



École Pratique
des Hautes Études



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

ÉCOLE PRATIQUE DES HAUTES ÉTUDES

Sciences de la Vie et de la Terre

MÉMOIRE

présenté

par

Sylvie WAROT

pour l'obtention du Diplôme de l'École Pratique des Hautes Études

Caractérisation moléculaire et isolements reproducteurs chez des auxiliaires de lutte biologique

soutenu le 22 novembre 2018

devant le jury suivant :

<i>Dr Thierry DUPRESSOIR</i>	–	Président
<i>Dr Nicolas RIS</i>	–	Tuteur scientifique
<i>Pr Claudine MONTGELARD</i>	–	Tuteur pédagogique
<i>Dr Jean-Claude STREITO</i>	–	Rapporteur
<i>Dr Didier FORCIOLI</i>	–	Examineur

Mémoire préparé sous la direction de :

Nicolas RIS

Intitulé de la structure d'accueil : INRA - Institut Sophia-Agrobiotech, Sophia Antipolis

Directeur : Philippe Castagnone

et de

Claudine MONTGELARD

Intitulé de la structure d'accueil EPHE : CNRS – CEFE, Montpellier

Directeur : Richard JOFFRE

EPHE (Sciences de la Vie et de la Terre)

Groupe de Recherche et d'Enseignement Thématiques de l'EPHE : Biogéographie et Écologie des Vertébrés (BEV)

REMERCIEMENTS

Je tiens à remercier Claudine Montgelard d'avoir accepté de superviser mon travail en tant que tutrice pédagogique, pour ses précieux conseils pour la rédaction du mémoire et pour le suivi rigoureux de ce projet.

Je remercie également Nicolas Ris, en tant que tuteur scientifique et chef d'équipe pour m'avoir permis de réaliser ce projet en me libérant du temps au maximum malgré les contraintes fortes. Il m'a été d'une grande aide ainsi bien pour les analyses que pour la rédaction du mémoire. C'est un réel plaisir de travailler sous sa direction.

J'aimerais remercier Jean-Claude Streito en tant que rapporteur et Didier Forcioli en tant qu'examineur pour avoir accepté de juger ce travail ainsi que Thierry Dupressoir pour avoir accepté de présider le jury.

Un grand merci à Astrid Cruaud pour sa disponibilité pendant la formation en phylogénie multilocus, pour tous ses conseils et ses explications, ainsi qu'à Jean-Yves Rasplus, responsable du projet ANR TriPTIC principal financement de ces travaux.

Je remercie également Hélène Henri et Laurence Mouton pour les expérimentations complémentaires sur *Wolbachia* et pour leur disponibilité.

Je remercie les membres de mon équipe, Alexandre, Gaby, Isabelle, Marcel, Michela, Nadine et NicoB, pour m'avoir soutenu, pour m'avoir ménagé pendant la période de rédaction et particulièrement Anaël et Géraldine qui m'ont également fourni les trichogrammes. Je remercie nonante fois Hélène Wera et bien sûr Camille Duraj pour leur bonne humeur et pour m'avoir prêté main-forte pour les expériences de croisements.

Merci beaucoup à Roger Boll qui m'a apporté une écoute attentive tout au long de cette formation, pour nos discussions passionnantes et entre autres pour sa patience à m'apprendre les techniques de baby-foot.

Un grand merci à mes parents, mes frères et ma sœur pour leur soutien et leur encouragement et plus particulièrement à mon frère Thierry Fricaux pour ses conseils avisés pendant nos pauses café indispensables du matin.

Merci également à mes amis Sandra Berger et Bertrand Depincé qui m'ont apporté une aide précieuse en acceptant de relire et corriger ce mémoire.

Un merci ne suffit pas pour Yves mon mari, qui m'a débuggé tous mes problèmes informatiques (sauf python3 😊), corrigé ce mémoire, et surtout soutenu dans ce projet en gérant le quotidien. Et pour finir un gros bisous-câlin à mes loulous, Kélian et Lilwen.

1	INTRODUCTION GENERALE	6
1.1	Décrire le vivant	6
1.2	Définition de la Taxonomie intégrative	6
1.3	Présentation du modèle biologique	7
1.4	Taxonomie intégrative chez les trichogrammes : Etat d'avancement	9
1.5	Présentation des objectifs de l'étude	10
2	MATERIELS ET METHODES	11
2.1	Barcoding moléculaire Cytochrome oxydase I	11
2.1.1	Matériel biologique	11
2.1.2	Extraction d'ADN, PCR, et séquençage de la <i>COI</i>	12
2.1.3	Analyse	12
2.2	Caractérisation moléculaire multilocus	14
2.2.1	Matériel biologique	14
2.2.2	Choix des marqueurs	14
2.2.3	PCR et séquençage	15
2.2.4	Analyses phylogénétiques	16
2.2.4.1	Analyse gène par gène	18
2.2.4.1.1	Vérification des séquences	18
2.2.4.1.2	Modèles d'évolution et analyse phylogénétique	19
2.2.4.2	Concaténation des gènes et évaluation des partitions	21
2.2.5	Analyses complémentaires	23
2.2.5.1	Réseaux d'haplotypes	23
2.2.5.2	Cartographie intra-clusters	23
2.3	Détection des endosymbiotes bactériens du genre <i>Wolbachia</i>	23
2.3.1	Matériel biologique	23
2.3.2	Détection moléculaire de <i>Wolbachia</i>	24
2.3.3	Expérimentations complémentaires	24
2.3.3.1	PCR quantitative et caractérisation MLST	24
2.3.3.2	Traitements par antibiothérapie	25
2.4	Tests de croisement	25
2.4.1	Matériel biologique utilisé	25
2.4.2	Réalisation des croisements	26
3	CARACTERISATION DE LA BIODIVERSITE DES TRICHOGRAMMES VIA LE BARCODING <i>COI</i>	28
3.1	Introduction	28
3.2	Résultats	29
3.2.1	Diversité haplotype	29
3.2.2	Structuration moléculaire	31
3.2.3	Congruence avec les anciennes identifications	37
3.2.4	Expérimentations de croisements	37

