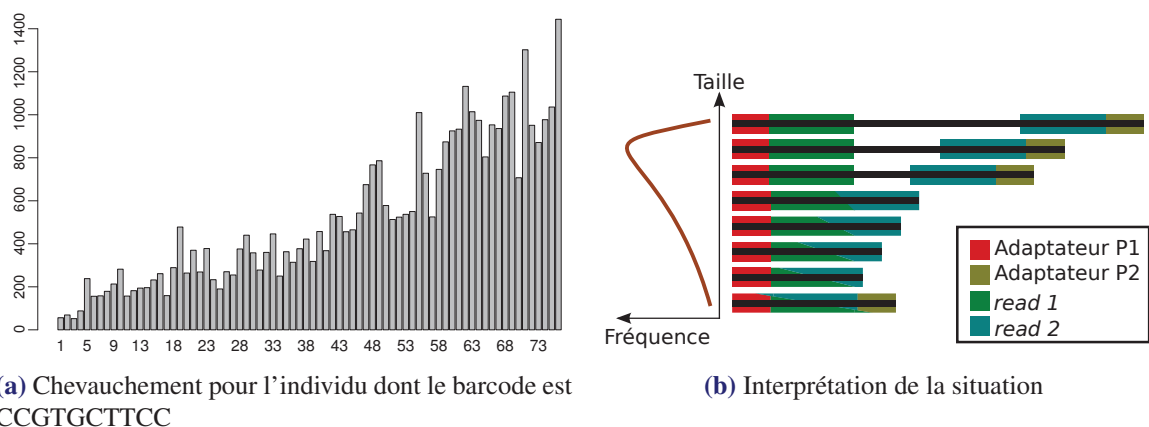
A la fin du protocole, reprendre dans 40µl d'eau ultra pure (Fermentas). Elution finale : centrifuger 1 mn 30

## Chevauchement des *read*

Lorsqu'on prend le reverse complément des 10 dernières bases du *read 2* des données brutes, un peu moins de 10 % ont une correspondance sur le *read 1* (543 089 sur 6 000 000 de lectures prises au hasard), dont au moins 20 % correspondent exactement à de l'adaptateur P1. En répétant l'opération avec les lectures des deux drosophiles, on obtient respectivement 30 000 sur 393 000 et 53 000 sur 671 000, ordre de grandeur comparable aux données de *B. tabaci* (chiffres non présentés ici), ce qui semble indiquer un problème non spécifique à *B. tabaci*.

La position de ce chevauchement sur le *read 1* a été mesurée (voir figure A.2). Sur cette figure est représentée la fréquence de chevauchements entre les deux *read* pour chaque position dans le *read 1*. L'individu représenté a été choisi au hasard, mais le patron est globalement le même pour les autres. Des chevauchements commençant à la position 1 indique un chevauchement total entre les deux *read*, plus le chevauchement commence tard dans la séquence, plus celui-ci est faible. On voit ainsi que le nombre de chevauchement par position augmente avec la position dans la séquence. Dit autrement, la majorité des chevauchements est très faible, et pour quelques uns, les deux *read* se chevauchent entièrement.

Ce problème pourrait être expliqué par une taille de fragments d'ADN trop courte lors de la préparation de la banque, ainsi les deux *reads* se chevauchent au lieu d'être distants d'environ 200 bases, mais ceci n'explique pas le problème de qualité. Par contre si une grosse proportion des fragments est identique à chaque position, la machine ne peut séquencer correctement car le signal est identique partout et renvoie donc des bases indéterminées (N). On peut imaginer qu'une bonne partie des fragments soient de la taille minimum acceptable par le séquenceur et ainsi lors de la lecture du *read 2*, à partir de la 20<sup>ème</sup> base, elles sont identiques pour beaucoup de fragments, engendrant des problèmes de lecture.



**FIGURE A.2** – Phénomène de chevauchement des deux *reads*. À gauche, fréquence des chevauchements pour chaque position pour un individu représentatif du phénomène général. À droite, schéma expliquant cette situation avec des fragments de plus en plus petit de haut en bas : les trois premiers sont de bonne taille, mais les autres sont trop petits et les deux *reads* se chevauchent. La courbe de fréquence n'est pas réelle, ce n'est qu'une hypothèse.





## Analyse des témoins drosophiles

Deux drosophiles avaient été insérées dans cet essai, plus précisément ce sont deux répliques d'un même individu avec deux barcodes différents. On s'attend donc à trouver le même nombre de locus pour ces deux individus si nous sommes répétables, ce nombre étant connu puisqu'il y a un génome de référence. Cette analyse a été réalisée avec STACKS (paramètres : -m 2 -M 6 -H -p 1 -r -d) pour une question de rapidité et de facilité de mise en œuvre.

Le nombre de locus trouvés est de 4 234 et 4 595 pour respectivement 350 000 et 620 000 lectures. Ainsi, pour quasiment deux fois plus de lectures, on obtient seulement 350 locus en plus. La couverture moyenne est respectivement de 88x et 140x. Les consensus des locus ont ensuite été mappés sur le génome de référence avec BWA (Li & Durbin 2009) en utilisant les paramètres par défaut. 3 806 et 3 993 locus respectivement ont un *hit* sur le génome, représentant 3 761 et 3 917 « vrais » locus. En effet, plusieurs locus créés par STACKS peuvent en fait provenir d'un même et vrai locus, artificiellement éclaté comme nous l'avions dit. Dans une deuxième étape, ce sont les lectures qui ont directement été mappées (voir tableau A.1). Le nombre de locus obtenu est supérieur au nombre obtenu par STACKS et approche les 80 % de locus retrouvés.

La couverture observée est largement inférieure à celle de l'essai pilote qui était d'environ 250x, on a donc plus d'information avec moins de lectures ce qui est un très bon résultat.

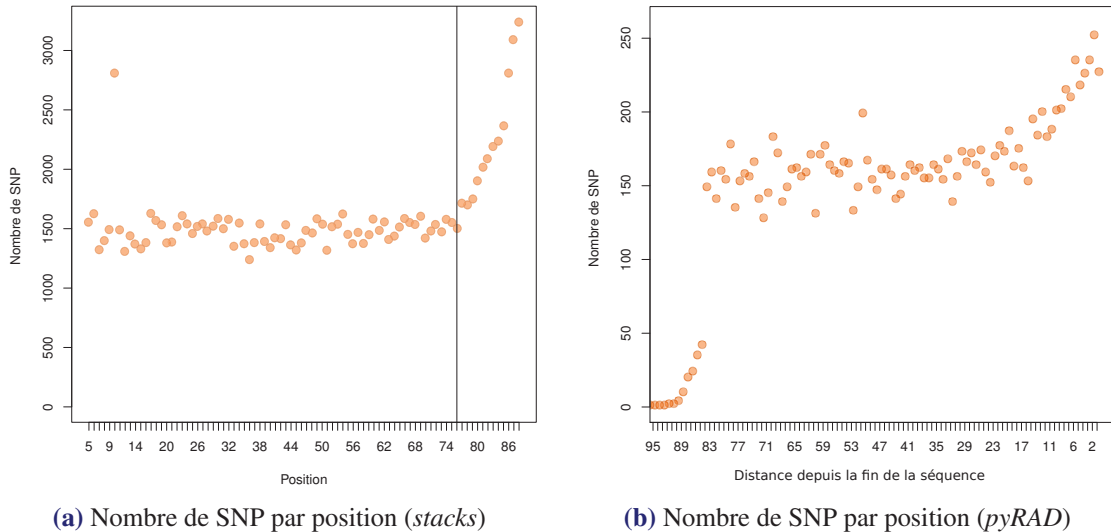
Individu	<i>read 1</i>			<i>reads 2</i>	
	Lectures	Locus	Couverture	Lectures	Locus
CACAATCCCC	619 476	4 717	130x	518 386	55 968
ATGAGCGC	353 892	4 338	80x	297 890	31 999

**TABLE A.1** – Résultats du *mapping* des séquences des deux drosophiles. Pour chaque *reads*, le nombre de lectures qui ont une similarité est indiqué, puis le nombre de position (ou locus) que cela représente. Le nombre de locus formés par le *read 2* est bien plus important car l'extrémité de celui-ci est formé par cassure aléatoire de l'ADN.



## Résoudre le problème d'augmentation de la densité de SNP en fin de locus

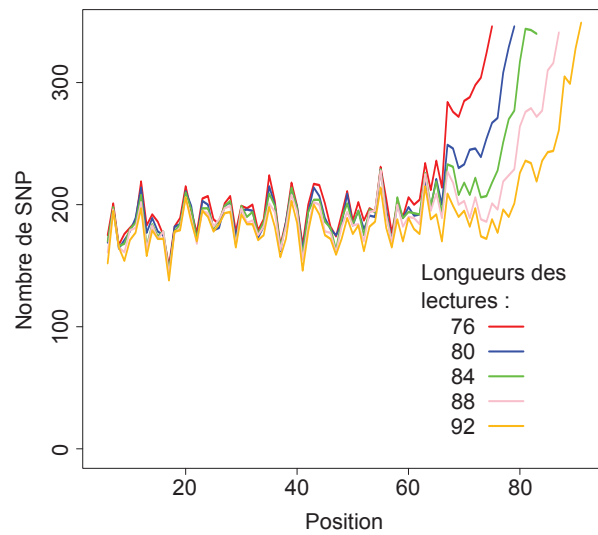
Il s'est avéré que la densité de SNP le long des locus était hétérogène : celle-ci augmente plus ou moins fortement sur la fin (voir figure A.3).



**FIGURE A.3** – Nombre de SNP le long des locus. En a), SNP-calling réalisé avec *stacks*. En b), réalisé avec *pyRAD*, les positions sont ici exprimées en distance depuis la fin de la séquence puisque les locus n'ont pas tous la même taille (il y a des insertions et des délétions).

La première hypothèse avancée pour expliquer ce phénomène est la baisse de qualité observée en fin de séquence (voir figure 4.7a page 146), ce qui est classiquement observé dans la littérature (Eaton & Ree 2013). Si cette hypothèse est la bonne, il suffit de couper les lectures en amont, préalablement au *clustering*, pour écarter la partie de mauvaise qualité, c'est donc ce que nous avons fait, avec les deux programmes : *STACKS* et *pyRAD* (voir figure A.4, en ce qui concerne les résultats obtenus avec *pyRAD*). Sur cette figure est présenté le nombre de SNP à chaque position pour plusieurs longueurs de lectures. On voit que quelle que soit la longueur des lectures, le nombre de SNP augmente en fin de séquence. La qualité des bases ne semble donc pas être en cause. Une deuxième possibilité est la non prise en compte des insertions et délétions par *STACKS* comme nous l'avons vu page 143. D'après cette explication, toute insertion ou délétion en toute fin de séquence peut induire des SNP sur les bases restantes, à condition qu'il n'en reste pas trop (inférieure au nombre total de mésappariements autorisés). Même si cette hypothèse explique en partie ce phénomène, cela ne peut l'expliquer en totalité puisque les mêmes tests conduits avec *pyRAD* montrent les mêmes patrons, bien que l'amplitude du problème soit moindre. Nous n'avons pas d'hypothèses pour l'expliquer dans ce cas.

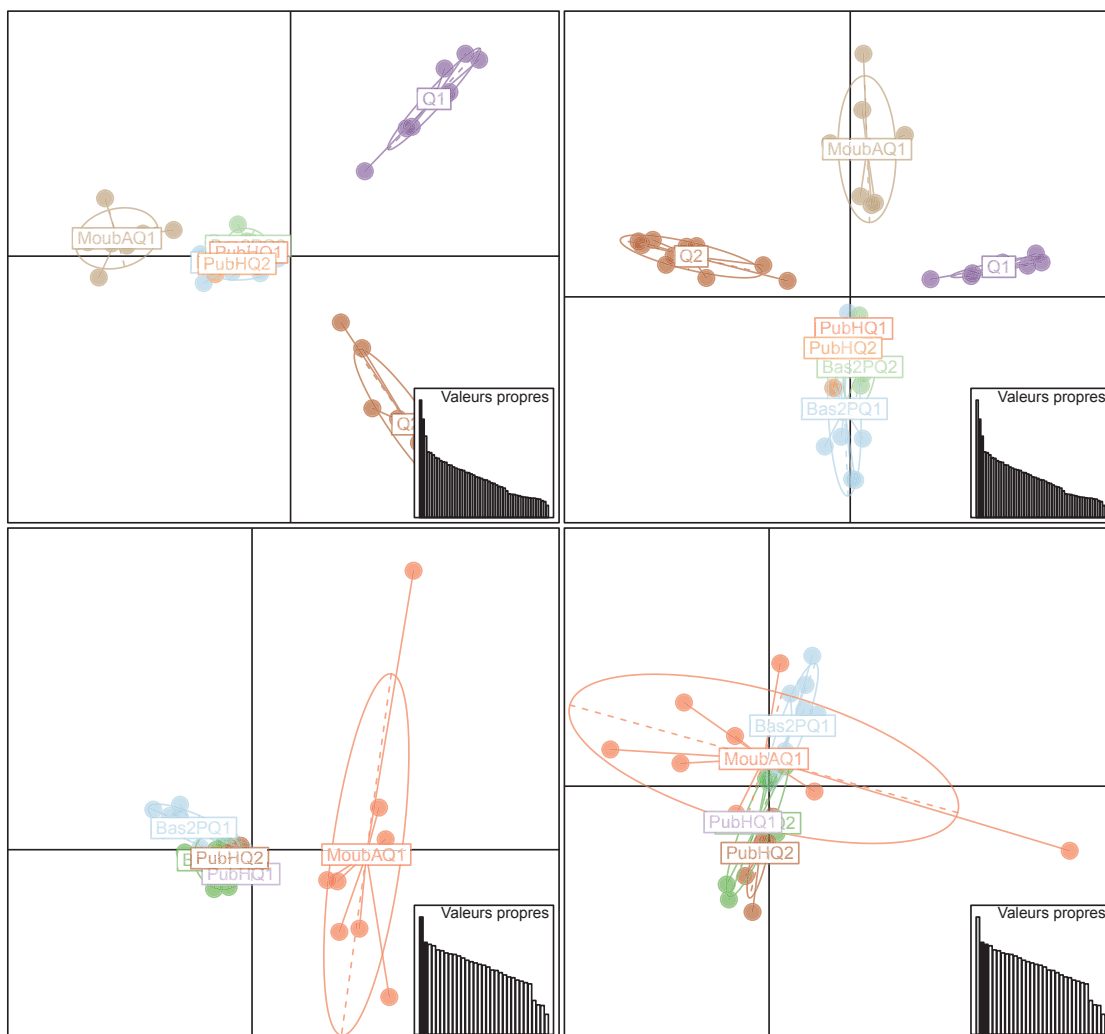
La solution retenue pour résoudre ce problème est de couper les locus, mais seulement à la fin du processus, après le *clustering* et non de couper les séquence en amont. Les SNP dont la position était supérieure à 79 dans la séquence ont donc été écartés suite à différents tests.



**FIGURE A.4** – Densité de SNP en fonction de la position sur le locus, pour plusieurs longueurs de lectures. Les lectures ont été coupées avant de faire le *clustering*, ici avec STACKS. Quelle que soit la longueur, le nombre de SNP augmente en fin de locus. Des valeurs bien plus faibles de longueur de séquence ont été essayées (jusqu’à 30 paires de bases de long) et donnent les mêmes résultats.