3.3	Discussion	39
4	VERS UNE PHYLOGENIE MULTILOCUS DU GENRE TRICHOGRAMMA	42
4.1	Introduction	42
4.2	Résultats	43
4.2.1	Analyse séparée des marqueurs	43
4.2.2	Analyse multilocus	44
4.2.2.1	Comparaison des schémas de partitions :	44
4.2.2.2	Structuration :	49
4.2.2.2.1	Cas d'étude « cacoeciae-embryophagum »	49
4.2.2.2.2	Cas d'étude « brassicae/euproctidis »	52
4.2.2.2.3	Cas d'étude « cluster I-daumalae-evanescens »	54
4.2.2.2.4	Cas d'étude cluster V-misB :	54
4.2.2.3	Infections à <i>Wolbachia</i>	55
4.3	Discussion	57
4.3.1	Confrontation entre structuration moléculaire et classification moléculaire	57
4.3.2	Précisions concernant les complexes d'espèces majeurs	57
4.3.3	Infections à <i>Wolbachia</i>	58
5	CONCLUSION ET PERSPECTIVES	60
5.1	Bilan synthétique des résultats et apports par rapport à l'état de l'art	60
5.2	Complémentarité des approches moléculaires	61
5.3	Limite de la couverture taxonomique	62
5.4	La caractérisation morphologique, le « chaînon manquant »	63
5.5	Vers un couplage « Marqueurs – Traits »	64
6	BIBLIOGRAPHIE	65
7	TABLEAUX VOLUMINEUX	72
8	SUPPLÉMENTS :	85
8.1	Suppléments S1 : Résultats généraux liés à l'ensemble des croisements	85
8.2	Suppléments S2 : Présence de filles par catégorie de croisement	88
8.3	Suppléments S3 : Résultats des croisements interspécifiques entre complexes	89
8.4	Autres suppléments	90

1 INTRODUCTION GENERALE

1.1 Décrire le vivant

Face à la biodiversité du vivant, les biologistes se sont attachés depuis plusieurs siècles à décrire les organismes et à les classer les uns par rapport aux autres en fonction de similarités croissantes, une discipline à part entière nommée la taxonomie. A l'échelle la plus large, la classification actuelle distingue ainsi 6 Règnes : les Archées (procaryotes unicellulaires à histones), les Bactéries (procaryotes unicellulaires sans histones), les Protistes (eucaryotes unicellulaires), les Champignons, les Végétaux et les Animaux. Chaque règne est ensuite subdivisé en rangs taxonomiques successifs : Embranchement, Classe, Ordre, Famille, Genre et Espèce. A l'échelle la plus fine, un autre critère que la ressemblance morphologique intervient puisque la définition la plus connue de l'espèce est celle proposée par Ernst Mayr en 1942 (Mayr, 1999) qui la définit comme « des groupes de populations naturelles, effectivement ou potentiellement interfécondes, qui sont génétiquement isolées d'autres groupes similaires ». D'une façon générale, le travail de classification est loin d'être achevé. Ainsi, selon (Mora *et al.*, 2011), environ 1 500 000 espèces d'animaux sur terre ont été cataloguées alors que leur nombre total est estimé à environ 7 770 000. Dans ce contexte et face aux limites des approches morphologiques et du concept d'inter-fécondité, le système d'identification par la morphologie a peu à peu évolué du fait de sa complexité et du manque d'experts vers des identifications moléculaires plus accessibles et plus rapides. Dans un contexte de lutte biologique ¹, la description du vivant s'avère être une activité particulièrement importante afin d'identifier précisément les ravageurs et leurs ennemis naturels potentiellement utilisables dans ce domaine.

1.2 Définition de la Taxonomie intégrative

D'après de nombreux auteurs (voir par exemple Dayrat, 2005 ; Will *et al.*, 2005), la meilleure méthode pour étudier la biodiversité serait d'associer la taxonomie classique à des informations complémentaires incluant des données écologiques, géographiques, moléculaires, etc., une démarche portant le nom de « taxonomie intégrative ». En effet, si chacune des méthodes prises séparément peut aboutir à des erreurs d'identification et de délimitation, leur combinaison permet des identifications des espèces plus fiables et, au final, souvent plus rapides. Le « *DNA barcoding* » est une composante importante en taxonomie

¹ Lutte biologique : Méthode qui utilise des organismes vivants afin de réduire à un niveau acceptable le taux d'un organisme nuisible. Les principales stratégies de lutte biologique sont la lutte biologique dite « classique », par inoculation, par inondation, et par conservation (Eilenberg *et al.*, 2001).

intégrative puisqu'il permet une identification moléculaire des taxons par le séquençage d'un marqueur moléculaire suffisamment discriminant, comme une partie de la sous-unité 1 de la Cytochrome oxydase (*COI*) (Hebert *et al.*, 2003). L'intérêt et les limites du barcoding sont détaillés dans un prochain paragraphe (§ 3.1). La taxonomie intégrative a déjà prouvé son utilité dans le domaine de la lutte biologique où les connaissances précises hôtes – ravageurs sont particulièrement importantes. Des études intégratives sur des parasitoïdes ont notamment permis l'identification de 11 nouvelles espèces à l'intérieur d'un complexe d'espèces d'*Eupelmus* (Al Khatib *et al.*, 2014) et ont également montré l'existence de plusieurs espèces cryptiques avec une tendance à la spécificité des hôtes dans le cas du parasitoïde *Pnigalio soemius* présumé polyphage (Gebiola *et al.*, 2012).

De plus, afin de définir une frontière interspécifique précise, il paraît intéressant d'associer phylogénie et isolements reproducteurs notamment dans le cas d'espèces proches, de population en cours de spéciation ou d'espèces présentant une grande diversité haplotypique. Plusieurs études comparant la phylogénie et les isolements reproducteurs chez des parasitoïdes ont montré des variabilités de reproduction importantes non seulement dans des complexes d'espèces (Heraty *et al.*, 2007) mais également au niveau intraspécifique notamment chez les *Trichogramma* (Benvenuto *et al.*, 2012; Pinto *et al.*, 1991). Parallèlement, des barrières reproductives intraspécifiques peuvent également être observées en présence d'endosymbiotes bactériens manipulant la reproduction de son hôte (Stouthamer *et al.*, 1999).

1.3 Présentation du modèle biologique

Le travail présenté dans ce mémoire vise à contribuer à une démarche de taxonomie intégrative chez un taxon particulier, le genre *Trichogramma*. D'une façon générale, les trichogrammes sont des micro-hyménoptères (environ 0,7 mm) parasitoïdes oophages idiobiontes². Ce mode de développement se caractérise :

- d'une part, par un développement parasite à l'état pré-imaginal au détriment d'œufs d'autres espèces (l'embryon hôte étant rapidement tué) ;
- d'autre part, par une vie adulte « libre » (dispersion possible par la marche, le saut ou le vol) et une reproduction soit par haplo-diploïdie³, soit par thélytoquie⁴.

Parmi les parasitoïdes d'œufs, les trichogrammes ont reçu depuis plusieurs décennies une attention particulière avec plus de 2000 références trouvées dans Web of Science. Cette

² : Parasitoïde qui tue son hôte afin d'y pondre son œuf.

³ : Les femelles diploïdes sont issues d'ovocytes fécondés tandis que les mâles haploïdes d'ovocytes non fécondés.

⁴ : Les femelles se reproduisent de manière asexuée pour donner seulement des femelles.

particularité est expliquée par la pertinence de ce taxon à la fois pour des fins appliquées et pour la recherche fondamentale.

Au niveau appliqué, *Trichogramma* a été identifié comme un réservoir d'agents de lutte biologique car les espèces peuvent se développer au détriment d'œufs d'environ 200 espèces de Lépidoptères (Cônoli *et al.*, 2010; Hassan, 1993; Lenteren and Bueno, 2003; Parra and Zucchi, 2004; Pintureau, 2009; Smith, 1996) voire d'autres ordres puisque ces espèces sont connues pour s'attaquer à dix familles d'insectes. En France, les trichogrammes sont déjà commercialisés pour la lutte de différents ravageurs comme par exemple l'utilisation de *T. brassicae* contre la pyrale du maïs (*Ostrinia nubilalis*) (<http://biotop-solutions.fr/agriculture/professionnels/grandes-cultures.html>), de *T. evanescens* contre les chenilles de papillon de la famille des Noctuelles ou de *T. achaeae* contre la mineuse de la tomate (*Tuta absoluta*) en serres (<http://biotop-solutions.fr/agriculture/professionnels/maraichage-serres.html>). Le genre *Trichogramma* est également utilisé comme un modèle expérimental en écotoxicologie (Desneux *et al.*, 2006).

Dans le cadre de la recherche fondamentale, *Trichogramma* a été utilisé pour aborder des questions relatives à l'écologie comportementale, aux interactions hôte-symbiote (Pintureau *et al.*, 2000; Rousset *et al.*, 1992; Stouthamer, 1993; Stouthamer *et al.*, 1999), à la « thermal biology » c'est-à-dire la compréhension des effets de la température ambiante sur un animal (Hance *et al.*, 2007), et à la dynamique des populations.

Malgré ces acquis et ce statut de « taxon modèle », une méconnaissance demeure concernant la systématique de ce genre qui comprendrait environ 200 espèces (Nagarkatti and Nagaraja, 1977; Pintureau, 2008) dont 40 en Europe. Cette tâche est en effet compliquée à cause de problèmes techniques liés à la couverture de l'échantillonnage, au nombre limité et à la subjectivité éventuelle des caractères morphologiques (organes génitaux masculins et antennes) utilisés pour la « systématique traditionnelle ». D'autres difficultés peuvent résulter de la rapidité des processus de spéciation observés chez les hyménoptères (Cruaud *et al.*, 2012; Danforth *et al.*, 2006; Ronquist and Liljeblad, 2001; Wilson and Holldobler, 2005). Au niveau intraspécifique, peu d'information sur la structuration génétique existe même pour des espèces largement distribuées et fréquemment commercialisées (par exemple *T. cacoeciae* ou *T. evanescens* dans la partie occidentale de l'écozone paléarctique). Cette connaissance insuffisante de la diversité inter et intraspécifique limite probablement simultanément la pertinence et la fiabilité des travaux appliqués pour le développement de nouveaux agents de

lutte biologique ainsi que des recherches plus fondamentales (données pour choisir les meilleurs « modèles » répondant aux questions de recherche).

1.4 Taxonomie intégrative chez les trichogrammes : Etat d'avancement

Au cours du temps, de nombreuses techniques d'identification ont été mises au point afin de caractériser les trichogrammes. Historiquement, ce sont évidemment les caractères morphologiques (genitalia mâles et antennes Figure 1) qui ont été utilisés par un petit nombre de spécialistes dans le domaine (Pintureau, 2008). Par la suite, de nouvelles techniques sont apparues pour objectiver, faciliter ou « démocratiser » les identifications comme des méthodes biochimiques (Kazmer, 1991; Pinto et al., 1993; Pintureau and Keita, 1989), l'utilisation des PCR RAPD Random Amplified Polymorphic DNA (Vanlerberghe-Masutti, 1994), le séquençage du gène Internal Transcribed Spacer 1, (Orrego and Agudelo-Silva, 1993) puis du gène Internal Transcribed Spacer 2 (Stouthamer et al., 1999) et plus récemment, l'utilisation du marqueur Cytochrome oxydase 1 (Ercan and Öztem, 2013; del Pino et al., 2013). Parallèlement, des études d'isolements reproducteurs ont également déjà été menées afin de mieux comprendre certains complexes d'espèces (Pinto *et al.*, 1991; Stouthamer *et al.*, 2000). Si les bases d'une taxonomie intégrative existent donc, force est toutefois de constater qu'il n'y a pas de réelles synthèses de ces travaux.