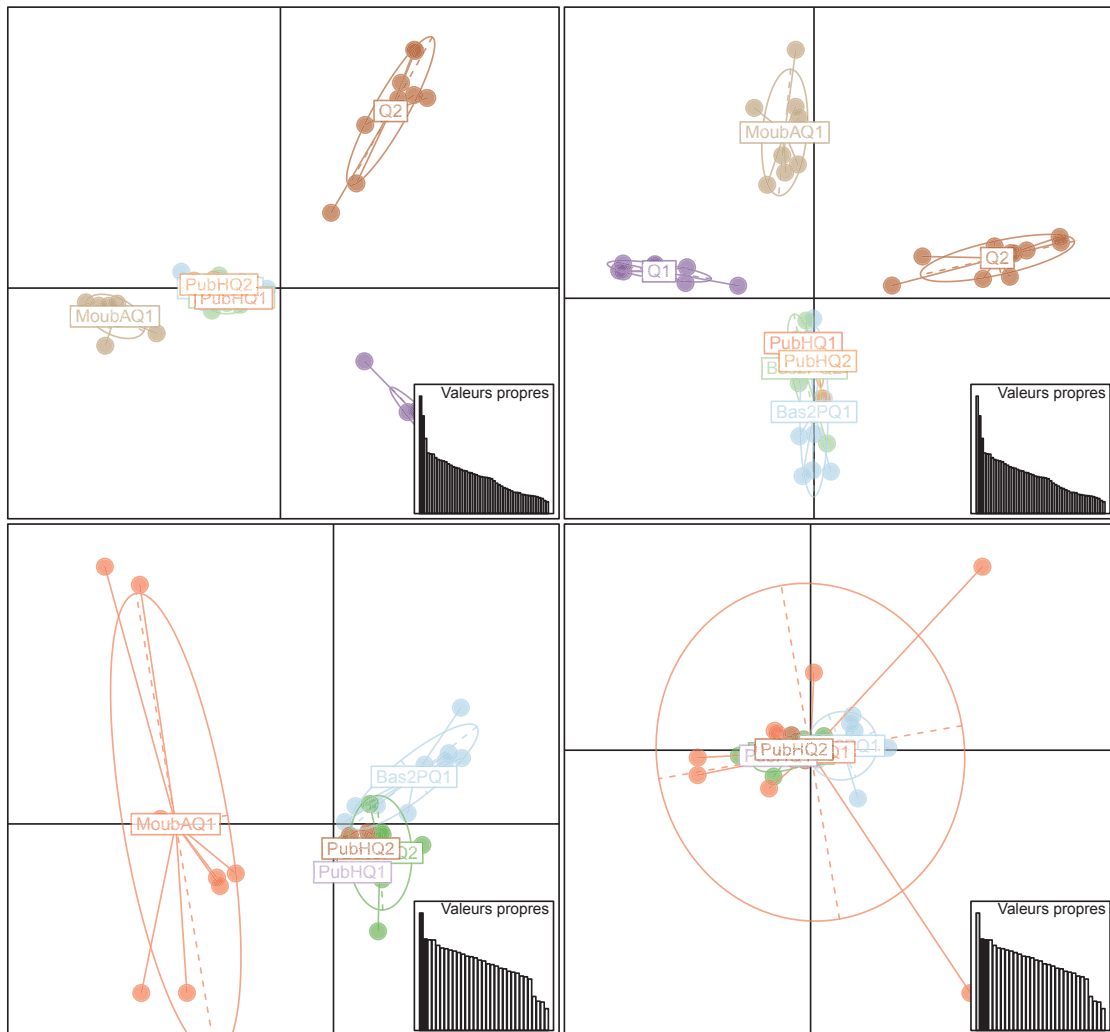
## **Analyse en composante principale des données de *RADSeq***

Pourcentage d'individus minimum par locus : 2

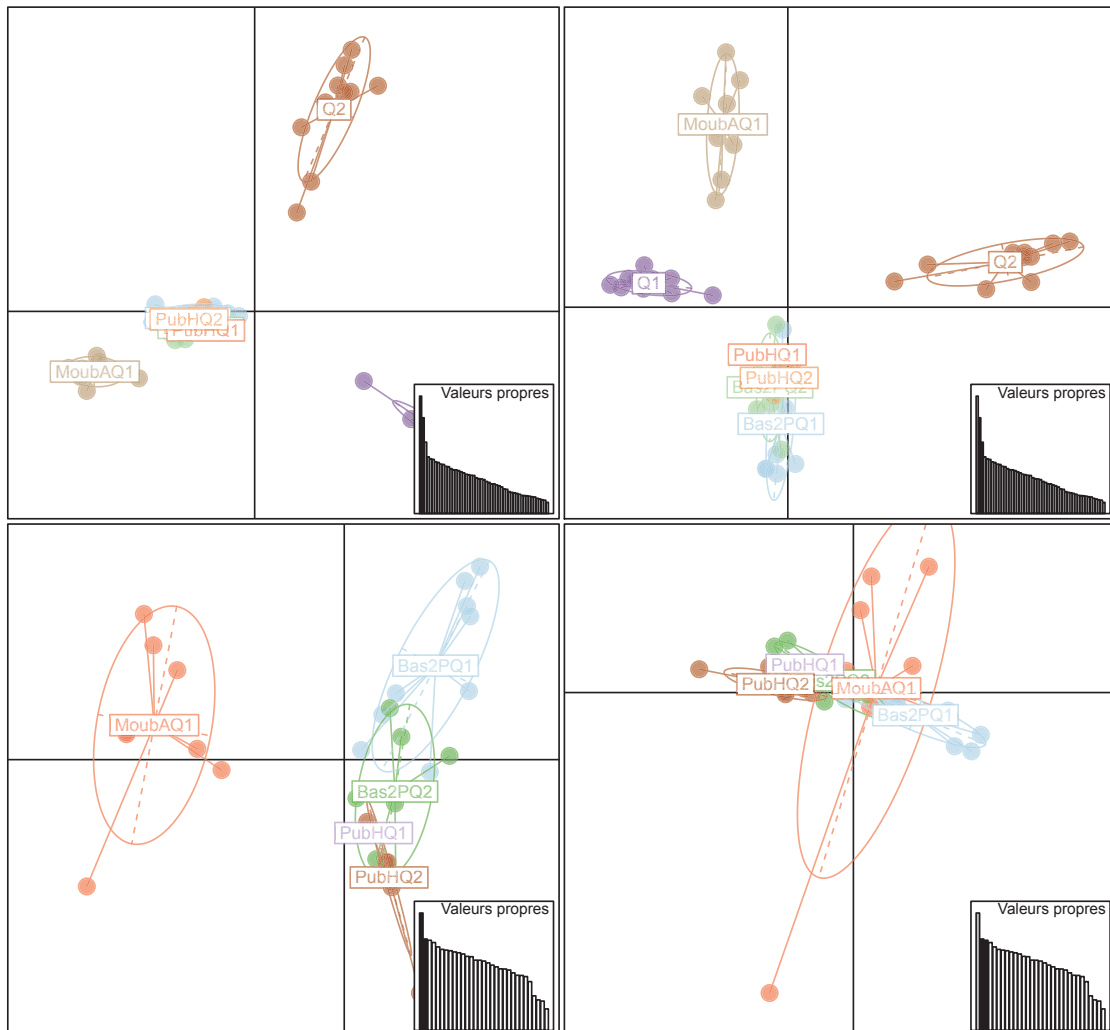




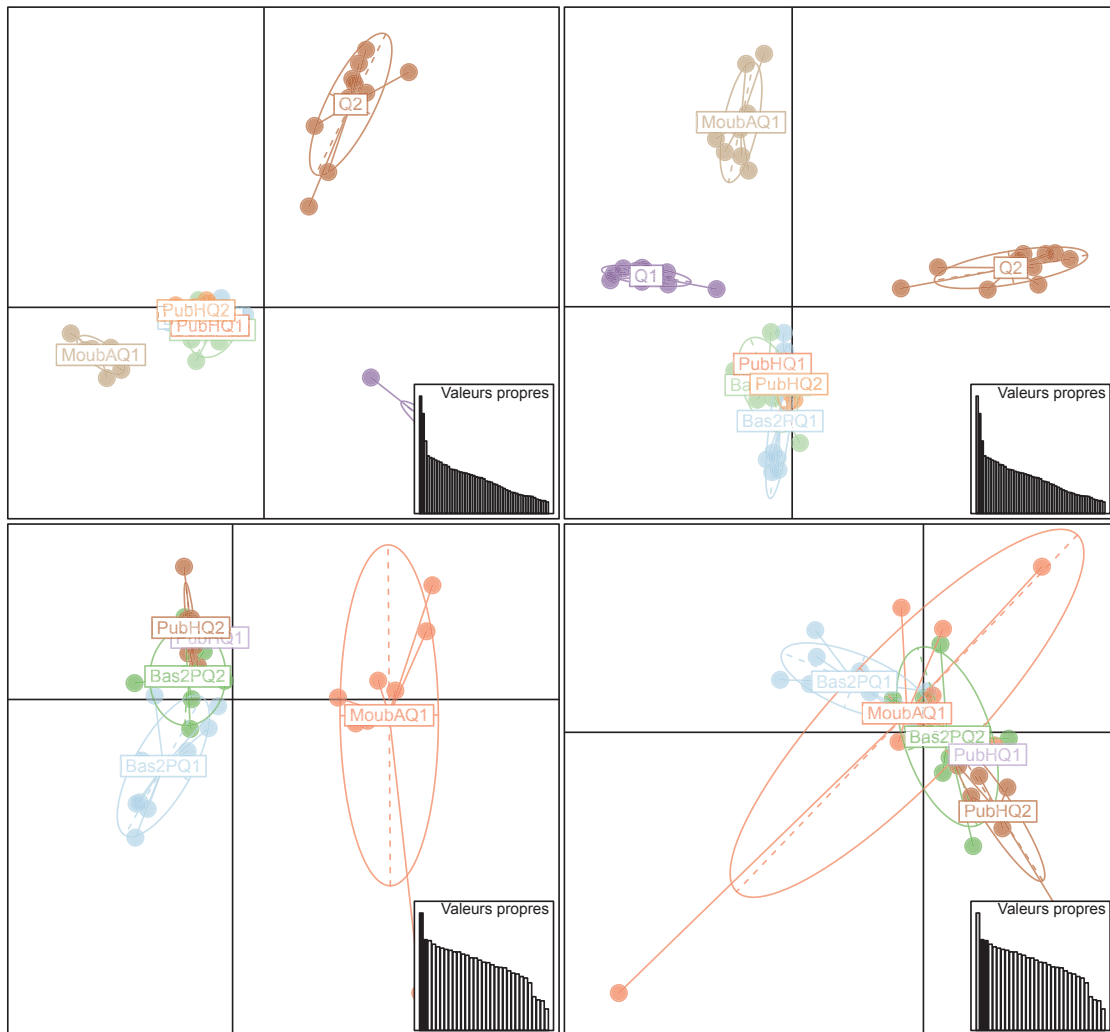
Pourcentage d'individus minimum par locus : 3



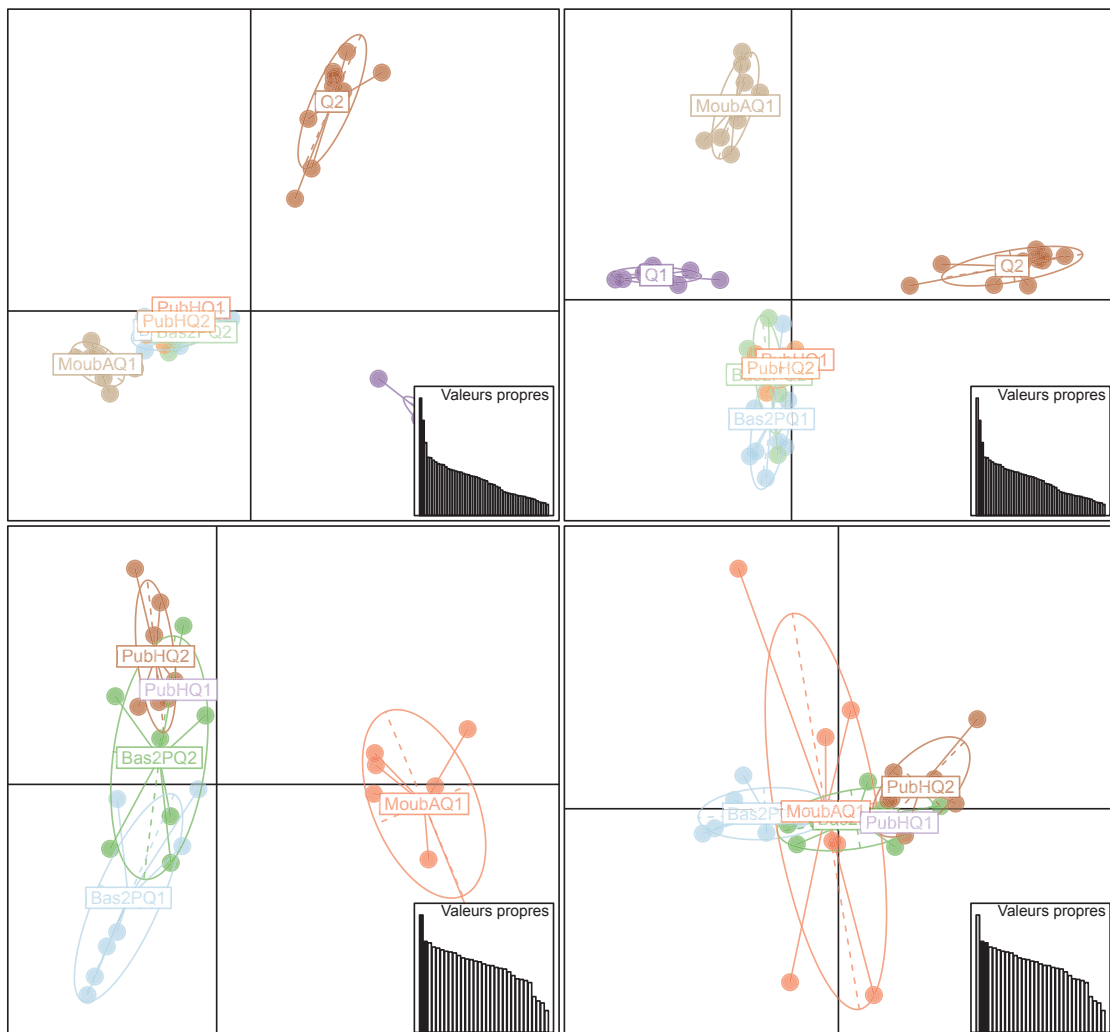
Pourcentage d'individus minimum par locus : 13



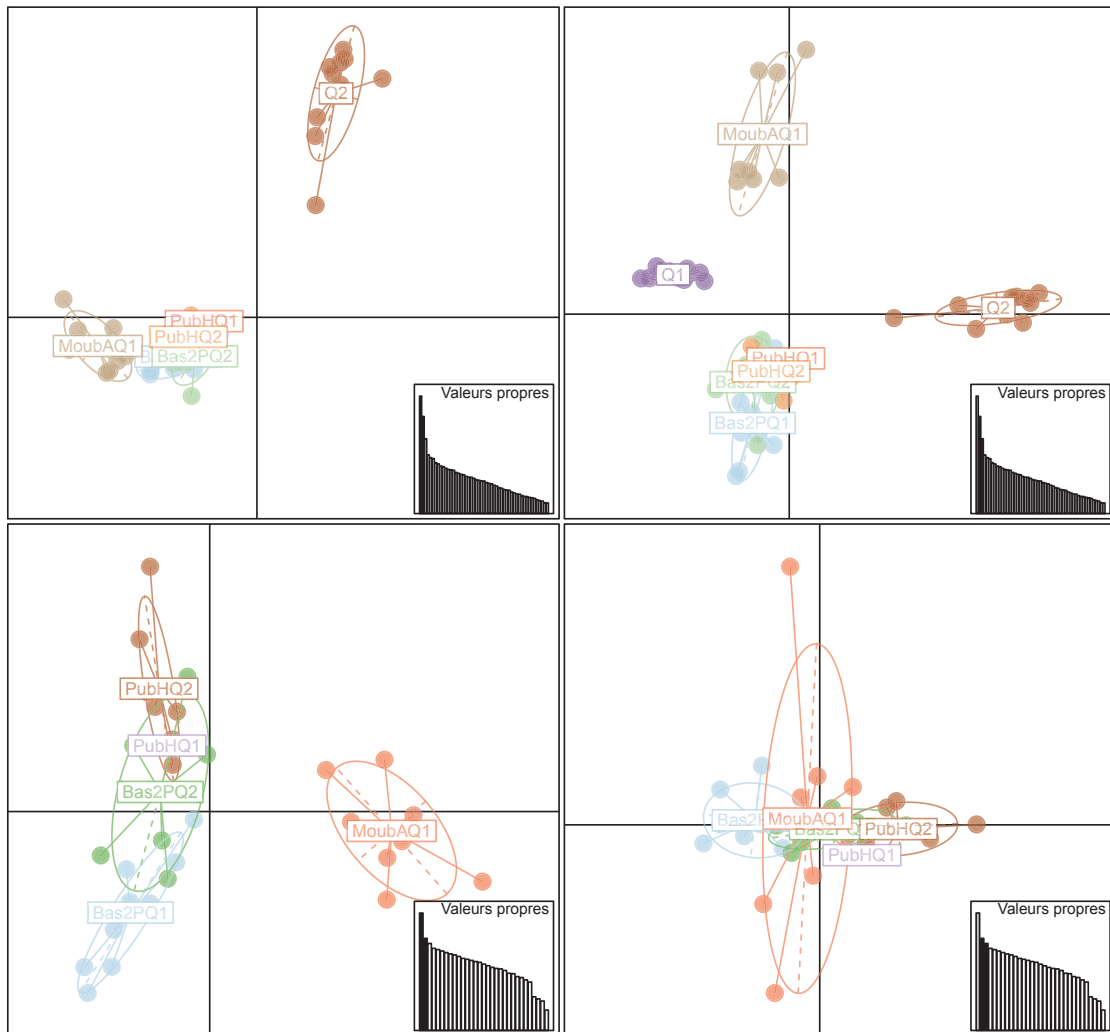
Pourcentage d'individus minimum par locus : 26



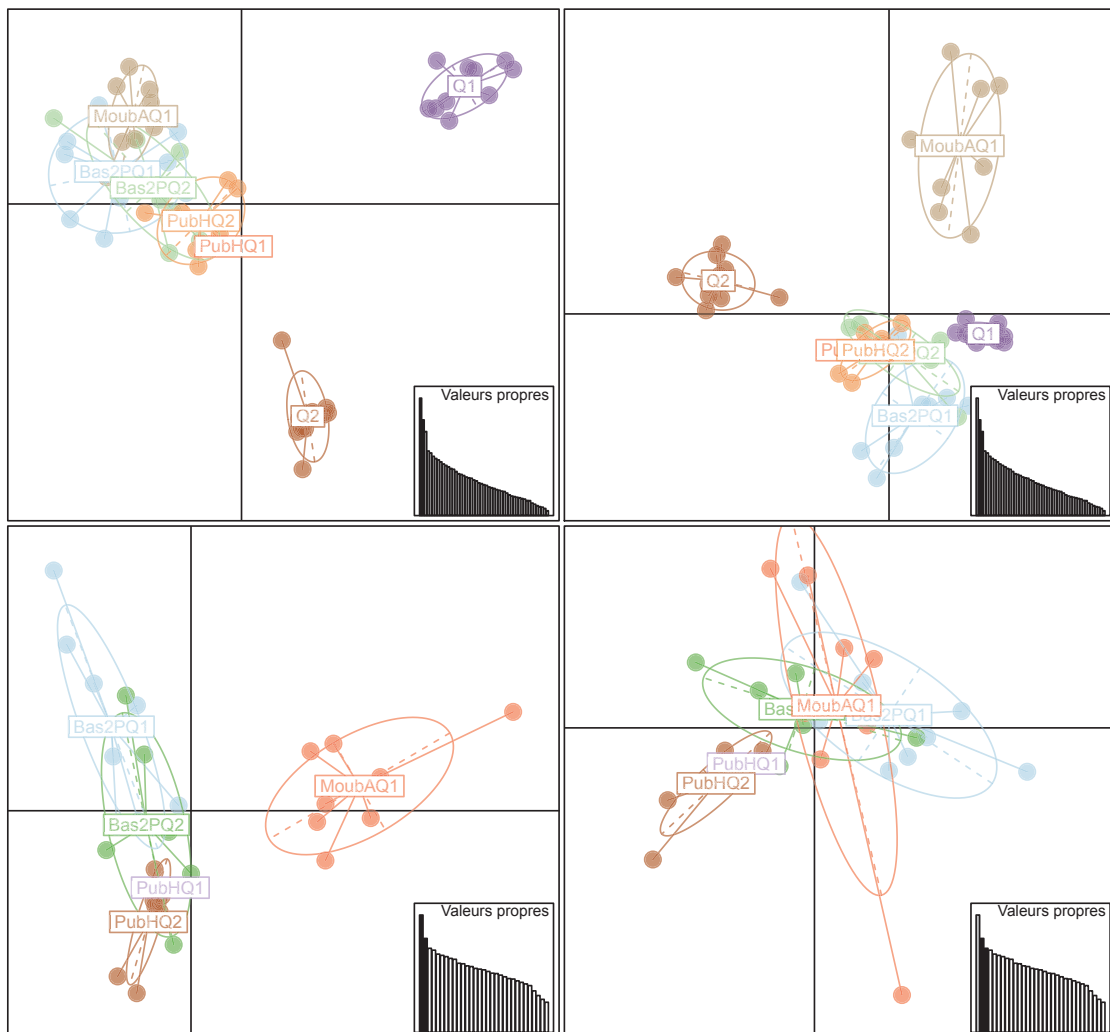
Pourcentage d'individus minimum par locus : 31



Pourcentage d'individus minimum par locus : 36



Pourcentage d'individus minimum par locus : 42





# Références bibliographiques

- Abdullahi I, Winter S, Atiri GI, Thottappilly G (2003) Molecular characterization of whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae) populations infesting cassava. *Bulletin of entomological research*, **93**, 97–106.
- Ahmed MZ, De Barro PJ, Olleka A, *et al.* (2012) Use of consensus sequences and genetic networks to identify members of the *Bemisia tabaci* cryptic species complex in Egypt and Syria. *Journal of Applied Entomology*, **136**, 510–519.
- Ahmed MZ, De Barro PJ, Ren SX, Greeff JM, Qiu BL (2013) Evidence for horizontal transmission of secondary endosymbionts in the *Bemisia tabaci* cryptic species complex. *PLoS One*, **8**, e53084.
- Ahmed MZ, Li SJ, Xue X, *et al.* (2015) The Intracellular Bacterium *Wolbachia* Uses Parasitoid Wasps as Phoretic Vectors for Efficient Horizontal Transmission. *PLoS Pathogens*, **10**, e1004672.
- Ahmed MZ, Ren SX, Xue X, Li XX, Jin Gh, Qiu BL (2010) Prevalence of endosymbionts in *Bemisia tabaci* populations and their in vivo sensitivity to antibiotics. *Current microbiology*, **61**, 322–8.
- Akman L, Yamashita A, Watanabe H, *et al.* (2002) Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nature genetics*, **32**, 402–7.
- Alemandri V, De Barro P, Bejerman N, *et al.* (2012) Species within the *Bemisia tabaci* (Hemiptera: Aleyrodidae) complex in soybean and bean crops in Argentina. *Journal of economic entomology*, **105**, 48–53.
- Alizon S, Hurford A, Mideo N, Van Baalen M (2009) Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *Journal of Evolutionary Biology*, **22**, 245–259.
- Allen JM, Reed DL, Perotti MA, Braig HR (2007) Evolutionary relationships of "*Candidatus* riesia spp.," endosymbiotic Enterobacteriaceae living within hematophagous primate lice. *Environmental Microbiology*, **73**, 1659–1664.
- Alon M, Alon F, Nauen R, Morin S (2008) Organophosphates' resistance in the B-biotype of *Bemisia tabaci* (Hemiptera: Aleyrodidae) is associated with a point mutation in an ace1-type acetylcholinesterase and overexpression of carboxylesterase. *Insect biochemistry and molecular biology*, **38**, 940–9.
- Alon M, Benting J, Lueke B, Ponge T, Alon F, Morin S (2006) Multiple origins of pyrethroid resistance in sympatric biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Insect biochemistry and molecular biology*, **36**, 71–9.
- Anbutsu H, Fukatsu T (2006) Tissue-specific infection dynamics of male-killing and nonmale-killing spiroplasmas in *Drosophila melanogaster*. *FEMS Microbiology Ecology*, **57**, 40–46.
- Arnold J (1993) Cytonuclear Disequilibria in Hybrid Zones. *Annual Review of Ecology and Systematics*, **24**, 521–554.
- Arnold ML (1997) *Natural hybridization and evolution*. Oxford University Press.
- Asmussen MA, Arnold J (1991) The effects of admixture and population subdivision on cytonuclear disequilibria. *Theoretical population biology*, **39**, 273–300.
- Asplen MK, Anfora G, Biondi A, *et al.* (2015) Invasion biology of spotted wing *Drosophila* (*Drosophila suzukii*): a global perspective and future priorities. *Journal of Pest Science*, **88**, 469–494.
- Baird NA, Etter PD, Atwood TS, *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, **3**, e3376.
- Baldo L, Ayoub Na, Hayashi CY, Russell Ja, Stahlhut JK, Werren JH (2008) Insight into the routes of *Wolbachia* invasion: High levels of horizontal transfer in the spider genus *Agelenopsis* revealed by *Wolbachia* strain and mitochondrial DNA diversity. *Molecular Ecology*, **17**, 557–569.
- Bandi C, Sironi M, Damiani G, *et al.* (1995) The Establishment of Intracellular Symbiosis in an Ancestor of Cockroaches and Termites. *Proceedings of the Royal Society B: Biological Sciences*, **259**, 293–299.
- Barat M, Tarayre M, Atlan A (2008) Genetic divergence and ecological specialisation of seed weevils (*Exapion spp.*) on gorses (*Ulex spp.*). *Ecological Entomology*, **33**, 328–336.
- Basit M, Sayyed aLIH, Saeed S (2012) Lack of Fitness Costs Associated With Acetamiprid Resistance in *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Journal of economic entomology*, **105**, 1401–1406.
- Basten CJ, Asmussen Ma (1997) The exact test for cytonuclear disequilibrium. *Genetics*, **146**, 1165–71.
- Baumann L, Baumann P (2005) Cospeciation between the primary endosymbionts of mealybugs and their hosts. *Current Microbiology*, **50**, 84–87.
- Baumann P (2005) Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annual review of microbiology*, **59**, 155–89.
- Baxter SW, Davey JW, Johnston JS, *et al.* (2011) Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS One*, **6**, e19315.
- Beck EA, Thompson AC, Sharbrough J, Brud E, Llopart A (2015) Gene flow between *Drosophila yakuba* and *Drosophila santomea* in subunit V of cytochrome c oxidase: A potential case of cytonuclear cointrogression. *Evolution*, **69**, 1973–1986.
- Bedford ID, Briddon RW, Brown JK, Rosell RC, Markham PG (1994) Geminivirus transmission and biological characterisation of *Bemisia tabaci* (Gennadius) biotypes from different geographic regions. *Annals of Applied Biology*, **125**, 311–325.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. *Laboratoire génome, populations, interactions, CNRS UMR*, **5000**, 1996–2004.
- Bellows TS, Perring TM, Gill RJ, Headrick DH (1994) Description of a species of *Bemisia* (Homoptera: Aleyrodidae). *Annals of the Entomological Society of America*, **87**, 195–206.
- Benazzo A, Ghirotto S, Vilaça ST, Hoban S (2015) Using ABC and microsatellite data to detect multiple introductions of invasive species from a single source. *Heredity*, **115**, 262–272.
- Bennett AE (2013) Can plant-microbe-insect interactions enhance or inhibit the spread of invasive species? *Functional Ecology*, **27**, 661–671.
- Bethke JA, Byrne FJ, Hodges GS, McKenzie CL, Shatters Jr RG (2009) First record of the Q biotype of the sweetpotato whitefly, *Bemisia tabaci*, in Guatemala. *Phytoparasitica*, **37**, 61–64.
- Bickford D, Lohman DJ, Sodhi NS, *et al.* (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, **22**, 148–155.
- Bing XL, Ruan YM, Rao Q, Wang XW, Liu SS (2013a) Diversity of secondary endosymbionts among different putative species of the whitefly *Bemisia tabaci*. *Insect science*, **20**, 194–206.
- Bing XL, Xia WQ, Gui JD, Yan GH, Wang XW, Liu SS (2014) Diversity and evolution of the *Wolbachia* endosymbionts of *Bemisia* (Hemiptera: Aleyrodidae) whiteflies. *Ecology and Evolution*, **4**, 2714–2737.
- Bing XL, Yang J, Zchori-Fein E, Wang XW, Liu SS (2013b) Characterization of a newly discovered symbiont of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Environmental Microbiology*

- 79, 569–575.
- Bird J, Brown JK (1998) Displacement of the Sida race of *Bemisia tabaci* by the B biotype of the *B. tabaci* complex in Puerto Rico. In *Proceedings of the Second International Whitefly and Geminivirus Workshop, San Juan, Puerto Rico*.
- Birky CW (1995) Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences*, **92**, 11331–11338.
- Blackman RL, Cahill M (1998) The karyotype of *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Bulletin of Entomological Research*, **88**, 213.
- Bossdorf O, Lipowsky A, Prati D (2008) Selection of preadapted populations allowed *Senecio inaequidens* to invade Central Europe. *Diversity and Distributions*, **14**, 676–685.
- Boykin LM, Armstrong KF, Kubatko L, de Barro P (2012) Species delimitation and global biosecurity. *Evolutionary Bioinformatics*, **8**, 1–37.
- Boykin LM, Bell CD, Evans G, Small I, De Barro PJ (2013) Is agriculture driving the diversification of the *Bemisia tabaci* species complex (Hemiptera: Sternorrhyncha: Aleyrodidae)? Dating, diversification and biogeographic evidence revealed. *BMC evolutionary biology*, **13**, 228.
- Boykin LM, Shatters RG, Rosell RC, et al. (2007) Global relationships of *Bemisia tabaci* (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Molecular phylogenetics and evolution*, **44**, 1306–19.
- Braendle C, Miura T, Bickel R, Shingleton AW, Kambhampati S, Stern DL (2003) Developmental origin and evolution of bacteriocytes in the aphid-*Buchnera* symbiosis. *PLoS Biology*, **1**.
- Brown J, Frohlich D, Rosell R (1995) The sweetpotato or silverleaf whiteflies: biotypes of *Bemisia tabaci* or a species complex? *Annual Review of Entomology*, **40**, 511–534.
- Brown JK, Idris aM (2005) Genetic Differentiation of Whitefly *Bemisia tabaci* Mitochondrial Cytochrome Oxidase I, and Phylogeographic Concordance with the Coat Protein of the Plant Virus Genus Begomovirus. *Annals of the Entomological Society of America*, **98**, 827–837.
- Brownlie JC, Cass BN, Riegler M, et al. (2009) Evidence for metabolic provisioning by a common invertebrate endosymbiont, *Wolbachia pipentis*, during periods of nutritional stress. *PLoS Pathogens*, **5**.
- Brumin M, Kontsedalov S, Ghanim M (2011) *Rickettsia* influences thermotolerance in the whitefly *Bemisia tabaci* B biotype. *Insect Science*, **18**, 57–66.
- Brumin M, Levy M, Ghanim M (2012) Transovarial transmission of *Rickettsia* spp. and organ-specific infection of the whitefly *Bemisia tabaci*. *Environmental Microbiology*, **78**, 5565–74.
- Buchner P (1953) *Endosymbiose der Tiere mit Pflanzlichen Mikroorganismen*. Birkhauser, Verlag.
- Burban C, Fishpool L, Fauquet C, Fargette D, Thouvenel J (1992) Host-associated biotypes within West African populations of the whitefly *Bemisia tabaci* (Genn.) (Hom., Aleyrodidae). *Journal of Applied Entomology*, **113**, 416–423.
- Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PDN (2008) DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservacion Guanacaste, Costa Rica. *Proceedings of the National Academy of Sciences*, **105**, 6350–5.
- Byrne F, Gorman K (2000) The role of B-type esterases in conferring insecticide resistance in the tobacco whitefly, *Bemisia tabaci* (Genn.). *Pest management science*, **874**, 867–874.
- Byrne FJ, Cahill M, Denholm I, Devonshire AL (1995) Biochemical identification of interbreeding between B-type and non B-type strains of the tobacco whitefly *Bemisia tabaci*. *Biochemical genetics*, **33**, 13–23.
- Cardoza YJ, Klepzig KD, Raffa KF (2006) Bacteria in oral secretions of an endophytic insect inhibit antagonistic fungi. *Ecological Entomology*, **31**, 636–645.
- Caspi-Fluger A, Inbar M, Mozes-Daube N, Mouton L, Hunter MS, Zchori-Fein E (2011) *Rickettsia* 'in' and 'out': two different localization patterns of a bacterial symbiont in the same insect species. *PLoS One*, **6**, e21096.
- Caspi-Fluger A, Inbar M, Mozes-Daube N, et al. (2012) Horizontal transmission of the insect symbiont *Rickettsia* is plant-mediated. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 1791–6.
- Caspi-Fluger A, Inbar M, Steinberg S, et al. (2014) Characterization of the symbiont *Rickettsia* in the mirid bug *Nesidiocoris tenuis* (Reuter) (Heteroptera: Miridae). *Bulletin of Entomological Research*, **104**, 681–688.
- Cass BN, Himler AG, Bondy EC, et al. (2015a) Conditional fitness benefits of the *Rickettsia* bacterial symbiont in an insect pest. *Oecologia*.
- Cass BN, Yallouz R, Bondy EC, et al. (2015b) Dynamics of the Endosymbiont *Rickettsia* in an Insect Pest. *Microbial Ecology*.
- Catchen J, Hohenlohe Pa, Bassham S, Amores A, Cresko Wa (2013) Stacks: an analysis tool set for population genomics. *Molecular ecology*, **22**, 3124–40.
- Cervera MT, Cabezas Ja, Simón B, Martínez-Zapater JM, Beitia F, Cenis JL (2000) Genetic relationships among biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae) based on AFLP analysis. *Bulletin of entomological research*, **90**, 391–396.
- Chang KP, Musgrave aJ (1972) Multiple symbiosis in a leafhopper, *Heliochara communis* fitch (Cicadellidae: Homoptera): envelopes, nucleoids and inclusions of the symbiotes. *Journal of Cell Science*, **11**, 275–293.
- Chattopadhyay B, Garg KM, K VKA, D PSD, Ramakrishnan U, Kandula S (2012) Sibling species in South Indian populations of the rufous horse-shoe bat *Rhinolophus rouxii*. *Conservation Genetics*, **13**, 1435–1445.
- Chaubey R, Andrew RJ, Naveen NC, Rajagopal R, Ahmad B, Ramamurthy VV (2015) Morphometric Analysis of Three Putative Species of *Bemisia tabaci* (Hemiptera: Aleyrodidae) Species Complex from India. *Annals of the Entomological Society of America*, **108**, 600–612.
- Chen DQ, Campbell BC, Purcell aH (1996) A new *Rickettsia* from a herbivorous insect, the pea aphid *Acyrtosiphon pisum* (Harris). *Current Microbiology*, **33**, 123–128.
- 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*. *Journal of Molecular Evolution*, **48**, 49–58.
- Chermi B, Braham M, Cenis J, Alonso C, Beitia F (1997) Sur la présence en Tunisie des biotypes 'B' et 'non B' de *Bemisia tabaci* (Homoptera: Aleyrodidae) et de leurs parasitoïdes associés. *Bulletin OILB/SROP*, **20**, 108–113.
- Chiel E, Gottlieb Y, Zchori-Fein E, et al. (2007) Biotype-dependent secondary symbiont communities in sympatric populations of *Bemisia tabaci*. *Bulletin of entomological research*, **97**, 407–13.
- Chiel E, Zchori-Fein E, Inbar M, et al. (2009) Almost there: transmission routes of bacterial symbionts between trophic levels. *PLoS One*, **4**, e4767.
- Chowda-Reddy R, Kirankumar M (2012) *Bemisia tabaci* Phylogenetic Groups in India and the Relative Transmission Efficacy of *Tomato leaf curl Bangalore virus* by an Indigenous and an Exotic Population. *Journal of Integrative Agriculture*, **11**, 235–248.
- Chrudimský T, Husník F, Nováková E, Hypša V (2012) *Candidatus Sodalis melophagi* sp. nov.: Phylogenetically Independent Comparative Model to the Tsetse Fly Symbiont *Sodalis glossinidia*. *PLoS One*, **7**, e40354.
- Chu D, Jiang T, Liu GX, et al. (2007) Biotype status and distribution of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Shandong province of China based on mitochondrial DNA markers. *Environmental Entomology*, **36**, 1290–5.
- Chu D, Wan FH, Tao YL, Liu GX, Fan ZX, Bi YP (2008) Genetic differentiation of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) biotype Q based on mitochondrial DNA markers. *Insect Science*, **15**, 115–123.
- Chu D, Zhang YJ, Brown JK, et al. (2006) The introduction of the exotic Q biotype of *Bemisia tabaci* from the mediterranean region into china on ornamental crops. *Florida Entomologist*, **89**, 168–