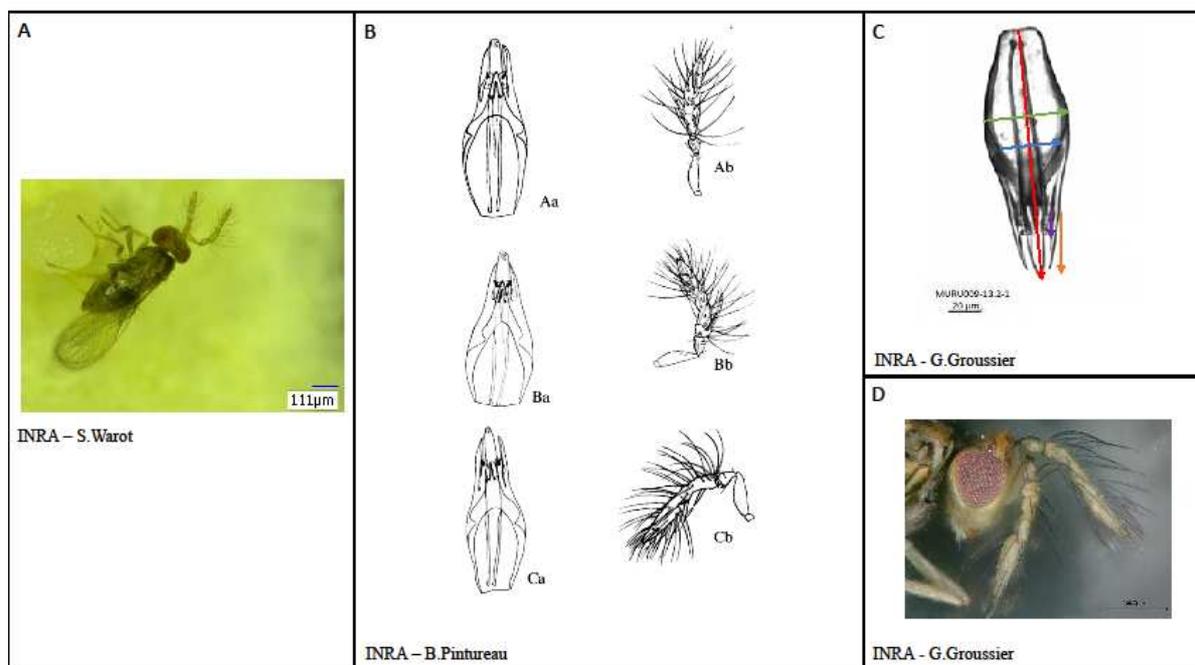


Figure 1 : **A** : Trichogramme mâle ; **B** : Genitalia (a) et antennes (b) mâles de trois espèces de *Trichogramma* du groupe *evanescens*, *T. brassicae* (Aa, Ab), *T. euproctidis* (Ba, Bb) et *T. evanescens* (Ca, Cb) ; **C** : Genitalia d'un mâle de trichogramme ; **D** : Antennes d'un mâle de trichogramme.

1.5 Présentation des objectifs de l'étude

C'est dans ce contexte que mon équipe a lancé l'initiative de créer un Centre de Ressources Biologiques dédié aux trichogrammes, le CRB EP-Coll (Marchand *et al.*, 2017 - <https://doi.org/10.15454/AY4LMT>), afin non seulement de pouvoir fournir des souches fiables aux chercheurs intéressés mais également de mieux documenter la diversité de ce taxon notamment dans la zone ouest-paléarctique. Dans le cadre de ma formation EPHE, je me suis ainsi plus particulièrement consacrée à l'acquisition et à l'analyse de données de « DNA barcoding ». Ces résultats ont ensuite été complétés par des données sur d'autres marqueurs moléculaires pour préciser les relations évolutives entre les taxons (Cruaud *et al.*, 2013). Enfin, des expériences de croisements ont été réalisées pour mieux délimiter les périmètres de compatibilités reproductives, les résultats de cette approche étant eux-mêmes éclairés par l'étude des statuts d'infection à *Wolbachia*

2 MATERIELS ET METHODES

Cette section du manuscrit reprend l'ensemble des matériels et méthodes, les résultats étant présentés dans les sections suivantes.

2.1 Barcoding moléculaire Cytochrome oxydase I

Ce travail a été réalisé sur la période 2013-2018. L'ensemble des activités expérimentales a été réalisée par moi-même. Dans le cadre de la formalisation du CRB EP-Coll en vue de sa certification ISO9001, les activités de caractérisation moléculaire *COI* des trichogrammes sont formalisées par le processus PS-RDLB-003_R1 et le mode opératoire MO-RDLB-009. L'ensemble des documents est centralisé en interne sur un espace partagé du CRB EP-Coll.

2.1.1 Matériel biologique

Les souches utilisées dans ces travaux proviennent soit de collections « historiques », soit d'un échantillonnage plus récent (2013-2017), les collections historiques étaient auparavant conservées dans les laboratoires "Biologie Fonctionnelle, Insectes et Interactions" (Villeurbanne, France), dans l'"Unité Expérimentale de Lutte Biologique" (Valbonne, France) ou dans l'"Unité de Recherches Intégrées en Horticulture" (Sophia-Antipolis, France). La plupart des échantillonnages récents ont été organisés dans le cadre de deux projets de recherche, d'un côté, le projet intersectoriel "COLBICS" (FP7-PEOPLE-IAPP 2013-2016 – Resp.: Malausa T.) impliquant la société privée BIOLINE AGROSCIENCES et l'organisme public de recherche INRA et, d'autre part, le projet de recherche publique "TRIPTIC" (ANR 2016-2019 – Resp.: Rasplus JY.). Des détails sur le plan d'échantillonnage, dont les objectifs étaient (i) de tester des méthodes de piégeage, (ii) collecter des trichogrammes dans différents habitats naturels ou agricoles et ainsi documenter la diversité naturelle et la distribution des trichogrammes et (iii) alimenter le CRB Ep-Coll, sont disponibles dans un autre article (Muru *et al.*, in prep). Certaines souches ont finalement été envoyées par des collègues distants suite à des recherches bibliographiques auprès de chercheurs publiant sur cette thématique ou grâce à des interactions avec des personnes en collaboration avec l'équipe. Au final, plus de 3100 individus provenant d'environ 600 souches de *Trichogramma* ont été utilisés pour la caractérisation moléculaire avec le gène *COI*, une souche étant définie ici comme la combinaison d'une date, d'un emplacement et d'un micro-habitat précis. Les souches de référence listées dans le Tableau 1 (voir §7) donnent un aperçu de l'aire géographique couverte et de la diversité des espèces.