- 174.
- Claridge AW, Mifsud G, Dawson J, Saxon MJ (2004) Use of infrared digital cameras to investigate the behaviour of cryptic species. *Wildlife Research*, **31**, 645–650.
- Cockerham CC, Weir BS (1977) Digenic descent measures for finite populations. *Genetical Research*, **30**, 121–147.
- Colautti RI, Macisaac HJ (2004) A neutral terminology to define ‘invasive’ species. *Diversity and Distributions*, **10**, 135–141.
- Colborn J, Rabtree R, Haklee JB, Feiler E, Bowen BW (2001) The Evolutionary Enigma of Bonefishes (*Albula spp.*): Cryptic Species and Ancient Separations in a Globally Distributed Shorefish. *Evolution*, **55**, 807–820.
- Colston SM, Fullmer MS, Beka L, Lamy B, Gogarten JP (2014) Bioinformatic Genome Comparisons for Taxonomic and Phylogenetic Assignments Using *Aeromonas* as a Test Case. *mBio*, **5**, 1–13.
- Costa HS, Brown JK (1991) Variation in biological characteristics and esterase patterns among populations of *Bemisia tabaci*, and the association of one population with silverleaf symptom induction. *Entomologia Experimentalis et Applicata*, **61**, 211–219.
- Costa HS, Brown JK, Sivasupramaniam S, Bird J (1993) Regional distribution, insecticide resistance, and reciprocal crosses between the A-biotype and B-biotype of *Bemisia tabaci*. *Insect Science and its Application*, **14**, 255–266.
- Costa HS, Toscano NC, Henneberry TJ (1996) Mycocyte inclusion in the oocytes of *Bemisia argentifolii* (Homoptera: Aleyrodidae). *Annals of the Entomological Society of America*, **89**, 694–699.
- Costa HS, Westcot DM, Ullman DE, Rosell R, Brown JK, Johnson MW (1995) Morphological variation in *Bemisia* endosymbionts. *Protoplasma*, **189**, 194–202.
- Crowder DW, Eilers-Kirk C, Tabashnik BE, Carrière Y (2009) Lack of fitness costs associated with pyriproxyfen resistance in the B biotype of *Bemisia tabaci*. *Pest Management Science*, **65**, 235–240.
- Currie CR, Scott JA, Summerbell RC, Malloch D (1999) Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. *Nature*, **398**, 701–704.
- Dale C, Moran Na (2006) Molecular Interactions between Bacterial Symbionts and Their Hosts. *Cell*, **126**, 453–465.
- Dalmon a, Halkett F, Granier M, Delatte H, Peterschmitt M (2008) Genetic structure of the invasive pest *Bemisia tabaci*: evidence of limited but persistent genetic differentiation in glasshouse populations. *Heredity*, **100**, 316–25.
- Damiani C, Ricci I, Crotti E, et al. (2008) Paternal transmission of symbiotic bacteria in malaria vectors. *Current Biology*, **18**, R1087–8.
- Darby AC, Birkle LM, Turner SL, Douglas AE (2001) An aphid-borne bacterium allied to the secondary symbionts of whitefly. *FEMS Microbiology Ecology*, **36**, 43–50.
- Darby AC, Chandler SM, Welburn SC, Douglas AE (2005) Aphid-symbiotic bacteria cultured in insect cell lines. *Applied and Environmental Microbiology*, **71**, 4833–4839.
- Dasmahapatra KK, Walters JR, Briscoe AD, et al. (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, **487**, 94–98.
- Davey JW, Davey JL, Blaxter ML, Blaxter MW (2011) RADSeq: next-generation population genetics. *Briefings in functional genomics*, **9**, 416–23.
- Davis MA (2009) *Invasion biology*. Oxford University Press on Demand.
- Davis MA, Chew MK, Hobbs RJ, et al. (2011) Don’t judge species on their origins. *Nature*, **474**, 153–154.
- Davis MJ, Ying Z, Brunner BR, Pantoja A, Ferwerda FH (1998) *Rickettsial* relative associated with papaya bunchy top disease. *Current Microbiology*, **36**, 80–84.
- Davis TS (2015) The Ecology of Yeasts in the Bark Beetle Holobiont: A Century of Research Revisited. *Microbial Ecology*, **69**, 723–732.
- De Barro P, Trueman J, Frohlich D (2005) *Bemisia argentifolii* is a race of *B. tabaci* (Hemiptera: Aleyrodidae): the molecular genetic differentiation of *B. tabaci* populations around the world. *Bulletin of Entomological Research*, **95**, 193.
- De Barro PJ, Hart PJ (2000) Mating interactions between two biotypes of the whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Australia. *Bulletin of entomological research*, **90**, 103–112.
- De Barro PJ, Liu SS, Boykin LM, Dinsdale AB (2011) *Bemisia tabaci*: a statement of species status. *Annual Review of Entomology*, **56**, 1–19.
- De Bary A (1879) The phenomenon of symbiosis. *Karl J. Trubner, Strasbourg, Germany*.
- De Carvalho D, Ingvarsson PK, Joseph J, et al. (2010) Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Molecular ecology*, **19**, 1638–50.
- De Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology*, **56**, 879–886.
- De Souza DJ, Bézier A, Depoix D, Drezen JM, Lenoir A (2009) Blochmannia endosymbionts improve colony growth and immune defence in the ant *Camponotus fellah*. *BMC Microbiology*, **9**, 29.
- De Vooght L, Caljon G, Van Hees J, Van Den Abbeele J (2015) Paternal Transmission of a Secondary Symbiont during Mating in the Viviparous Tsetse Fly. *Molecular Biology and Evolution*, **32**, 1977–1980.
- Degnan PH, Lazarus AB, Brock CD, Wernegreen JJ (2004) Host-symbiont stability and fast evolutionary rates in an ant-bacterium association: cospeciation of *Camponotus* species and their endosymbionts, *Candidatus Blochmannia*. *Systematic biology*, **53**, 95–110.
- Degnan PH, Moran Na (2008) Evolutionary genetics of a defensive facultative symbiont of insects: exchange of toxin-encoding bacteriophage. *Molecular ecology*, **17**, 916–29.
- Delatte H, Baudin R, Becker N, et al. (2014) Species and endosymbiont diversity of *Bemisia tabaci* (Homoptera: Aleyrodidae) on vegetable crops in Senegal. *Insect Science*, **S. n/a–n/a**.
- Delatte H, David P, Granier M, et al. (2006) Microsatellites reveal extensive geographical, ecological and genetic contacts between invasive and indigenous whitefly biotypes in an insular environment. *Genetical research*, **87**, 109–24.
- Delatte H, Reynaud B, Granier M, et al. (2005) A new silverleaf-inducing biotype Ms of *Bemisia tabaci* (Hemiptera: Aleyrodidae) indigenous to the islands of the south-west Indian Ocean. *Bulletin of entomological research*, **95**, 29–35.
- Demichelis S, Bosco D, Manino A, Marian D, Caciagli P (2000) Distribution of *Bemisia tabaci* (Hemiptera: Aleyrodidae) biotypes in Italy. *The Canadian Entomologist*, **132**, 519–527.
- Dennehy TJ, Degain Ba, Harpold VS, et al. (2010) Extraordinary resistance to insecticides reveals exotic Q biotype of *Bemisia tabaci* in the New World. *Journal of economic entomology*, **103**, 2174–2186.
- Dickey AM, Hall PM, Shatters RG, Mckenzie CL (2013) Evolution and homoplasy at the Bem6 microsatellite locus in three sweetpotato whitefly (*Bemisia tabaci*) cryptic species. *BMC research notes*, **6**, 249.
- Dinsdale A, Cook L, Riginos C, Buckley YM, De Barro P (2010) Refined Global Analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) Mitochondrial Cytochrome Oxidase 1 to Identify Species Level Genetic Boundaries. *Annals of the Entomological Society of America*, **103**, 196–208.
- Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular ecology*, **17**, 431–49.
- Dobzhansky T (1968) *Evolutionary Biology: Volume 2*, Kap. On Some Fu, S. 1–34. Springer US, Boston, MA.
- Douglas AE (2009) The microbial dimension in insect nutritional ecology. *Functional Ecology*, **23**, 38–47.
- Dunbar HE, Wilson ACC, Ferguson NR, Moran NA (2007) Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS Biology*, **5**, 1006–1015.
- Duron O, Bouchon D, Boutin S, et al. (2008) The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC biology*, **6**, 27.
- Duron O, Lagnel J, Raymond M, Bourtzis K, Fort P, Weill M (2005) Transposable element polymorphism of *Wolbachia* in the



- mosquito *Culex pipiens*: Evidence of genetic diversity, superinfection and recombination. *Molecular Ecology*, **14**, 1561–1573.
- Duron O, Wilkes TE, Hurst GDD (2010) Interspecific transmission of a male-killing bacterium on an ecological timescale. *Ecology Letters*, **13**, 1139–1148.
- Earl Da, VonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.
- Eaton DaR (2014) PyRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1–6.
- Eaton DaR, Ree RH (2013) Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (Pedicularis: Orobanchaceae). *Systematic Biology*, **62**, 689–706.
- Eberl G (2015) Nos bactéries et nous: un subtil équilibre. *Pour la science*, **5**, 27–33.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, **32**, 1792–7.
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–1.
- Elbaz M, Lahav N, Morin S (2010) Evidence for pre-zygotic reproductive barrier between the B and Q biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Bulletin of entomological research*, **100**, 581–590.
- Elbert A, Nauen R (2000) Resistance of *Bemisia tabaci* (Homoptera: Aleyrodidae) to insecticides in southern Spain with special reference to neonicotinoids. *Pest Management Science*, **64**, 60–64.
- Ellstrand NC, Schierenbeck Ka (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences*, **97**, 7043–50.
- Elton CS (1958) *The ecology of invasions by plants and animals*, Bd. 18. Methuen, London.
- Erdogan C, Moores GD, Oktay Gurkan M, Gorman KJ, Denholm I (2008) Insecticide resistance and biotype status of populations of the tobacco whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) from Turkey. *Crop Protection*, **27**, 600–605.
- Etter P, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA (2011) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Methods in Molecular Biology*, **772**, 1–19.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology*, **14**, 2611–20.
- Everett KDE, Thao M, Horn M, Dyszynski GE, Baumann P (2005) Novel chlamydiae in whiteflies and scale insects: Endosymbionts 'Candidatus Fritschea bemisiae' strain Falk and 'Candidatus Fritschea eriococci' strain Elm. *International Journal of Systematic and Evolutionary Microbiology*, **55**, 1581–1587.
- Facon B, Pointier JP, Jarne P, Sarda V, David P (2008) High genetic variance in life-history strategies within invasive populations by way of multiple introductions. *Current Biology*, **18**, 363–7.
- Ferrari J, Darby AC, Daniell TJ, Godfray HCJ, Douglas AE (2004) Linking the bacterial community in pea aphids with host-plant use and natural enemy resistance. *Ecological Entomology*, **29**, 60–65.
- Ferrari J, Vavre F (2011) Bacterial symbionts in insects or the story of communities affecting communities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 1389–400.
- Fiala-Medioni A (1988) Synthèse sur les adaptations structurales liées à la nutrition des mollusques bivalves des sources hydrothermales profondes. *Oceanologica Acta*, **8**, 173–179.
- Firdaus S, Vosman B, Hidayati N, Jaya Supena ED, GF Visser R, van Heusden AW (2013) The *Bemisia tabaci* species complex: Additions from different parts of the world. *Insect Science*, **20**, 723–733.
- Fitzpatrick BM (2014) Symbiote transmission and maintenance of extra-genomic associations. *Frontiers in Microbiology*, **5**, 46.
- Foll M, Gaggiotti OE, Daub JT, Vatsiou A, Excoffier L (2014) Widespread signals of convergent adaptation to high altitude in Asia and America. *American Journal of Human Genetics*, **95**, 394–407.
- Fontes FVM, Colombo Ca, Lourencao aL (2012) Structure of genetic diversity of *Bemisia tabaci* (Genn.) (Hemiptera: Aleyrodidae) populations in Brazilian crops and locations. *Scientia Agricola*, **69**, 47–53.
- Frank AB (1876) *Ueber die biologischen Verhältnisse des Thalus einiger Krustenflechten: Vorläufige Mittheilung*. K[ö]nigl. Botanische Gesellschaft in Regensburg.
- Frantz A, Calcagno V, Mieuze L, Plantegenest M, Simon JCJC (2009) Complex trait differentiation between host-populations of the pea aphid *Acyrtosiphon pisum* (Harris): implications for the evolution of ecological specialisation. *Biological Journal of the Linnean Society*, **97**, 718–727.
- Fuchs J, Ericson PGP, Bonillo C, Couloux A, Pasquet E (2015) The complex phylogeography of the Indo-Malayan *Alophoxus* bulbuls with the description of a putative new ring species complex. *Molecular Ecology*, **24**, 5460–5474.
- Fujiie A, Sawas AE, Hadi MA, Sawas AB, Ueda S, Natsuaki KT (2007) First record on the occurrence of q biotype *Bemisia tabaci* on potatoes in Syria. *Journal of ISSAAS*, **13**, 1–8.
- Fujiwara a, Maekawa K, Tsuchida T (2015) Genetic groups and endosymbiotic microbiota of the *Bemisia tabaci* species complex in Japanese agricultural sites. *Journal of Applied Entomology*, **139**, 55–66.
- Fytrou A, Schofield PG, Kraaijeveld AR, Hubbard SF (2006) *Wolbachia* infection suppresses both host defence and parasitoid counter-defence. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 791–796.
- Galil BS (2007) Loss or gain? Invasive aliens and biodiversity in the Mediterranean Sea. *Marine Pollution Bulletin*, **55**, 314–322.
- Galtier N, Nabholz B, Glémin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular ecology*, **18**, 4541–50.
- Garnier S, Alibert P, Audiot P, Prieur B, Rasplus JY (2004) Isolation by distance and sharp discontinuities in gene frequencies: implications for the phylogeography of an alpine insect species, *Carabus solieri*. *Molecular ecology*, **13**, 1883–97.
- Gascuel O (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution*, **14**, 685–695.
- Gauthier N, Clouet C, Perrakis A, Kapantaidaki D, Peterschmitt M, Tsagkarakou A (2014) Genetic structure of *Bemisia tabaci* Med populations from home-range countries, inferred by nuclear and cytoplasmic markers: impact on the distribution of the insecticide resistance genes. *Pest management science*, **70**, 1477–1491.
- Gauthier N, Dalleau-Clouet C, Bouvret ME (2008) Twelve new polymorphic microsatellite loci and PCR multiplexing in the whitefly, *Bemisia tabaci*. *Molecular ecology resources*, **8**, 1004–7.
- Gautier M, Foucaud J, Gharbi K, et al. (2013) Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Molecular ecology*, **22**, 3766–79.
- Gehrer L, Vorburger C (2012) Parasitoids as vectors of facultative bacterial endosymbionts in aphids. *Biology Letters*, **8**, 613–615.
- Gennadius P (1889) Disease of tobacco plantations in the Trikonika. The aleurodid of tobacco. *Ellenike Georgia*, **5**, 1–3.
- Genton BJ, Shykoff Ja, Giraud T (2005) High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of introduction. *Molecular ecology*, **14**, 4275–85.
- Ghanim M, Kotsedalov S (2009) Susceptibility to insecticides in the Q biotype of *Bemisia tabaci* is correlated with bacterial symbiont densities. *Pest management science*, **65**, 939–42.
- Gherna RL, Werren JH, Weisburg W, et al. (1991) *Arsenophonus nasoniae* gen. nov, sp. nov, the causative agent of the son-killer trait in the parasitic wasp *Nasonia vitripennis*. *International Journal of Systematic Bacteriology*, **41**, 563–565.
- Ghosh S, Bouvaine S, Maruthi M (2015) Prevalence and genetic diversity of endosymbiotic bacteria infecting cassava whiteflies in Africa. *BMC Microbiology*, **15**, 1–17.
- Gillis NK, Walters LJ, Fernandes FC, Hoffman Ea (2009) Higher genetic diversity in introduced than in native populations of the muscel *Mytella charruana*: evidence of population admixture at introduction sites. *Diversity and Distributions*, **15**, 784–795.