2.1.2 Extraction d'ADN, PCR, et séquençage de la *COI*

L'ADN génomique a été extrait à partir de tissus frais (individus récemment tués) ou congelés (individus conservés dans de l'éthanol à -20°C) en utilisant le kit prepGEM® Insect (ZYGEM, PIN0500). Les individus non broyés ont été placés individuellement dans 15 µl de mélange réactionnel et incubés pendant 3 heures à 75°C puis pendant 5 minutes à 95°C. Les extraits ont ensuite été conservés à -20°C.

Un fragment (environ 700 pb) de la sous-unité 1 de la Cytochrome oxydase (*COI*) a été amplifié en utilisant les amorces LCO1490 et HCO2198 (Folmer *et al.*, 1994). La Cytochrome oxydase 1 est un gène universel, assez variable pour différencier les espèces entre elles tout en permettant de définir des amorces relativement « universelles » et donc d'amplifier des séquences comparables chez des taxons même relativement éloignés (familles voire des ordres).

Plus précisément, 1µl d'ADN a été utilisé pour la réaction de PCR (effectuée dans un volume total de 25µl). La PCR a été effectuée avec le QIAGEN Multiplex PCR Master Mix (n° de cat./ID: 206145) avec une concentration finale de 3mM de MgCl₂ et 0,125µl de chacune des deux solutions d'amorce (100µM). Les conditions de PCR étaient les suivantes : 95°C pendant 15 minutes ; 40 cycles à 95°C pendant 30 s, 50°C pendant 90 s, 72°C pendant 1 min ; et une étape d'élongation finale à 60°C pendant 30 minutes. La taille des produits de PCR a été analysée en utilisant un kit Qiaxcel DNA Fast analysis (QIAGEN S.A.S) sur le système Qiaxcel Advanced (QIAGEN S.A.S) (instructions d'utilisation expliquées dans un document interne codifié INS-RDLB-003). Les produits de PCR ont été séquencés, soit avec les deux amorces, soit avec l'amorce HCO2198 uniquement en utilisant la méthode de Sanger par les sociétés BECKMAN COULTER puis GENEWIZ (Essex, GB) (mode opératoire expliqué dans le document interne codifié MO-RDLB-021).

2.1.3 Analyse

Toutes les séquences ont été éditées, nettoyées puis alignées avec ClustalW en utilisant MEGA version 7.0.01 (Kumar *et al.*, 2016). L'analyse finale présentée ici a porté sur une portion courte de 271pb. Ce choix délibéré a été motivé par la haute qualité et la fiabilité de cette partie, même avec un séquençage unidirectionnel (HCO uniquement). Il peut donc être utilisé avec confiance non seulement par la communauté scientifique mais aussi par d'autres utilisateurs finaux (entreprises privées notamment). *A posteriori*, cette région s'est avérée suffisamment informative pour la plupart des taxons et/ou des objectifs (voir 3.2_Résultats). Les limites de ce barcode court peuvent être localisées en utilisant les deux séquences dégénérées suivantes : 5'-TNATNNTNGGNANNCCNGANATNGC-3' et 5'-

CNATTNTNAATATAAAAATTTANAA-3' (par exemple en commençant aux positions 207 et 453 dans l'accession GenBank MG932137 - Tableau 1 - voir §7).

Sur la base de l'ensemble de données obtenues, les haplotypes COI ont été identifiés et triés en utilisant DnaSP version 5.10.01 (Librado and Rozas, 2009). Les séquences obtenues ont été mises à disposition dans Genbank (Tableau 1 - voir §7)

Les haplotypes ont été catégorisés comme suit (voir le Tableau 1 - voir §7) :

- Catégorie A : haplotypes observés seulement chez un individu ;
- Catégorie B : haplotypes observés plus d'une fois mais seulement dans un échantillon de population ;
- Catégorie C : haplotypes observés plus d'une fois et présents dans 2 à 10 échantillons de population ;
- Catégorie D : haplotypes observés dans plus de 10 échantillons ;
- Catégorie E : haplotypes récupérés sur la base de données GenBank, sans information de collecte facilement récupérable.

Chaque haplotype COI a été traduit en une séquence d'acides aminés en utilisant MEGA7 et les haplotypes présentant des variations non synonymes sur leur séquence entière ont été respectivement marqués comme "varprot" sur le Tableau 1 (voir §7).

La reconstruction d'une phylogénie à partir de l'ensemble des données s'est avérée difficile à interpréter en raison de artefacts (séquences isolées et pseudogènes) avec parfois des petites distances génétiques entre certains individus. Nous avons donc choisi la reconstruction d'un réseau sous forme de cycles à partir de toutes les données afin de mieux visualiser les délimitations entre taxons et entre individus. Bien que la méthode vise habituellement à analyser des séquences intraspécifiques, elle a également été utilisée pour étudier des ensembles de données interspécifiques (Fulton *et al.*, 2012; Yamakawa and Imai, 2013). Le réseau d'haplotypes COI a donc été établi sur l'ensemble des haplotypes recensés dans le Tableau 1 (voir §7) à l'aide du logiciel PopARTs version 1.7 avec la méthode Median-Joining⁵ (paramètre par défaut : epsilon = 0). En parallèle, un arbre phylogénétique a été construit avec la méthode du Neighbor-Joining⁶ dérivé d'une matrice de distance Tamura-Nei⁷ choisi à partir de l'outil « Find best model » (MEGA7) (500 réplifications de Bootstrap⁸), en

⁵ Median-Joining : Méthode de construction de réseau non enraciné qui combine le réseau Minimum-Spanning (rapide mais restrictif avec un seul nœud possible) et le réseau quasi-median (complexe et lourd), au sein d'un même réseau.

⁶ Neighbor-Joining (1987): Méthode rapide de reconstruction d'arbre basée sur le principe du minimum d'évolution en minimisant à chaque étape la longueur totale des branches entre taxons voisins.

⁷ Distance Tamura-Nei (1993) : Modèle évolutif prenant en compte de 6 paramètres pour quantifier les distances évolutives. Elle distingue les taux de transitions entre purines et entre pyrimidines, les taux de transversions, les différences de taux de substitution entre les nucléotides et de l'inégalité des fréquences des nucléotides.