- Giron D, Frago E, Glevarec G, Pieterse CMJ, Dicke M (2013) Cytokinin as key regulators in plant-microbe-insect interactions: Connecting plant growth and defence. *Functional Ecology*, **27**, 599–609.
- Giska I, Sechi P, Babik W (2015) Deeply divergent sympatric mitochondrial lineages of the earthworm *Lumbricus rubellus* are not reproductively isolated. *BMC Evolutionary Biology*, **15**, 217.
- Giustina W, Martinez M, Bertaux F (1989) *Bemisia tabaci*: the new enemy of glasshouse crops in Europe. *Phytoma*, **406**, 48–52.
- Gleditsch JM, Carlo TA (2011) Fruit quantity of invasive shrubs predicts the abundance of common native avian frugivores in central Pennsylvania. *Diversity and Distributions*, **17**, 244–253.
- Gnankiné O, Mouton L, Henri H, et al. (2013a) Distribution of *Bemisia tabaci* (Homoptera: Aleyrodidae) biotypes and their associated symbiotic bacteria on host plants in West Africa. *Insect Conservation and Diversity*, **6**, 411–421.
- Gnankiné O, Mouton L, Savadogo A, et al. (2013b) Biotype status and resistance to neonicotinoids and carbosulfan in *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Burkina Faso, West Africa. *International journal of pest management*, **59**, 95–102.
- Gnankiné O, Traoré D, Sanon A, Ouedraogo AP (2007) Traitements insecticides et dynamique des populations de *Bemisia tabaci* Gennadius en culture cotonnière au Burkina Faso. *Cahiers Agricoles*, **16**, 101–109.
- Gomez-Menor J (1943) *Contribucion al conocimiento de los Aleyrodidos de Espana (Hem. Homop.)*, S. 173–209. Sinauer Associates Incorporated.
- Gonella E, Crotti E, Rizzi A, et al. (2012) Horizontal transmission of the symbiotic bacterium *Asaia* sp. in the leafhopper *Scaphoideus titanus* Ball (Hemiptera: Cicadellidae). *BMC microbiology*, **12** Suppl 1, S4.
- Goodsell JA, Kats LB (1999) Effect of introduced mosquitofish on pacific treefrogs and the role of alternative prey. *Conservation Biology*, **13**, 921–924.
- Gorsane F, Ben Halima A, Ben Khalifa M, Bel-Kadhi MS, Fakhfakh H (2011) Molecular characterization of *Bemisia tabaci* populations in Tunisia: genetic structure and evidence for multiple acquisition of secondary symbionts. *Environmental Entomology*, **40**, 809–17.
- Goto S, Anbutsu H, Fukatsu T (2006) Asymmetrical interactions between *Wolbachia* and *Spiroplasma* endosymbionts coexisting in the same insect host. *Environmental Microbiology*, **72**, 4805–10.
- Gotoh T, Noda H, Ito S (2007) *Cardinium* symbionts cause cytoplasmic incompatibility in spider mites. *Heredity*, **98**, 13–20.
- Gottlieb Y, Ghanim M, Chiel E, et al. (2006) Identification and Localization of a *Rickettsia* sp. in *Bemisia tabaci* (Homoptera: Aleyrodidae). *Environmental Microbiology*, **72**, 3646–3652.
- Gottlieb Y, Ghanim M, Gueguen G, et al. (2008) Inherited intracellular ecosystem: symbiotic bacteria share bacteriocytes in whiteflies. *The FASEB journal*, **22**, 2591–9.
- Gottlieb Y, Zchori-Fein E, Mozes-Daube N, et al. (2010) The transmission efficiency of tomato yellow leaf curl virus by the whitefly *Bemisia tabaci* is correlated with the presence of a specific symbiotic bacterium species. *Journal of virology*, **84**, 9310–7.
- Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular ecology*, **5**, 184–186.
- Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution*, **27**, 221–4.
- Grille G, Gauthier N, Buenahora J, Basso C, Bonato O (2011) First report of the Q biotype of *Bemisia tabaci* in Argentina and Uruguay. *Phytoparasitica*, **39**, 235–238.
- Gueguen G (2009) *Les communautés endosymbiotiques des insectes vecteurs de virus : diversité bactérienne, effets phénotypiques, conséquences écologiques et épidémiologiques*. Dissertation, Université Claude Bernard Lyon 1.
- Gueguen G, Rolain JM, Zchori-Fein E, Vavre F, Fleury F, Raoult D (2009) Molecular detection and identification of *Rickettsia* endosymbiont in different biotypes of *Bemisia tabaci*. *Clinical microbiology and infection*, **15** Suppl 2, 271–2.
- Gueguen G, Vavre F, Gnankiné O, et al. (2010) Endosymbiont meta-communities, mtDNA diversity and the evolution of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. *Molecular ecology*, S. 4365–4378.
- Guevara-Coto JA, Barboza-Vargas N, Hernandez-Jimenez E, Hammond RW, Ramirez-Fonseca P (2011) *Bemisia tabaci* Biotype Q is present in Costa Rica. *European Journal of Plant Pathology*, **131**, 167–170.
- Guirao P, Beitia F, Cenis JL (1997) Biotype determination of spanish populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Bulletin of entomological research*, **87**, 587–593.
- Gunning RV, Byrne FJ, Devonshire AL (1997) Electrophoretic analysis of non-B and B-biotype *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) in Australia. *Australian Journal of Entomology*, **36**, 245–249.
- Guo LT, Wang SL, Wu QJ, Zhou XG, Xie W, Zhang YJ (2015) Flow cytometry and K-mer analysis estimates of the genome sizes of *Bemisia tabaci* B and Q (Hemiptera: Aleyrodidae). *Frontiers in Physiology*, **6**, 1–7.
- Guo Xj, Rao Q, Zhang F, Luo C, Zhang Hy, Gao Xw (2012) Diversity and Genetic Differentiation of the Whitefly *Bemisia tabaci* Species Complex in China Based on mtCOI and cDNA-AFLP Analysis. *Journal of Integrative Agriculture*, **11**, 206–214.
- Gurevitch J, Padilla DK (2004) Are invasive species a major cause of extinctions? *Trends in Ecology & Evolution*, **19**, 470–474.
- Haack Ra, Hérard F, Sun J, Turgeon JJ (2010) Managing invasive populations of Asian longhorned beetle and citrus longhorned beetle: a worldwide perspective. *Annual Review of Entomology*, **55**, 521–546.
- Hagimori T, Abe Y, Date S, Miura K (2006) The first finding of a *Rickettsia* bacterium associated with parthenogenesis induction among insects. *Current microbiology*, **52**, 97–101.
- 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. *Applied and Environmental Microbiology*, **73**, 7531–7535.
- Haygood MG, Distel DL (1993) Bioluminescent Symbionts of Flashlight Fishes and Deep-Sea Anglerfishes Form Unique Lineages Related to the Genus *Vibrio*.
- Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences*, **101**, 14812–14817.
- Hedges LM, Brownlie JC, O'Neill SL, Johnson KN (2008) *Wolbachia* and virus protection in insects. *Science*, **322**, 702.
- Hekkala E, Shirley MH, Amato G, et al. (2011) An ancient icon reveals new mysteries: Mummy DNA resurrects a cryptic species within the Nile crocodile. *Molecular Ecology*, **20**, 4199–4215.
- Henri H, Terraz G, Gnankiné O, Fleury F, Mouton L (2013) Molecular characterization of genetic diversity within the Africa/Middle East/Asia Minor and Sub-Saharan African groups of the *Bemisia tabaci* species complex. *International journal of pest management*, **59**, 329–338.
- Herrera S, Reyes-Herrera PH, Shank TM (2015) Predicting genome sizes and restriction enzyme recognition-sequence probabilities across the eukaryotic tree of life. *bioRxiv*, **7**, 007781.
- Heuvel JFJM, Verbeek M, Van Der Wilk F (1994) Endosymbiotic bacteria associated with circulative transmission of potato leafroll virus transmission by *Myzus persicae*. *Journal of General Virology*, **75**, 2559–2565.
- Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH (2008) How many species are infected with *Wolbachia*? - A statistical analysis of current data. *FEMS Microbiology Letters*, **281**, 215–220.
- Himler AG, Adachi-Hagimori T, Bergen JE, et al. (2011) Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. *Science*, **332**, 254–6.
- Hohenlohe Pa, Bassham S, Etter PD, Stiffler N, Johnson Ea, Cresko Wa (2010) Population genomics of parallel adaptation in three-spine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.
- Holt RD, Gomulkiewicz R, Barfield M (2003) The phenomenology of

- niche evolution via quantitative traits in a 'black-hole' sink. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 215–24.
- Horowitz AR, Denholm I, Gorman KJ, Cenis JL, Kontsedalov S, Ishaaya I (2003) Biotyping Q of *Bemisia tabaci* identified in Israel. *Phytoparasitica*, **31**, 94–98.
- Horowitz AR, Kontsedalov S, Denholm I, Ishaaya I (2002) Dynamics of insecticide resistance in *Bemisia tabaci*: a case study with the insect growth regulator pyriproxyfen. *Pest management science*, **58**, 1096–1100.
- Horowitz AR, Kontsedalov S, Khasdan V, Ishaaya I (2005) Biotypes B and Q of *Bemisia tabaci* and their relevance to neonicotinoid and pyriproxyfen resistance. *Archives of insect biochemistry and physiology*, **58**, 216–25.
- Hothorn T, Bretz F, Westfall P (2008) Simultaneous Inference in General Parametric Models. *Biometrical Journal*, **50**, 346–363.
- Houndété TA, Kétoh GK, Hema OSA, Brévault T, Glietho IA, Martin T (2010) Insecticide resistance in field populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae) in West Africa. *Pest Management Science*, **66**, 1181–1185.
- Hoy Ma, Jeyapragash A (2005) Microbial diversity in the predatory mite *Metaseiulus occidentalis* (Acari: Phytoseiidae) and its prey, *Tetranychus urticae* (Acari: Tetranychidae). *Biological Control*, **32**, 427–441.
- Hsieh CH, Ko CC, Chung CH, Wang HY (2014) Multilocus approach to clarify species status and the divergence history of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. *Molecular Phylogenetics and Evolution*, **76**, 172–180.
- Hsieh CH, Wang CH, Ko CC (2006) Analysis of *Bemisia tabaci* (Hemiptera: Aleyrodidae) Species Complex and Distribution in Eastern Asia Based on Mitochondrial DNA Markers. *Annals of the Entomological Society of America*, **99**, 768–775.
- Hsieh CH, Wang CH, Ko CC (2007) Evidence from molecular markers and population genetic analyses suggests recent invasions of the Western North Pacific region by biotypes B and Q of *Bemisia tabaci* (Gennadius). *Environmental Entomology*, **36**, 952–61.
- Hu HY, Li ZX (2015) A novel *Wolbachia* strain from the rice moth *Corcyra cephalonica* induces reproductive incompatibility in the whitefly *Bemisia tabaci*: sequence typing combined with phenotypic evidence. *Environmental Microbiology*, **7**, 508–515.
- Hu J, De Barro P, Zhao H, Wang J, Nardi F, Liu SS (2011) An extensive field survey combined with a phylogenetic analysis reveals rapid and widespread invasion of two alien whiteflies in China. *PLoS One*, **6**, e16061.
- Huerta-Sánchez E, Jin X, Asan, *et al.* (2014) Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, **512**, 194–7.
- Hufbauer Ra, Facon B, Ravigné V, *et al.* (2012) Anthropogenically induced adaptation to invade (AIAI): Contemporary adaptation to human-altered habitats within the native range can promote invasions. *Evolutionary Applications*, **5**, 89–101.
- Hunter MS, Perlman SJ, Kelly SE (2003) A bacterial symbiont in the Bacteroidetes induces cytoplasmic incompatibility in the parasitoid wasp *Encarsia pergandiella*. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 2185–2190.
- Hurst GDD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 1525–34.
- Jaenike J, Unckless R, Cockburn SN, Boelio LM, Perlman SJ (2010) Adaptation via symbiosis: recent spread of a *Drosophila* defensive symbiont. *Science*, **329**, 212–5.
- Jakobsson M, Rosenberg NA (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.
- Janson EM, Stireman JO, Singer MS, Abbot P (2008) Phytophagous Insect–Microbe Mutualisms and Adaptive Evolutionary Diversification. *Evolution*, **62**, 997–1012.
- Jayasekera S, Thomas A, Kar A, Ramamurthy VV (2010) Host correlated morphometric variations in the populations of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae). *Oriental Insects*, **44**, 193–204.
- Jeyapragash A, Hoy MA (2000) Long PCR improves *Wolbachia* DNA amplification: wsp sequences found in 76% of sixty-three arthropod species. *Insect Molecular Biology*, **9**, 393–405.
- Jiang YX (1999) Probing and Feeding Behavior of Two Distinct Biotypes of *Bemisia tabaci* (Homoptera: Aleyrodidae) on Tomato Plants. *Ecology and Behaviour*, **92**, 357–366.
- Jing X, Wong ACN, Chaston JM, Colvin J, McKenzie CL, Douglas AE (2014) The bacterial communities in plant phloem-sap-feeding insects. *Molecular Ecology*, **23**, 1433–1444.
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–5.
- Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**, 92–103.
- Jombart T, Pontier D, Dufour AB (2009) Genetic markers in the playground of multivariate analysis. *Heredity*, **102**, 330–41.
- Kaiser W, Huguet E, Casas J, Commin C, Giron D (2010) Plant green-island phenotype induced by leaf-miners is mediated by bacterial symbionts. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 2311–2319.
- Kaltenpoth M (2009) Actinobacteria as mutualists: general healthcare for insects? *Trends in Microbiology*, **17**, 529–535.
- Kaltenpoth M, Göttler W, Herzner G, Strohm E (2005) Symbiotic bacteria protect wasp larvae from fungal infestation. *Current Biology*, **15**, 475–479.
- Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, **2**, e281.
- Kaplan I, Denno RF (2007) Interspecific interactions in phytophagous insects revisited: a quantitative assessment of competition theory. *Ecology Letters*, **10**, 977–994.
- Karunker I, Benting J, Lueke B, *et al.* (2008) Over-expression of cytochrome P450 CYP6CM1 is associated with high resistance to imidacloprid in the B and Q biotypes of *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Insect biochemistry and molecular biology*, **38**, 634–44.
- Karut K, Kaydan MB, Castle SJ, Kazak C, Ulusoy MR (2014) Study on species composition of *Bemisia tabaci* (Gennadius, 1889) (Hemiptera: Aleyrodidae) on cotton in Cukurova plain, Turkey. *Turkiye Entomoloji Dergisi*, **38**, 43–50.
- Karut K, Tok B (2014) Secondary endosymbionts of Turkish *Bemisia tabaci* (Gennadius) populations. *Phytoparasitica*, **42**, 413–419.
- Keller S, Gilbert K, Fields P, Taylor D (2012) Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, *Silene latifolia*. *Molecular ecology*, **21**, 4721–4734.
- Keller SR, Taylor DR (2010) Genomic admixture increases fitness during a biological invasion. *Journal of evolutionary biology*, **23**, 1720–31.
- Khasdan V, Levin I, Rosner A, *et al.* (2005) DNA markers for identifying biotypes B and Q of *Bemisia tabaci* (Hemiptera: Aleyrodidae) and studying population dynamics. *Bulletin of Entomological Research*, **95**, 605–613.
- Klasson L, Walker T, Sebahia M, *et al.* (2008) Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group. *Molecular Biology and Evolution*, **25**, 1877–1887.
- Kliot A, Cilia M, Czosnek H, Ghanim M (2014) Implication of the bacterial endosymbiont *Rickettsia* spp. in interactions of the whitefly *Bemisia tabaci* with tomato yellow leaf curl virus. *Journal of virology*, **88**, 5652–60.
- Knowlton N (1986) Cryptic and Sibling Species among the Decapod Crustacea. *Journal of Crustacean Biology*, **6**, 356–363.
- Koga R, Meng XY, Tsuchida T, Fukatsu T (2012) Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, E1230–7.
- Kolar CS, Lodge DM (2001) Progress in invasion biology: predicting invaders. *Trends in Ecology & Evolution*, **16**, 199–204.
- Kolbe JJ, Glor RE, Rodríguez Schettino L, Lara AC, Larson A, Losos



- JB (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 177–81.
- Kontsedalov S, Zchori-Fein E, Chiel E, Gottlieb Y, Inbar M, Ghanim M (2008) The presence of *Rickettsia* is associated with increased susceptibility of *Bemisia tabaci* (Homoptera: Aleyrodidae) to insecticides. *Pest Management Science*, **64**, 789–792.
- Küchler SM, Dettner K, Kehl S (2010) Molecular characterization and localization of the obligate endosymbiotic bacterium in the birch catkin bug *Kleidocerys resedae* (Heteroptera: Lygaeidae, Ischnorhynchinae). *FEMS Microbiology Ecology*, **73**, 408–418.
- Lang FB, Gray MW, Burger G (1999) MITOCHONDRIAL GENOME EVOLUTION AND THE ORIGIN OF EUKARYOTES.
- Lapointe DA, Atkinson CT, Samuel MD (2012) Ecology and conservation biology of avian malaria. *Annals of the New York Academy of Sciences*, **1249**, 211–226.
- Lavergne S, Molofsky J (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences*, **104**, 3883–8.
- Le Roux JJ, Wicczorek AM, Wright MG, Tran CT (2007) Super-genotype: global monoclonality defies the odds of nature. *PLoS One*, **2**, e590.
- Leclercq A, Hartelt K, Schuster C, Jung K, Kleespies RG (2011) Multilocus sequence typing (MLST) for the infra-generic taxonomic classification of entomopathogenic *Rickettsiella* bacteria. *FEMS Microbiology Letters*, **324**, 125–34.
- Lee CE (1999) Rapid and repeated invasions of fresh water by the copepod *Eurytemora affinis*. *Evolution*, **53**, 1423–1434.
- Lee W, Park J, Lee GS, Lee S, Akimoto Si (2013) Taxonomic status of the *Bemisia tabaci* complex (Hemiptera: Aleyrodidae) and reassessment of the number of its constituent species. *PLoS One*, **8**, e63817.
- Legg JP, Sseruwagi P, Boniface S, et al. (2014) Spatio-temporal patterns of genetic change amongst populations of cassava *Bemisia tabaci* whiteflies driving virus pandemics in East and Central Africa. *Virus Research*, **186**, 61–75.
- Lewontin R, Birch L (1966) Hybridization as a source of variation for adaptation to new environments. *Evolution*, **20**, 315–336.
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–60.
- Liu Ss, Colvin J, De Barro PJ (2012) Species Concepts as Applied to the Whitefly *Bemisia tabaci* Systematics: How Many Species Are There? *Journal of Integrative Agriculture*, **11**, 176–186.
- Liu SS, De Barro PJ, Xu J, et al. (2007) Asymmetric mating interactions drive widespread invasion and displacement in a whitefly. *Science*, **318**, 1769–72.
- Llopart A, Herrig D, Brud E, Stecklein Z (2014) Sequential adaptive introgression of the mitochondrial genome in *Drosophila yakuba* and *Drosophila santomea*. *Molecular Ecology*, **23**, 1124–1136.
- Lo N, Bandi C, Watanabe H, Nalepa C, Beninati T (2003) Evidence for coeladogenesis between diverse dictyopteran lineages and their intracellular endosymbionts. *Molecular Biology and Evolution*, **20**, 907–913.
- Lo N, Evans TA (2007) Phylogenetic diversity of the intracellular symbiont *Wolbachia* in termites. *Molecular Phylogenetics and Evolution*, **44**, 461–466.
- Lockwood JL, Cassey P, Blackburn T (2005) The role of propagule pressure in explaining species invasions. *Trends in Ecology & Evolution*, **20**, 223–8.
- Login FH, Balmund S, Vallier A, et al. (2011) Antimicrobial Peptides Keep Insect Endosymbionts Under Control. *Science*, **334**, 362–365.
- Lombaert E, Guillemaud T, Cornuet JM, Malausa T, Facon B, Estoup A (2010) Bridgehead effect in the worldwide invasion of the bio-control harlequin ladybird. *PLoS One*, **5**, e9743.
- Lozier JD (2014) Revisiting comparisons of genetic diversity in stable and declining species: assessing genome-wide polymorphism in North American bumble bees using RAD sequencing. *Molecular ecology*, **23**, 788–801.
- Luan JB, Chen W, Hasegawa DK, et al. (2015) Metabolic coevolution in the bacterial symbiosis of whiteflies and related plant sap-feeding insects. *Genome Biol Evol*, **7**, 2635–2647.
- Luck GW, Daily GC, Ehrlich PR (2003) Population diversity and ecosystem services. *Trends in Ecology & Evolution*, **18**, 331–336.
- Lugo AE (2004) The outcome of alien tree invasions in Puerto Rico. *Frontiers in Ecology and the Environment*, **2**, 265–273.
- Luo C, Jones CM, Devine G, Zhang F, Denholm I, Gorman K (2010) Insecticide resistance in *Bemisia tabaci* biotype Q (Hemiptera: Aleyrodidae) from China. *Crop Protection*, **29**, 429–434.
- Ma J, Amos CI (2012) Principal components analysis of population admixture. *PLoS One*, **7**.
- Mahadav A, Kontsedalov S, Czosnek H, Ghanim M (2009) Thermotolerance and gene expression following heat stress in the whitefly *Bemisia tabaci* B and Q biotypes. *Insect Biochemistry and Molecular Biology*, **39**, 668–676.
- Malausa T, Fenis A, Warot S, et al. (2011) DNA markers to disentangle complexes of cryptic taxa in mealybugs (Hemiptera: Pseudococcidae). *Journal of Applied Entomology*, **135**, 142–155.
- Mallet J, Beltrán M, Neukirchen W, Linares M (2007) Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC evolutionary biology*, **7**, 28.
- Mallet J, Besansky N, Hahn MW (2015) How reticulated are species? *BioEssays*, **37**, 140–149.
- Margulis L (1970) *Origin of eukaryotic cells*. Yale University Press New Haven.
- Martin JH, Mound La (2007) An annotated check list of the world's whiteflies (Insecta: Hemiptera: Aleyrodidae). *Zootaxa*, **1492**, 1–84.
- Martin SH, Dasmahapatra KK, Nadeau NJ, et al. (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome research*, **23**, 1817–1828.
- Marubayashi JM, Kliot A, Yuki VA, et al. (2014) Diversity and Localization of Bacterial Endosymbionts from Whitefly Species Collected in Brazil. *PLoS One*, **9**, e108363.
- Maruthi MN, Colvin J, Thwaites RM, Banks GK, Gibson G, Seal SE (2004) Reproductive incompatibility and cytochrome oxidase I gene sequence variability amongst host-adapted and geographically separate *Bemisia tabaci* populations (Hemiptera: Aleyrodidae). *Systematic Entomology*, **29**, 560–568.
- Mateos M, Castrezana SJ, Nankivell BJ, Estes AM, Markow TA, Moran NA (2006) Heritable endosymbionts of *Drosophila*. *Genetics*, **174**, 363–376.
- McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.
- McKenzie CL, Bethke Ja, Byrne FJ, et al. (2012) Distribution of *Bemisia tabaci* (Hemiptera: Aleyrodidae) biotypes in North America after the Q invasion. *Journal of economic entomology*, **105**, 753–66.
- McNeill MR, Richards NK, White JA, Laugraud A (2014) Hidden arsenal: endosymbionts in arthropods, their role and possible implications for biological control success. *New Zealand Plant Protection*, **67**, 204–212.
- McVean G (2009) A genealogical interpretation of principal components analysis. *PLoS Genetics*, **5**.
- Merville A, Venner S, Henri H, et al. (2013) Endosymbiont diversity among sibling weevil species competing for the same resource. *BMC evolutionary biology*, **13**, 28.
- Mira A, Moran N (2002) Estimating Population Size and Transmission Bottlenecks in Maternally Transmitted Endosymbiotic Bacteria. *Microbial Ecology*, **44**, 137–143.
- Mitsuhashi W, Saiki T, Wei W, Kawakita H, Sato M (2002) Two novel strains of *Wolbachia* coexisting in both species of mulberry leafhoppers. *Insect Molecular Biology*, **11**, 577–584.
- Montllor C, Maxmen A, Purcell A (2002) Facultative bacterial endosymbionts benefit pea aphids *Acyrtosiphon pisum* under heat stress. *Ecological Entomology*, **27**, 189–195.
- Moran Na, Degnan PH (2006) Functional genomics of *Buchnera* and the ecology of aphid hosts. *Molecular ecology*, **15**, 1251–61.
- Moran Na, Dunbar HE (2006) Sexual acquisition of beneficial symbionts in aphids. *Proceedings of the National Academy of Sci-*



- ences, **103**, 12803–6.
- Moran Na, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics*, **42**, 165–190.
- Moran NA, Tran P, Gerardo NM (2005) Symbiosis and insect diversification: An ancient symbiont of sap-feeding insects from the bacterial phylum Bacteroidetes. *Applied and Environmental Microbiology*, **71**, 8802–8810.
- Moreno S, Beltrán JF, Cotilla I, et al. (2007) Long-term decline of the European wild rabbit (*Oryctolagus cuniculus*) in south-western Spain. *Wildlife Research*, **34**, 652–658.
- Morin S, Williamson MS, Goodson SJ, Brown JK, Tabashnik BE, Dennehy TJ (2002) Mutations in the *Bemisia tabaci* para sodium channel gene associated with resistance to a pyrethroid plus organophosphate mixture. *Insect biochemistry and molecular biology*, **32**, 1781–91.
- Mouton L, Gnaninké O, Henri H, et al. (2015) Detection of genetically isolated entities within the Mediterranean species of *Bemisia tabaci*: new insights into the systematics of this worldwide pest. *Pest management science*.
- Mouton L, Thierry M, Henri H, et al. (2012) Evidence of diversity and recombination in *Arsenophonus* symbionts of the *Bemisia tabaci* species complex. *BMC microbiology*, **12 Suppl 1**, S10.
- Moya A, Guirao P, Cifuentes D, Beitia F, Cenis L (2001) Genetic diversity of Iberian populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae) based on random amplified polymorphic DNA–polymerase chain reaction. *Molecular Ecology*, **10**, 891–897.
- Munson MA, Baumann P, Clark MA, et al. (1991) Evidence for the establishment of aphid-eubacterium endosymbiosis in an ancestor of four aphid families. *Journal of Bacteriology*, **173**, 6321–6324.
- Murat C, Diez J, Luis P, et al. (2004) Polymorphism at the ribosomal DNA ITS and its relation to postglacial re-colonization routes of the Perigord truffle *Tuber melanosporum*. *New Phytologist*, **164**, 401–411.
- Nathan R, Getz WM, Revilla E, et al. (2008) A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, **105**, 19052–19059.
- Nathan R, Giuggioli L (2013) A milestone for movement ecology research. *Movement Ecology*, **1**, 1.
- Nauen R, Vontas J, Kausmann M, Wölfel K (2013) Pymetrozine is hydroxylated by CYP6CM1, a cytochrome P450 conferring neonicotinoid resistance in *Bemisia tabaci*. *Pest management science*, **69**, 457–61.
- Nirgianaki A, Banks GK, Frohlich DR, et al. (2003) *Wolbachia* infections of the whitefly *Bemisia tabaci*. *Current Microbiology*, **47**, 93–101.
- Norris LC, Main BJ, Lee Y, et al. (2015) Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proceedings of the National Academy of Sciences*, **112**, 201418892.
- Nováková E, Hyspa V, Moran Na (2009) *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. *BMC microbiology*, **9**, 143.
- Oliveira M, Henneberry T, Anderson P (2001) History, current status, and collaborative research projects for *Bemisia tabaci*. *Crop Protection*, **20**, 709–723.
- Oliver KM, Degnan PH, Burke GR, Moran Na (2010) Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annual Review of Entomology*, **55**, 247–66.
- Oliver KM, Moran Na, Hunter MS (2006) Costs and benefits of a superinfection of facultative symbionts in aphids. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 1273–80.
- Oliver KM, Russell JA, Moran Na, Hunter MS (2003) Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the National Academy of Sciences*, **100**, 1803–7.
- Pais R, Lohs C, Wu Y, Wang J, Aksoy S (2008) The obligate mutualist *Wigglesworthia glossinidia* influences reproduction, digestion, and immunity processes of its host, the tsetse fly. *Applied and Environmental Microbiology*, **74**, 5965–5974.
- Pan H, Li X, Ge D, et al. (2012) Factors affecting population dynamics of maternally transmitted endosymbionts in *Bemisia tabaci*. *PLoS One*, **7**, e30760.
- Pandey N, Rajagopal R (2015) Molecular characterization and diversity analysis of bacterial communities associated with *Dialeurolonga malleswaramensis* (Hemiptera: Aleyrodidae) adults using 16S rDNA amplicon pyrosequencing and FISH. *Insect science*, **S**, 1–8.
- Papayiannis L, Brown J, Seraphides N, Hadjistyli M, Ioannou N, Katis N (2009) A real-time PCR assay to differentiate the B and Q biotypes of the *Bemisia tabaci* complex in Cyprus. *Bulletin of Entomological Research*, **99**, 573.
- Park J, Jahan SMH, Song WG, et al. (2012) Identification of biotypes and secondary endosymbionts of *Bemisia tabaci* in Korea and relationships with the occurrence of TYLCV disease. *Journal of Asia-Pacific Entomology*, **15**, 186–191.
- Parrella G, Nappo AG, Manco E, Greco B, Giorgini M (2013) Invasion of the Q2 mitochondrial variant of Mediterranean *Bemisia tabaci* in southern Italy: possible role of bacterial endosymbionts. *Pest management science*, **70**, 1514–1523.
- Parrella G, Scassillo L, Giorgini M (2012) Evidence for a new genetic variant in the *Bemisia tabaci* species complex and the prevalence of the biotype Q in southern Italy. *Journal of Pest Science*, **85**, 227–238.
- Pascual S, Callejas C (2004) Intra- and interspecific competition between biotypes B and Q of *Bemisia tabaci* (Hemiptera: Aleyrodidae) from Spain. *Bulletin of Entomological Research*, **94**, 369–375.
- Perlman SJ, Hunter MS, Zchori-Fein E (2006) The emerging diversity of *Rickettsia*. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 2097–106.
- Perotti MA, Allen JM, Reed DL, Braig HR (2007) Host-symbiont interactions of the primary endosymbiont of human head and body lice. *The FASEB journal*, **21**, 1058–1066.
- Perring TM, Cooper AD, Rodriguez RJ, Farrar CA, Bellows TS (1993) Identification of a whitefly species by genomic and behavioral studies. *Science*, **259**, 74–77.
- Perry W, Feder J, Dwyer G, Lodge D (2001) Hybrid zone dynamics and species replacement between *Orconectes* crayfishes in a northern Wisconsin lake. *Evolution*, **55**, 1153–1166.
- Petit RJ, Aguinagalde I, De Beaulieu JL, et al. (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.
- Pfenninger M, Schwenk K (2007) Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC evolutionary biology*, **7**.
- Pimentel D, Lach L, Zuniga R, Morrison D (2000) Environmental and economic costs of nonindigenous species in the United States. *BioScience*, **50**.
- Polston JE, De Barro P, Boykin LM (2014) Transmission specificities of plant viruses with the newly identified species of the *Bemisia tabaci* species complex. *Pest Management Science*, **70**, 1547–1552.
- Pons JM, Sonsthagen S, Dove C, Crochet Pa (2014) Extensive mitochondrial introgression in North American Great Black-backed Gulls (*Larus marinus*) from the American Herring Gull (*Larus smithsonianus*) with little nuclear DNA impact. *Heredity*, **112**, 226–39.
- Powell ME, Cuthbertson AGS, Bell HA, Boonham N, Morris J, Northing P (2012) First record of the Q Biotype of the sweetpotato whitefly, *Bemisia tabaci*, intercepted in the UK. *European Journal of Plant Pathology*, **133**, 797–801.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–59.
- Purcell aH, Suslow KG, Klein M (1994) Transmission via plants of an insect pathogenic bacterium that does not multiply or move in plants. *Microbial Ecology*, **27**, 19–26.
- Pysek P (1995) On the Terminology Used in Plant Invasion Studies. *Plant Invasions - General Aspects and Special Problems*, **S**, 71–81.
- R Core Team (2015) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

- Raina HS, Rawal V, Singh S, Daimei G, Shakarad M, Rajagopal R (2015) Elimination of *Arsenophonus* and decrease in the bacterial symbionts diversity by antibiotic treatment leads to increase in fitness of whitefly, *Bemisia tabaci*. *Infection, Genetics and Evolution*, **32**, 224–230.
- Rana VS, Singh ST, Priya NG, Kumar J, Rajagopal R (2012) *Arsenophonus* GroEL interacts with CLCuV and Is localized in midgut and salivary gland of whitefly *B. tabaci*. *PLoS One*, **7**.
- Rao Q, Rollat-Farnier PA, Zhu DT, et al. (2015) Genome reduction and potential metabolic complementation of the dual endosymbionts in the whitefly *Bemisia tabaci*. *BMC Genomics*, **16**, 1–13.
- Rao Q, Wang S, Zhu DT, Wang XW, Liu SS (2012) Draft genome sequence of *Rickettsia* sp. strain MEAM1, isolated from the whitefly *Bemisia tabaci*. *Journal of Bacteriology*, **194**, 4741–4742.
- Ratzka C, Liang C, Dandekar T, Gross R, Feldhaar H (2011) Immune response of the ant *Camponotus floridanus* against pathogens and its obligate mutualistic endosymbiont. *Insect Biochemistry and Molecular Biology*, **41**, 529–536.
- Rauch N, Nauen R (2004) Characterization and molecular cloning of a glutathione S-transferase from the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Insect biochemistry and molecular biology*, **34**, 321–9.
- Raymond M, Rousset F (1995) An exact test for population differentiation. *Evolution*, **49**, 1280–1283.
- Rey O, Estoup A, Vonshak M, et al. (2012) Where do adaptive shifts occur during invasion? A multidisciplinary approach to unravelling cold adaptation in a tropical ant species invading the Mediterranean area. *Ecology Letters*, **15**, 1266–1275.
- Reynaud P (2000) L'aleurode *Bemisia tabaci* en France. Situation actuelle et possibilités de développement. *Phytoma*, **527**, 18–21.
- Ricciardi A, Neves RJ, Rasmussen JB (1998) Impending extinctions of North American freshwater mussels (Unionoida) following the zebra mussel (*Dreissena polymorpha*) invasion. *Journal of Animal Ecology*, **67**, 613–619.
- Richardson DM, Allsopp N, D'Antonio CM, Milton SJ, Rejmánek M (2000) Plant invasions—the role of mutualisms. *Biological Reviews*, **75**, 65–93.
- Rieppel O (2002) A case of dispersing chameleons. *Nature*, **415**, 744–745.
- Rieseberg LH (2009) Evolution: Replacing Genes and Traits through Hybridization.
- Rigaud T, Soutygrosset C, Raimond R, Mocquard JP, Juchault P (1991) Feminizing endocytobiosis in the terrestrial crustacean *Armadillidium vulgare* Latr (isopoda) - recent acquisitions. *Endocytobiosis and Cell Research*, **7**, 259–273.
- Rius M, Darling JA (2014) How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology & Evolution*, **29**, 233–242.
- Rocha EPC, Danchin A (2002) Base composition bias might result from competition for metabolic resources. *Trends in Genetics*, **18**, 291–294.
- Roditakis E, Roditakis NE, Tsagkarakou A (2005) Insecticide resistance in *Bemisia tabaci* (Homoptera: Aleyrodidae) populations from Crete. *Pest Management Science*, **61**, 577–582.
- Roditakis E, Tsagkarakou A, Vontas J (2006) Identification of mutations in the para sodium channel of *Bemisia tabaci* from Crete, associated with resistance to pyrethroids. *Pesticide Biochemistry and Physiology*, **85**, 161–166.
- Rollat-Farnier PA, Santos-García D, Rao Q, et al. (2015) Two host clades, two bacterial arsenals: Evolution through gene losses in facultative endosymbionts. *Genome Biology and Evolution*, **7**, 839–855.
- Rosenberg N, Pritchard J, Weber J, et al. (2002) Genetic structure of human populations. *Science*, **298**, 2381–2385.
- Rua P, Simon B, Cifuentes D, Martínez-Mora C, Cenis JL (2006) New insights into the mitochondrial phylogeny of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in the Mediterranean Basin. *Journal of Zoological Systematics and Evolutionary Research*, **44**, 25–33.
- Rubin BER, Ree RH, Moreau CS (2012) Inferring phylogenies from RAD sequence data. *PLoS One*, **7**, 1–12.
- Rudgers JA, Mattingly WB, Koslow JM (2005) Mutualistic fungus promotes plant invasion into diverse communities. *Oecologia*, **144**, 463–471.
- Russell Ja, Latorre A, Sabater-Muñoz B, Moya A, Moran Na (2003) Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. *Molecular ecology*, **12**, 1061–75.
- Russell Ja, Moran Na (2006) Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 603–610.
- Sacchi L, Genchi M, Clementi E, et al. (2008) Multiple symbiosis in the leafhopper *Scaphoideus titanus* (Hemiptera: Cicadellidae): Details of transovarial transmission of *Cardinium* sp. and yeast-like endosymbionts. *Tissue and Cell*, **40**, 231–242.
- Sáez AG, Lozano E (2005) Body doubles. *Nature*, **433**, 111.
- Sakai AK, Allendorf FW, Holt JS, et al. (2001) The population biology of invasive species. *Annual Review of Ecology and Systematics*, **32**, 305–332.
- Sakurai M, Koga R, Tsuchida T, Meng Xy, Fukatsu T (2005) Symbiont in the Pea Aphid. *Environmental Microbiology*, **71**, 4069–4075.
- Saleh D, Laarif A, Clouet C, Gauthier N (2012) Spatial and host-plant partitioning between coexisting *Bemisia tabaci* cryptic species in Tunisia. *Population Ecology*, **54**, 261–274.
- Sandström JP, Russell Ja, White JP, Moran Na (2001) Independent origins and horizontal transfer of bacterial symbionts of aphids. *Molecular Ecology*, **10**, 217–228.
- Santos-García D, Farnier PA, Beitia F, et al. (2012) Complete genome sequence of "Candidatus Portiera aleyrodidarum" BT-QVLC, an obligate symbiont that supplies amino acids and carotenoids to *Bemisia tabaci*. *Journal of Bacteriology*, **194**, 6654–6655.
- Santos-García D, Latorre A, Moya A, et al. (2014) Small but powerful, the primary endosymbiont of moss bugs, *Candidatus Evansia muelleri*, holds a reduced genome with large biosynthetic capabilities. *Genome biology and evolution*, **6**, 1875–93.
- Santos-García D, Vargas-Chavez C, Moya A, Latorre A, Silva FJ (2015) Genome evolution in the primary endosymbiont of whiteflies sheds light on their divergence. *Genome Biology and Evolution*, **7**, 873–888.
- Saxena RC, Barrion AA (1987) Biotypes of insect pests of agricultural crops. *International Journal of Tropical Insect Science*, **8**, 453–458.
- Scarborough CL, Ferrari J, Godfray HCJ (2005) Aphid protected from pathogen by endosymbiont. *Science*, **310**, 1781.
- Schlaepfer MA, Sax DF, Olden JD (2011) El Valor de Conservación Potencial de Especies No Nativas. *Conservation Biology*, **25**, 428–437.
- Shaw KL (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences*, **99**, 16122–16127.
- Shine R (2010) The ecological impact of invasive cane toads *Bufo marinus* in Australia. *The Quarterly Review of Biology*, **85**, 253–291.
- Silva-Sanchez C (1997) Behavior of the silverleaf whitefly (*Bemisia argentifolii* B&P) in the cotton crop on the coast of Hermosillo, Sonora. In *Proceedings of the International Cotton Pest Work Committee*, S. 34.
- Simberloff D, Gibbons L (2004) Now you see them, now you don't! - Population crashes of established introduced species. *Biological Invasions*, **6**, 161–172.
- Simberloff D, Martin JI, Genovesi P, et al. (2013) Impacts of biological invasions: what's what and the way forward. *Trends in Ecology & Evolution*, S. 1–9.
- Simmons AM, Harrison HF, Ling KS (2008) Forty-nine new host plant species for *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Entomological Science*, **11**, 385–390.
- Simón B, Cenis JL, De La Rúa P (2007) Distribution patterns of the Q and B biotypes of *Bemisia tabaci* in the Mediterranean Basin based on microsatellite variation. *Entomologia Experimentalis et*