⁸ Bootstrap (1985) : Test de fiabilité des branches d'un arbre qui utilise la technique du ré-échantillonnage aléatoire de nucléotides avec reconstruction de N arbres fictifs confrontés à l'arbre d'origine. Le résultat de bootstrap est exprimé en pourcentage sur chaque branche.

utilisant seulement les haplotypes les plus fréquents (catégories B-D) et ceux de GenBank (catégorie E). Des haplotypes de la catégorie A (très rares et, par conséquent possiblement des artefacts) ainsi que ceux présentant des variations non synonymes (pseudogènes et / ou artefacts putatifs) ont été intentionnellement exclus. Les out-groups permettant d'enraciner l'arbre sont des séquences de *Nasonia*, un hyménoptère parasitoïde de la famille des Pteromalidae.

2.2 Caractérisation moléculaire multilocus

Ces travaux ont été réalisés sur la période 2014 - 2018. L'ensemble des activités expérimentales a été réalisé par moi-même. L'analyse des données a fait l'objet d'une formation auprès d'Astrid Cruaud (CBGP) puis de discussions avec elle ainsi qu'avec Claudine Montgelard (CEFE).

2.2.1 Matériel biologique

A partir des travaux menés sur le marqueur COI (cf. § 3_Barcoding COI), 317 individus provenant de 173 souches de *Trichogramma* ont été utilisés pour une caractérisation multilocus. L'ensemble des souches, leurs origines et les informations associées sont listés dans le Tableau 2 (voir §7). Le choix des souches/individus a été motivé par :

- (i) la volonté de réutiliser autant que possible les souches voire les individus déjà utilisés pour la caractérisation COI ;
- (ii) le souci de représenter l'ensemble des clusters moléculaires identifiés ;
- (iii) la volonté de préciser autant que possible les complexes préalablement identifiés sur COI (« brassicae-euproctidis », « daumalae-evanescens », « cacoeciae-embryophagum »).

2.2.2 Choix des marqueurs

Six marqueurs moléculaires ont été retenus pour cette étude :

- deux gènes mitochondriaux qui codent pour des sous-unités d'un complexe de protéines transmembranaires de la chaîne respiratoire. On les retrouve en nombreuses copies dans le génome. Ceci facilite les procédés d'amplification des gènes mitochondriaux au sein d'un individu. Ces gènes mitochondriaux présentent une faible diversité entre individus d'une même espèce car les mitochondries sont transmises verticalement de mère en fille créant ainsi un goulot d'étranglement.

La Cytochrome oxydase 1 *COI* est décrite précédemment (cf. § 3 Barcoding COI).

Le Cytochrome B (*CytB*) qui code pour une sous-unité d'un complexe de protéines transmembranaires de la chaîne respiratoire (Jermiin and Crozier, 1994) permet de déterminer les relations au niveau des clades terminaux et donc de décrire l'évolution récente.

- deux gènes nucléaires codants : (i) *Wingless* codant pour une protéine jouant un rôle dans la formation de motifs des ailes chez les insectes (Brower and DeSalle, 1998) permet de résoudre des relations « intermédiaires » (niveau taxonomique similaire) ; (ii) la copie F2 de *EF1alpha*, un gène qui code pour une protéine impliquée pendant la phase de traduction (Brady *et al.*, 2006; Cruaud *et al.*, 2013; Danforth *et al.*, 2004) permet de récupérer des informations au niveau des nœuds profonds des ancêtres communs (au niveau des familles).

- deux marqueurs EPIC (« exon-primed, intron-crossing ») dont les amorces s'hybrident dans des parties exoniques et permettent d'amplifier une partie intronique du gène : la 30S Ribosomal Protein S4 (*RpS4*) et la 60S Ribosomal Protein L27a, (*Rpl27a*) (Lohse *et al.*, 2011). Ces gènes devraient permettre de résoudre des relations « intermédiaires » et ont été utilisés notamment pour construire une phylogénie des hyménoptères associés aux galles de chênes et de figes (Cruaud *et al.*, 2013). Ces six marqueurs ont déjà été utilisés conjointement ou en partie avec succès dans des analyses phylogénétiques sur (i) des espèces du genre *Eupelmus* (Hymenoptera : Eupelmidae) (Al Khatib *et al.*, 2014; Al khatib *et al.*, 2016) et (ii) des Agaonidae (Hymenoptera : Chalcidoidea), pollinisateurs des Ficus (Cruaud *et al.*, 2010).

2.2.3 PCR et séquençage

Au final, 317 extraits d'ADN ont été utilisés pour réaliser les amplifications par PCR des 6 marqueurs. Pour chacun des marqueurs, 1 µl d'ADN a été utilisé pour un volume de réaction de PCR de 25 µl. Les réactions de polymérisation en chaîne (PCR) ont été réalisées avec le « PCR Multiplex Master Mix » de QIAGEN (Cat No. ID: 206145) avec une concentration finale de 3 mM MgCl₂ et 0,5 µM d'amorces listées dans le Tableau 3 en fonction du locus à amplifier. Les conditions de PCR étaient les suivantes : 95 ° C pendant 15 minutes, 40 cycles à 95 ° C pendant 30 secondes, T_m°C (cf. Tableau3) pendant 90 secondes, 72 ° C pendant 1 minute, avec une élongation finale à 60°C pendant 30 minutes. La taille des produits de PCR a été estimée avec le kit «QIAxcel DNA Fast Analysis» (QIAGEN S.A.S) sur le système Qiaxcel Advanced. Les produits de PCR ont été séquencés dans les deux sens à l'aide des amorces Forward et Reverse utilisées pour chacun des loci, cette activité étant assurée dans un premier temps par BECKMAN-COULTER (Essex, GB) puis par GENEWIZ (Essex, GB).