- Applicata*, **124**, 327–336.
- Simón B, Cenis JL, Demichelis S, Rapisarda C, Caciagli P, Bosco D (2003) Survey of *Bemisia tabaci* (Hemiptera: Aleyrodidae) biotypes in Italy with the description of a new biotype (T) from *Euphorbia characias*. *Bulletin of entomological research*, **93**, 259–64.
- Simon B, Moriones E, Soria C, Beitia F, Bosco D, Cenis J (1999) Variación genética de poblaciones de *Bemisia tabaci* (Gennadius) en la Cuenca Mediterránea occidental. In *Proceedings 7th Spanish National Congress*.
- Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 652–701.
- Simon JC, Carré S, Boutin M, et al. (2003) Host-based divergence in populations of the pea aphid: insights from nuclear markers and the prevalence of facultative symbionts. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 1703–12.
- Singh ST, Kumar J, Thomas A, Ramamurthy VV, Rajagopal R (2013) Detection and localization of *Rickettsia* sp in mealybug. *Environmental entomology*, **42**, 711–6.
- Singh ST, Priya NG, Kumar J, et al. (2012) Diversity and phylogenetic analysis of endosymbiotic bacteria from field caught *Bemisia tabaci* from different locations of North India based on 16S rDNA library screening. *Infection, Genetics and Evolution*, **12**, 411–9.
- Skaljic M, Zanic K, Ban SG, Kontsedalov S, Ghanim M (2010) Co-infection and localization of secondary symbionts in two whitefly species. *BMC microbiology*, **10**, 142.
- Skaljic M, Zanic K, Hrnčić S, Radonjić S, Perović T, Ghanim M (2013) Diversity and localization of bacterial symbionts in three whitefly species (Hemiptera: Aleyrodidae) from the east coast of the Adriatic Sea. *Bulletin of entomological research*, **103**, 48–59.
- Skoracka A, Kuczyski L, Santos De Mendona R, et al. (2012) Cryptic species within the wheat curl mite *Aceria tosichella* (Keifer) (Acari: Eriophyoidea), revealed by mitochondrial, nuclear and morphometric data. *Invertebrate Systematics*, **26**, 417–433.
- Slatkin M (2008) Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, **9**, 477–85.
- Sloan DB, Moran Na (2012) Endosymbiotic bacteria as a source of carotenoids in whiteflies. *Biology letters*, **8**, 986–989.
- Stansly PA, Naranjo SE, Brown JK, et al. (2010) *Bemisia*: *Bionomics and management of a global pest*. Springer.
- Stapley J, Reger J, Feulner PG, et al. (2010) Adaptation genomics: the next generation. *Trends in Ecology & Evolution*, **25**, 705–712.
- Streicher JW, Devitt TJ, Goldberg CS, Malone JH, Blackmon H, Fujita MK (2014) Diversification and asymmetrical gene flow across time and space: Lineage sorting and hybridization in polytypic barking frogs. *Molecular Ecology*, **23**, 3273–3291.
- Stromberg JC, Chew MK, Nagler PL, Glenn EP (2009) Changing perceptions of change: The role of scientists in tamarix and river management. *Restoration Ecology*, **17**, 177–186.
- Stuart BL, Inger RF, Voris HK (2006) High level of cryptic species diversity revealed by sympatric lineages of Southeast Asian forest frogs. *Biology letters*, **2**, 470–4.
- Su Q, Oliver KM, Pan H, et al. (2013a) Facultative Symbiont *Hamiltonella* Confers Benefits to *Bemisia tabaci* (Hemiptera: Aleyrodidae), an Invasive Agricultural Pest Worldwide. *Environmental Entomology*, **42**, 1265–71.
- Su Q, Pan H, Liu B, et al. (2013b) Insect symbiont facilitates vector acquisition, retention, and transmission of plant virus. *Scientific reports*, **3**, 1367.
- Su Q, Xie W, Wang S, Wu Q, Ghanim M, Zhang Y (2014a) Location of Symbionts in the Whitefly *Bemisia tabaci* Affects Their Densities during Host Development and Environmental Stress. *PLoS One*, **9**, e91802.
- Su Q, Xie W, Wang S, et al. (2014b) The endosymbiont *Hamiltonella* increases the growth rate of its host *Bemisia tabaci* during periods of nutritional stress. *PLoS One*, **9**, 1–6.
- Subandiyah S, Nikoh N, Tsuyumu S, Somowiyarjo S, Fukatsu T (2000) Complex Endosymbiotic Microbiota of the Citrus Psyllid *Diuraphis citri* (Homoptera: Psylloidea). *Zoological Science*, **17**, 983–989.
- Suh A, Smeds L, Ellegren H (2015) The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS Biology*, **13**, 1–18.
- Sun DB, Xu J, Luan JB, Liu SS (2011) Reproductive incompatibility between the B and Q biotypes of the whitefly *Bemisia tabaci* in China: genetic and behavioural evidence. *Bulletin of entomological research*, **101**, 211–220.
- Szklarzewicz T, Moskal A (2001) Ultrastructure, distribution, and transmission of endosymbionts in the whitefly *Aleurochiton aceris* Modeer (Insecta, Hemiptera, Aleyrodinae). *Protoplasma*, **218**, 45–53.
- Tahiri A, Halkett F, Granier M, Gueguen G, Peterschmitt M (2013) Evidence of gene flow between sympatric populations of the Middle East-Asia Minor 1 and Mediterranean putative species of *Bemisia tabaci*. *Ecology and Evolution*, **3**, 2619–2633.
- Tahiri a, Sekkat A, Bennani A, Granier M, Delvare G, Peterschmitt M (2006) Distribution of tomato-infecting begomoviruses and *Bemisia tabaci* biotypes in Morocco. *Annals of Applied Biology*, **149**, 175–186.
- Tajebe L, Guastella D, Cavalieri V, et al. (2014) Diversity of symbiotic bacteria associated with *Bemisia tabaci* (Homoptera: Aleyrodidae) in cassava mosaic disease pandemic areas of Tanzania. *Annals of Applied Biology*, **166**, n/a–n/a.
- Takahashi T, Nagata N, Sota T (2014) Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. *Molecular phylogenetics and evolution*, **80**, 1–8.
- Tassin J (2014) *Grande invasion (La) : Qui a peur des espèces invasives ?* Odile Jacob.
- Tay WT, Evans Ga, Boykin LM, De Barro PJ (2012) Will the real *Bemisia tabaci* please stand up? *PLoS One*, **7**, e50550.
- Taylor DR, Keller SR (2007) Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution*, **61**, 334–45.
- Teixeira L, Ferreira A, Ashburner M (2008) The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLoS Biology*, **6**, 2753–2763.
- Terraz G, Gueguen G, Arnó J, Fleury F, Mouton L (2014) Nuclear and cytoplasmic differentiation among Mediterranean populations of *Bemisia tabaci*: testing the biological relevance of cytotypes. *Pest management science*, **70**, 1503–1513.
- Thao M, Baumann P (2004a) Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. *Environmental Microbiology*, **70**, 3401–3406.
- Thao ML, Baumann L, Hess JM, et al. (2003) Phylogenetic evidence for two new insect-associated Chlamydia of the family Simkaniaceae. *Current microbiology*, **47**, 46–50.
- Thao ML, Baumann P (2004b) Evidence for multiple acquisition of *Arsenophonus* by whitefly species (Sternorrhyncha: Aleyrodidae). *Current microbiology*, **48**, 140–4.
- Thao ML, Moran NA, Abbot P, Brennan EB, Burckhardt DH, Baumann P (2000) Cospeciation of psyllids and their primary prokaryotic endosymbionts. *Applied and Environmental Microbiology*, **66**, 2898–2905.
- Thierry M, Becker N, Hajri A, Reynaud B, Lett JM, Delatte H (2011) Symbiont diversity and non-random hybridization among indigenous (Ms) and invasive (B) biotypes of *Bemisia tabaci*. *Molecular ecology*, **20**, 2172–87.
- Thierry M, Bile A, Grondin M, Reynaud B, Becker N, Delatte H (2015) Mitochondrial, nuclear, and endosymbiotic diversity of two recently introduced populations of the invasive <i> Bemisia tabaci</i> MED species in La Réunion. *Insect Conservation and Diversity*, **8**, 71–80.
- Thomas A, Kar A, Rebijith KB, Asokan R, Ramamurthy VV (2014) *Bemisia tabaci* ( Hemiptera : Aleyrodidae ) Species Complex from Cotton Cultivars : A Comparative Study of Population Density , Morphology , and Molecular Variations *Bemisia tabaci* ( Hemiptera : Aleyrodidae ) Species Complex From Cotton Cultivars : A Compa. *Annals of the Entomological Society of America*,



- 107, 389–398.
- Thomas-Orillard M (1996) A virus- *Drosophila* association: The first steps towards co-evolution? *Biodiversity and Conservation*, **5**, 1015–1021.
- Thompson K, Hodgson G J, Rich C G T (1995) Native and alien invasive plants: more of the same? *Ecography*, **18**, 390–402.
- Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**, 3907–3930.
- Tsagkarakou a, Mouton L, Kristoffersen JB, Dokianakis E, Grispu M, Bourtzis K (2012) Population genetic structure and secondary endosymbionts of Q *Bemisia tabaci* (Hemiptera: Aleyrodidae) from Greece. *Bulletin of entomological research*, **102**, 353–65.
- Tsagkarakou A, Nikou D, Roditakis E, Sharvit M, Morin S, Vontas J (2009) Molecular diagnostics for detecting pyrethroid and organophosphate resistance mutations in the Q biotype of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Pesticide Biochemistry and Physiology*, **94**, 49–54.
- Tsagkarakou a, Roditakis N (2003) Isolation and characterization of microsatellite loci in *Bemisia tabaci* (Hemiptera: Aleyrodidae). *Molecular Ecology Notes*, **3**, 196–198.
- Tsagkarakou a, Tsigenopoulos CS, Gorman K, Lagnel J, Bedford ID (2007) Biotype status and genetic polymorphism of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Greece: mitochondrial DNA and microsatellites. *Bulletin of entomological research*, **97**, 29–40.
- Tsuchida T, Koga R, Horikawa M, *et al.* (2010) Symbiotic bacterium modifies aphid body color. *Science*, **330**, 1102–1104.
- Ueda S, Brown J (2006) First report of the Q biotype of *Bemisia tabaci* in Japan by mitochondrial cytochrome oxidase I sequence analysis. *Phytoparasitica*, **34**, 405–411.
- Unterman BM, Baumann P, McLean DL (1989) Pea aphid symbiont relationships established by analysis of 16S rRNAs. *Journal of Bacteriology*, **171**, 2970–2974.
- Vassiliou V, Emmanouilidou M, Perrakis A, *et al.* (2011) Insecticide resistance in *Bemisia tabaci* from Cyprus. *Insect Science*, **18**, 30–39.
- Vassiliou V, Jagge C, Grispu M (2008) Note: Biotype status of *Bemisia tabaci* from various crops in Cyprus. *Phytoparasitica*, **36**, 400–404.
- Vavre F, Fleury F, Lepetit D, Fouillet P, Boulétreau M (1999) Phylogenetic evidence for horizontal transmission of *Wolbachia* in host-parasitoid associations. *Molecular biology and evolution*, **16**, 1711–23.
- Vitousek P, D’Antonio C (1997) Introduced species: a significant component of human-caused global change. *New Zealand Journal of Ecology*, **21**, 1–16.
- Vorburger C, Gehrler R, Rodríguez P (2010) A strain of the bacterial symbiont *Regiella insecticola* protects aphids against parasitoids. *Biology letters*, **6**, 109–111.
- Wade MJ (2007) The co-evolutionary genetics of ecological communities. *Nature Reviews Genetics*, **8**, 185–95.
- Wade MJ, Goodnight CJ (2006) Cyto-Nuclear Epistasis: Two-Locus Random Genetic Drift in Hermaphroditic and Dioecious Species. *Evolution*, **60**, 643.
- Wang XW, Luan JB, Li JM, Su YL, Xia J, Liu SS (2011) Transcriptome analysis and comparison reveal divergence between two invasive whitefly cryptic species. *BMC genomics*, **12**, 458.
- Wang Z, Yan H, Yang Y, Wu Y (2010) Biotype and insecticide resistance status of the whitefly *Bemisia tabaci* from China. *Pest Management Science*, **66**, 1360–1366.
- Webb D (1985) What are the criteria for presuming native status? *Watsonia*, **15**, 231–236.
- Weeks AR, Velten R, Stouthamer R (2003) Incidence of a new sex-ratio-distorting endosymbiotic bacterium among arthropods. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 1857–65.
- Weinert LA, Araujo-jnr EV, Ahmed MZ, Welch JJ (2015) The incidence of bacterial endosymbionts in terrestrial arthropods. *Proceedings of the Royal Society B: Biological Sciences*, **282**, 3–8.
- Weinert La, Tinsley MC, Temperley M, Jiggins FM (2007) Are we underestimating the diversity and incidence of insect bacterial symbionts? A case study in ladybird beetles. *Biology Letters*, **3**, 678–681.
- Weinert La, Werren JH, Aebi A, Stone GN, Jiggins FM (2009) Evolution and diversity of *Rickettsia* bacteria. *BMC biology*, **7**, 6.
- Weir B, Cockerham C (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Werren J, Hurst G, Zhang W, Breeuwer J, Stouthamer R, Majerus M (1994) *Rickettsial* relative associated with male killing in the ladybird beetle (*Adalia bipunctata*). *Journal of Bacteriology*, **176**, 388–394.
- Werren JH, Baldo L, Clark ME (2008) *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology*, **6**, 741–751.
- Whitfield CW, Behura SK, Berlocher SH, *et al.* (2006) Thrice out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera*. *Science*, **314**, 642–5.
- Wilcove DS, Rothstein D, Dubow J, Phillips A, Losos E (1998) Quantifying threats to imperiled species in the United States. *BioScience*, **48**, 607–615.
- Wilkes TE, Darby AC, Choi JH, Colbourne JK, Werren JH, Hurst GDD (2010) The draft genome sequence of *Arsenophonus nasoniae*, son-killer bacterium of *Nasonia vitripennis*, reveals genes associated with virulence and symbiosis. *Insect Molecular Biology*, **19**, 59–73.
- Williamson M, Fitter A (1996) The varying success of invaders. *Ecology*, **77**, 1661–1666.
- Wisniewski JM, Bockrath KD, Wares JP, Fritts AK, Hill MJ (2013) The Mussel–Fish Relationship: A Potential New Twist in North America? *Transactions of the American Fisheries Society*, **142**, 642–648.
- Xiao N, Pan LL, Zhang CR, Shan HW, Liu SS (2016) Differential tolerance capacity to unfavourable low and high temperatures between two invasive whiteflies. *Scientific Reports*, **6**, 24306.
- Xu J, De Barro PJ, Liu SS (2010) Reproductive incompatibility among genetic groups of *Bemisia tabaci* supports the proposition that the whitefly is a cryptic species complex. *Bulletin of entomological research*, **100**, 359–66.
- Xu J, Wang P (2015) Mitochondrial inheritance in basidiomycete fungi. *Fungal Biology Reviews*.
- Xue X, Li SJ, Ahmed MZ, De Barro PJ, Ren SX, Qiu BL (2012) Inactivation of *Wolbachia* reveals its biological roles in whitefly host. *PLoS One*, **7**, e48148.
- Ye XD, Su YL, Zhao QY, Xia WQ, Liu SS, Wang XW (2014) Transcriptomic analyses reveal the adaptive features and biological differences of guts from two invasive whitefly species. *BMC genomics*, **15**, 370.
- Zakas C, Jones K, Wares JP (2014) Homogeneous nuclear background for mitochondrial cline in northern range of *Notochthamalus scabrosus*. *G3*, **4**, 225–30.
- Zchori-Fein E, Brown JK (2002) Diversity of Prokaryotes Associated with *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae). *Annals of the Entomological Society of America*, **95**, 711–718.
- Zchori-Fein E, Gottlieb Y, Kelly SE, *et al.* (2001) A newly discovered bacterium associated with parthenogenesis and a change in host selection behavior in parasitoid wasps. *Proceedings of the National Academy of Sciences*, **98**, 12555–12560.
- Zchori-Fein E, Lahav T, Freilich S (2014) Variations in the identity and complexity of endosymbiont combinations in whitefly hosts. *Frontiers in Microbiology*, **5**, 310.
- Zchori-Fein E, Perlman S (2004) Distribution of the bacterial symbiont *Cardinium* in arthropods. *Molecular Ecology*, **13**, 2009–2016.
- Zhang L, Zhang Y, Zhang W (2005) Analysis of genetic diversity among different geographical populations and determination of biotypes of *Bemisia tabaci* in China. *Journal of Applied Entomology*, **129**, 121–128.
- Zhou W, Rousset F, O’Neill S (1998) Phylogeny and PCR-based classification of *Wolbachia* strains using wsp gene sequences. *Proceedings of the Royal Society B: Biological Sciences*, **265**, 509–515.

- Zhu Y, Fournier PE, Ereemeeva M, Raoult D (2005) Proposal to create subspecies of *Rickettsia conorii* based on multi-locus sequence typing and an emended description of *Rickettsia conorii*. *BMC microbiology*, **5**, 11.
- Zindel R, Gottlieb Y, Aebi A (2011) Arthropod symbioses: a neglected parameter in pest- and disease-control programmes. *Journal of Applied Ecology*, **48**, 864–872.
- Zug R, Hammerstein P (2012) Still a Host of Hosts for *Wolbachia*: Analysis of Recent Data Suggests That 40% of Terrestrial Arthropod Species Are Infected. *PLoS One*, **7**, e38544.

# Liste des figures

1	Etapas de l'invasion	10
2	Analyse multivariée de l'espace morphologique	14
3	Tri incomplet des lignées	17
4	Espèces cryptiques	18
5	Bactériomes	24
6	Phylogénie des endosymbiotes	24
7	Symbiotes et plantes chez le puceron	28
8	Symptômes de virus phytopathogènes	28
9	Incongruences entre mitochondrie et symbiotes	30
10	Manipulation de la reproduction	31
11	Pupes du genre <i>Bemisia</i>	39
12	<i>B. tabaci</i> : photos, cycle de vie et stylet	40
13	Distribution des distances mitochondriales	43
14	Phylogénie de <i>B. tabaci</i>	44
15	Distribution mondiale de <i>B. tabaci</i>	44
16	Divergences intraspécifiques	46
17	Différentes phylogénies de <i>B. tabaci</i>	48
18	Bactériomes	49
19	Exemples de cytotypes	51
20	Prevalence des bactéries chez Q1	52
21	Localisation d' <i>Arsenophonus</i>	53
22	Localisation de <i>Cardinium</i>	53
23	Localisation d' <i>Hamiltonella</i>	54
24	Localisation de <i>Rickettsia</i>	55
25	Localisation de <i>Wolbachia</i>	56
26	Élevage de <i>B. tabaci</i> au laboratoire : cages et boîtes de Pétri	57
27	Profils PCR RFLP	59
1.1	Remplacement de l'espèce AUSTRALIA par l'espèce MEAM 1 en Australie	65
1.2	Dynamique des espèces à l'échelle du bassin méditerranéen	66
1.3	Épisodes invasifs successifs par les espèces potentielles MEAM 1 et MED en Chine	66
1.4	Haplogroupes mitochondriaux en Méditerranée	69
1.5	Arbre phylogénétique <i>mtCOI</i>	70
1.6	Infections bactériennes autour du bassin méditerranéen	71
1.7	Distribution des <i>mtCOI</i> en France	73
1.8	Données symbiotiques des populations françaises	75
2.1	Distribution des cytotypes en Afrique de l'ouest	86
2.2	Infections des cytotypes présents en Afrique de l'ouest	87
2.3	Prélèvements effectués au Burkina-Faso	88



2.4	Structure génétique au Burkina-Faso	89
2.5	ACP sur les cytotypes de références	95
2.6	Vraisemblance des résultats de STRUCTURE des données européennes	96
2.7	Barplot structure	96
2.8	ACP sur les cytotypes en sympatrie	97
2.9	ACP, microsatellites global	98
2.10	Vraisemblance et données microsatellites complètes	99
2.11	Barplot structure	99
2.12	Probabilités d'assignement au groupe grec pour les individus des populations sympatriques	101
2.13	Distribution des distances mitochondriales	102
2.14	Phylogénie de l'espèce MED	103
3.1	Détection de la nouvelle association	111
3.2	<i>Rickettsia</i> dans le complexe <i>B. tabaci</i>	113
3.3	Souches de <i>Rickettsia</i>	115
3.4	Transfert de <i>Rickettsia</i> par la plante : schéma	117
4.1	Principe du <i>RADseq</i>	132
4.2	Nombre de lectures par espèce	135
4.3	Nombre de lectures en fonction de la quantité d'ADN initiale	136
4.4	Nombre de locus par individu, couverture	137
4.5	Locus en commun	139
4.6	Arbre de distance <i>COI</i>	142
4.7	Qualité et composition en bases des lectures	146
4.8	Quantité d'ADN, population, biotype et nombre de lectures	147
4.9	Nombre de locus	148
4.10	Hétérozygotie et couverture	149
4.11	Locus partagés	150
4.12	Détection des symbiotes dans les populations	151
4.13	Proxy de la densité en symbiote par population	153
4.14	ACP	154
4.15	Influence des données manquantes sur l'ACP	155
4.16	Structure : vraisemblance et $\Delta K$	156
4.17	Structure	158
4.18	Carte des populations qui vont être génotypées par <i>RADSeq</i>	160
A.2	Chevauchement des deux <i>reads</i>	229
A.3	Densité en SNP sur les locus	233
A.4	Densité de SNP et longueur de lecture	234

# Liste des tableaux

1	Amorces pour la détection des symbiotes	59
1.1	Échantillons méditerranéens (France exclue)	68
1.2	Échantillons en France	72
1.3	Plante hôte	74
2.1	Amorces microsattellites	85
2.2	Échantillons du Burkina-Faso	88
2.3	Gènes de résistance au Burkina-Faso	90
2.4	Croisements Q1 - Q2	93
2.5	Données microsattellites des populations européennes	94
2.6	Gènes de résistance	95
3.1	Echantillons génotypés pour le MLST	114
3.2	Transfert de <i>Rickettsia</i> dans la plante	118
4.1	Individus de l'expérience pilote	134
4.2	Estimation du nombre d'individus multiplexables	139
4.3	Populations utilisées	141
4.4	Séquençage	145
4.5	Lectures disponibles après filtrage	147
4.6	Lectures par symbiotes	152
4.7	Distances génétiques ( $F_{ST}$ )	157
4.8	Tableau d'AMOVA	159
A.1	Mapping des lectures de drosophiles	231



# Table des matières

<b>Résumé</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>Remerciements</b>	<b>ix</b>
<b>Sommaire</b>	<b>xiii</b>
<b>Introduction</b>	<b>1</b>
1 La génétique des invasions biologiques	3
1.1 Définition et notion d'espèce invasive	3
1.1.1 Qu'est ce qu'une espèce indigène ?	3
1.1.2 Les conséquences des invasions peuvent être plus subtiles, voire positives	5
1.1.3 Des effets variables dans le temps et l'espace	7
1.2 Déroulement d'une invasion	8
1.3 Qu'entend on par adaptation ?	10
1.3.1 L'admixture, facteur clé dans les processus adaptatifs	12
1.3.2 Augmentation du potentiel adaptatif	13
1.4 Complexe d'espèces et invasions	14
1.4.1 Comment définit-on une espèce ?	14
1.4.2 Complexe d'espèces et invasion	15
1.5 Interaction avec le compartiment cytoplasmique	19
2 La symbiose, actrice de l'invasion ?	20
2.1 Définition, brève histoire du concept de symbiose	20
2.2 Transmission de ces endosymbiotes	22
2.3 Les symbiotes obligatoires	22
2.3.1 Localisation tissulaire	23
2.3.2 Diversité et rôles de ces symbiotes	23
2.3.3 Les symbioses primaires sont des associations anciennes	25
2.4 Les symbiotes secondaires	26
2.4.1 Les effets phénotypiques de ces symbiotes	27
2.4.2 La transmission horizontale des symbiotes secondaires	29
2.4.3 Les multi-infections	29
2.5 Les manipulateurs de la reproduction	31
2.6 Quel rôle peut jouer la symbiose dans l'invasion biologique ?	32
2.6.1 Résistance contre des moyens de lutte	32
2.6.2 Augmentation du potentiel invasif de l'hôte	32
<b>Problématique générale</b>	<b>35</b>

<b>Le modèle d'étude <i>Bemisia tabaci</i></b>	<b>37</b>
1 Description écologique	39
2 Des races d'hôtes au complexe d'espèces	41
3 Les associations symbiotiques chez <i>Bemisia tabaci</i>	48
3.1 Le symbiote primaire	48
3.2 Les symbiotes secondaires	49
3.2.1 Généralités	49
3.2.2 Description des partenaires symbiotiques secondaires	51
<i>Arsenophonus</i>	51
<i>Cardinium</i>	52
<i>Hamiltonella</i>	53
<i>Rickettsia</i>	54
<i>Wolbachia</i>	55
4 Matériel et méthodes général	56
4.1 Élevage des insectes	57
4.2 Détermination du cytoype	58
<b>1 Description et dynamique spatio-temporelle des cytotypes</b>	<b>61</b>
1 Problématique	63
2 Historique des populations de <i>B. tabaci</i>	64
2.1 En France et dans le monde, deux invasions successives par deux espèces différentes	64
3 Description des cytotypes à l'échelle du bassin méditerranéen	67
3.1 Méthode	67
3.2 Résultats	67
4 Dynamique des cytotypes en France	70
4.1 Méthode	70
4.2 Répartition des haplogroupes	71
4.2.1 Répartition spatio-temporelle	71
4.2.2 Effets de la plante hôte sur la structuration	72
4.3 Description du compartiment symbiotique	74
5 Discussion	75
5.1 Répartition géographique des haplogroupes : compétition, remplacement, coexistence ?	75
5.2 Stabilité des associations : les cytotypes	78
5.2.1 Les associations de bactéries ne sont pas recombinaisonnées	78
5.2.2 Quels mécanismes expliquent cette stabilité ?	78
<b>2 Analyse du compartiment nucléaire des cytotypes</b>	<b>81</b>
1 Problématique	83
2 Matériels et méthodes	83
2.1 Allèles de résistance aux insecticides	84
2.2 Microsatellites	84
2.2.1 Génotypage	84
2.2.2 Analyse des données	85
3 Le cas des cytotypes ASL et Q1 en Afrique	86
3.1 Présentation de la situation	86
3.2 Échantillonnage	87
3.3 Résultats	88

	3.3.1	Analyse microsatellite	88
	3.3.2	Détection des allèles de résistance aux insecticides	89
	3.4	Conclusion sur ce premier exemple	90
4		Le cas des cytotypes Q1 et Q2 en Europe	92
	4.1	Présentation de la situation	92
	4.2	Croisements contrôlés	92
	4.2.1	Méthode	92
	4.2.2	Résultat	93
	4.3	Analyse du compartiment nucléaire	93
	4.3.1	Échantillonnage	93
	4.3.2	Allèles de résistance	94
	4.3.3	Analyse microsatellite	94
	4.4	Comparaison de populations mixtes et pures dans l'aire sympatrique	97
	4.5	Conclusion sur le modèle Q1-Q2 en Europe de l'ouest	100
5		Discussion	101
	5.1	Les bactéries n'induisent pas d'incompatibilités	101
	5.2	Taxonomie de l'espèce potentielle MED	101
	5.3	Limites des microsatellites, vers de nouveaux marqueurs	104
<b>3</b>		<b>Émergence d'un nouveau cytotype</b>	<b>107</b>
1		Problématique	109
2		Introduction	110
3		Identification de la souche de <i>Rickettsia</i> associée à Q1	112
	3.1	Méthode : MLST	112
	3.2	Résultats	113
4		Transfert expérimental	114
	4.1	Transfert de <i>Rickettsia</i> lors de l'accouplement	114
	4.2	Transfert de <i>Rickettsia</i> par l'intermédiaire de la plante hôte	116
	4.2.1	Méthode	116
		Mesure de la densité symbiotique par PCR quantitative	116
	4.2.2	Résultats	116
5		Discussion	118
	5.1	Le résultat d'un transfert horizontal	118
	5.2	Fréquences des transferts	120
	5.3	Stabilité de l'infection au niveau intra-individuel	121
	5.3.1	Compétition intra-hôte entre bactéries	122
	5.3.2	La capacité de <i>Rickettsia</i> à se transmettre à la génération suivante	123
	5.3.3	Interactions avec l'hôte	123
	5.4	Compétitivité du cytotype	124
<b>4</b>		<b>Développement et utilisation des marqueurs RADSeq pour le modèle <i>B. tabaci</i></b>	<b>127</b>
1		Introduction	129
	1.1	Objectifs	129
	1.2	Un outil pour la génomique des populations	129
	1.3	Concrètement, qu'est ce que le RADSeq ?	130
2		Expérience pilote, multiplexer de nombreux individus à faible coût	131
	2.1	Méthode	133
	2.1.1	Plan expérimental	133
	2.1.2	Analyse des données	134



2.2	Résultats du séquençage	135
2.2.1	Nombre de lectures par modèles biologiques	135
2.2.2	Estimation du nombre de locus chez <i>B. tabaci</i>	136
	<i>Clustering</i> intra-individuel	136
	<i>Clustering</i> inter-individus	137
2.3	Conclusion sur cette expérience pilote	139
3	Développement de l'outil pour <i>B. tabaci</i>	140
3.1	Objectifs de l'essai	140
3.2	Matériel et méthodes	140
3.2.1	Populations testées	140
3.2.2	Extraction d'ADN	141
3.2.3	Analyse des données	141
	Demultiplexage	141
	Séquences uniques	143
	Détection des séquences cytoplasmiques	143
	<i>Clustering</i> intra-individuel	143
	<i>Clustering</i> inter-individuel	144
	Export des données	144
	Analyses de génétique des populations	145
3.3	Résultats	145
3.3.1	Compte-rendu du séquençage	145
3.3.2	Démultiplexage et filtrage des lectures	146
3.3.3	<i>Clustering</i> : combien de locus ?	147
3.3.4	L'hétérozygotie est fonction du nombre de lectures	148
3.3.5	<i>Clustering</i> inter-individus, combien de locus partagés ?	149
3.3.6	Détection des symbiotes	150
3.3.7	Structure génétique	153
	Analyse en composante principale	153
	Nombre de groupes génétiques	155
	Estimation de la différenciation	157
3.3.8	Déséquilibre nucléo-cytoplasmique	157
4	Perspectives	160
	<b>Discussion générale</b>	<b>161</b>
1	Émergence d'un génotype très compétitif ?	162
1.1	Les deux cytotypes s'hybrident	162
1.2	Quid des bactéries ?	164
2	Vers une génomique nucléo-cytoplasmique ?	166
2.1	Détection de covariance nucléo-cytoplasmique	166
2.1.1	En populations naturelles	166
2.1.2	Un protocole expérimental pour tester la covariance entre génomes	168
3	Conclusion	168
	<b>Travaux publiés</b>	<b>169</b>
	<b>Annexes</b>	<b>219</b>
	<b>Références bibliographiques</b>	<b>243</b>

<b>Liste des figures</b>	<b>255</b>
<b>Liste des tableaux</b>	<b>257</b>
<b>Table des matières</b>	<b>263</b>