Tableau 3 : Descriptif des conditions de PCR pour les marqueurs moléculaires utilisés pour la phylogénie multilocus :

Locus	Nom des amorces	F/R	Séquence des amorces	Taille du fragment	Température d'hybridation	Source
COI	LCO1490	F	GGTCAACAAATCATAAAGATATTGG	710bp	48-50 C°	Folmer <i>et al.</i> (1994)
	HCO2198	R	TAAACTTCAGGGTGACCAAAAAATCA			
Cyt b	CB1	F	TATGTACTACCATGACGACAAATATC	484bp	46 C°	Jermin & Crozier (1994)
	CB2	R	ATTACACCTCCTAATTTATTAGGAAT			
Wingless	LepWg1a	F	GARTGYAARTGYCAYGGYATGTCTGG	664bp	53 C°	Brower&DeSalle (1998)
	LepWg2	R	ACTNCGCRCACCARTGGAATGTRCA			Danforth <i>et al.</i> (2004)
EF-1α	F2-557F	F	GAACGTGAACGTGGTTATYACSAT	561bp	55 C°	Cruaud <i>et al.</i> (2013)
	F2-1118R	R	TTACCTGAAGGGGAAGACGRAG			Brady <i>et al.</i> (2006)
RpS4	11F	F	BAARGCATGGATGTTTGACA	800bp	55 C°	Lohse <i>et al.</i> (2011)
	11R	R	GGTCWGGRTADCGRATRGT			
RpL27a	28Fb	F	CAAYTTYGACAARTACCATCCWG	600bp	52 C°	Lohse <i>et al.</i> (2011)
	28R	R	CCYTTKCCYARRAGTTTGTA			

2.2.4 Analyses phylogénétiques

Le schéma des différentes étapes réalisées pour les analyses phylogénétiques multilocus est représenté dans la Figure 2 avec les logiciels utilisés et les numéros des modes opératoires associés. La construction d'une phylogénie multilocus nécessite préalablement de travailler sur les alignements gène par gène avant leur concaténation afin de :

- nettoyer les séquences des primers voire les parties 5' et 3' peu lisibles ;
- détecter et supprimer les NUMTS (nuclear mitochondrial DNA segment), les paralogues, les séquences de mauvaise qualité, les séquences présentant des codons–stop ;
- vérifier la cohérence des individus entre les différents gènes et éliminer les éventuelles erreurs manifestes (erreur de numérotation, contaminations).

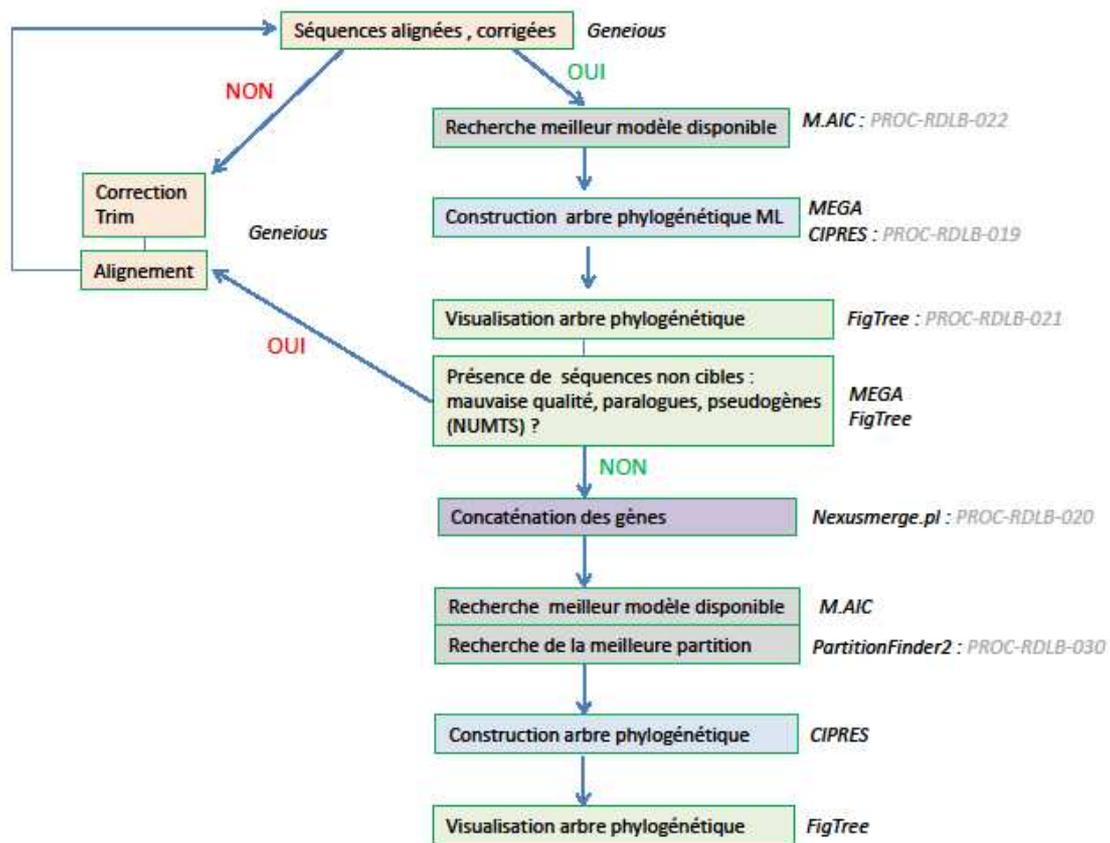


Figure 2 : Schéma des différentes étapes réalisées pour les analyses de phylogénie multilocus.

A partir de chacun des alignements gène par gène, la recherche du « meilleur » modèle l'évolution (MrAIC) sur la base de la parcimonie, c'est-à-dire en maximisant l'adéquation aux données (vraisemblance) tout en limitant le nombre de paramètres. En effet, il existe de nombreux modèles d'évolution plus ou moins complexes, du plus simple « Jukes et Cantor » 1969, impliquant un paramètre au plus complet le modèle GTR (General Time Reversible) 1994, impliquant neuf paramètres autorisant des taux différents pour les six types de substitutions (les transitions entre AG et CT, les transversions entre AC, AT, GC et GT), ce dernier étant le modèle le plus utilisé pour les séquences nucléotidiques.

A partir du meilleur modèle proposé par MrAIC, nous avons privilégié une reconstruction probabiliste par Maximum de Vraisemblance (ML) qui est décrite comme étant la meilleure méthode pour trouver un arbre le plus proche de la réalité, son désavantage étant par contre le temps nécessaire à l'analyse. Comme alternative, la méthode RaxML (Randomized Axelerated Maximum Likelihood) optimisée uniquement avec le modèle d'évolution GTR permet de construire rapidement un arbre ML. Comme le montre la Figure 2, nous avons privilégié les analyses RaxML sous CIPRES et pour les cas où le meilleur modèle n'était pas

GTR, l'arbre ML a été construit avec le logiciel MEGA. Tous les arbres obtenus ont été visualisés avec le logiciel FigTree qui permet d'annoter facilement un arbre phylogénétique.

Une fois les arbres analysés et comparés entre eux, nous sommes passés à l'étape de concaténation des séquences des différents gènes pour réaliser des « super-alignements » (script nexusmerge.pl, Figure 2). Ces supers alignements ont fait l'objet d'une nouvelle analyse MrAIC et d'une recherche de la meilleure partition. En effet, la meilleure façon de prendre en compte l'hétérogénéité du processus d'évolution entre les gènes (par exemple, gènes mitochondriaux *versus* gènes nucléaires) et entre les différents sites des gènes (par exemple : portions codantes *versus* non codantes ; positions synonymes *versus* non synonymes) est de partitionner les données d'alignement en groupant ensemble les données qui évoluent de la même façon. Deux types de schémas de partition ont été considérés :

- Un schéma de partition « à dire d'expert » proposé par A. Cruaud sur la base de ses études sur d'autres hyménoptères ;
- un schéma de partition sélectionné par le logiciel PartitionFinder2 (Figure 2).

Les schémas de partition sélectionnés sont alors implémentés à l'analyse RaxML sous CIPRES (Figure 2).

2.2.4.1 Analyse gène par gène

2.2.4.1.1 Vérification des séquences

Toutes les séquences ont été éditées, nettoyées, assemblées puis alignées avec ClustalW en utilisant soit MEGA7 (version 7.0.21 - (Kumar *et al.*, 2016)), soit Geneious (version R10.2.3 – (Kearse *et al.*, 2012)). Les séquences codantes ont été traduites afin d'identifier d'éventuelles variations dans les séquences d'acides aminés correspondantes et notamment identifier la présence de codons stop. Pour chaque marqueur, des arbres RaxML sont inférés et les topologies de chacun de ces arbres sont également comparées « visuellement » les unes aux autres afin notamment de détecter des incohérences « ponctuelles » (conflits entre marqueurs au niveau d'un individu attribuables notamment à des erreurs de traçabilité ou des contaminations), des « incohérences généralisées » (à l'échelle d'une souche : amplification de pseudogènes voire de paralogues) ou des phénomènes plus complexes (hybridation interspécifique).

2.2.4.1.2 Modèles d'évolution et analyse phylogénétique

Le meilleur modèle d'évolution pour chaque partition et pour les gènes concaténés a été identifié en utilisant le critère d'information d'Akaike AIC implémenté dans MrAIC.pl version 1.4.3 (Nylander, 2004). Les analyses phylogénétiques de chaque jeu de données sont réalisées en fonction des résultats du test AIC donnés dans le Tableau 4 :

- sous MEGA7 pour les reconstructions ML de *CytB* et *F2-Ef1alpha*, en utilisant respectivement les modèles de calcul de distance de Hasegawa Kishino Yano (HKY) et Jukes et Cantor (JC) avec une prise en compte d'une variation du taux d'évolution Gamma entre sites ;
- sur le portail CIPRES Science V 3.3 (Miller *et al.*, 2012) en utilisant l'outil « RAxML-HPC2 on XSEDE » (v.8.2.10) (Stamatakis, 2014) pour la reconstruction des autres gènes et des gènes concaténés. La reconstruction des arbres phylogénétiques a été réalisée en utilisant le programme RAxML (Randomized Axelerated Maximum Likelihood) basé sur le Maximum de Vraisemblance. Les reconstructions RaxML sont réalisées sous un modèle « Generalised Time Reversible » CAT (GTRCAT) avec 100 répétitions de bootstrap. Le GTRCAT est utilisé pour accélérer les calculs de données importantes, il s'agit d'une approximation de calcul largement utilisé pour des données sous le modèle GTR GAMMA. Il demande généralement 4 fois moins de temps et de mémoire.

L'édition des arbres a été réalisée grâce au logiciel Figtree (version 1.4.3 <http://tree.bio.ed.ac.uk/software/figtree/>).

Tableau 4 : Descriptif des différents alignements utilisés pour les phylogénies multilocus.

Le tableau donne des informations sur le résultat de la recherche du meilleur modèle AIC/AICc sur chacun des gènes et sur les différentes concaténations. Les 3 dernières colonnes donnent le nombre de sites utilisés dans les analyses, le nombre de sites variables et le nombre de sites informatifs de parcimonie avec en parenthèse le pourcentage des sites sur le nombre de sites total.

Locus	Test M.AIC	Best model	Méthodes et paramètres	Nbre positions	Nbre de sites	Nbre de sites
		M.AIC / AICc	Figures associées	analysées	variables	informatifs
COI	Minimum AIC model	GTRIG	RaxML GTR CAT (CIPRES)	600	247 (41%)	212 (35%)
	Minimum AICc model	JC69	Figure S6			
CytB	Minimum AIC model	HKYIG	ML HKY IG (MEGA7)	465	201 (43%)	179 (38%)
	Minimum AICc model	HKYIG	Figure S7			
F2_Ef1alpha	Minimum AIC model	SYMG	ML JC69 (G=5) (MEGA7)	540	121 (22%)	108 (20%)
	Minimum AICc model	JC69G	Figure S4			
RpL27a	Minimum AIC model	GTRG	RaxML GTR CAT (CIPRES)	427	160 (37%)	142 (33%)
	Minimum AICc model	GTRG	Figure S13			
RpS4	Minimum AIC model	GTRG	RaxML GTR CAT (CIPRES)	943	461 (49%)	356 (38%)
	Minimum AICc model	GTRG	Figure S8			
Wingless	Minimum AIC model	GTRIG	RaxML GTR CAT (CIPRES)	434	154 (35%)	139 (32%)
	Minimum AICc model	GTRIG	Figure S5			
Concaténation	Minimum AIC model	GTRIG	RaxML GTR CAT (CIPRES)	1065	448 (42%)	391 (37%)
Mitochondriale	Minimum AICc model	HKYIG	Figure S9 et Figure S10			
Concaténation	Minimum AIC model	GTRG	RaxML GTR CAT (CIPRES)	2344	896 (38%)	745 (32%)
Nucléaire	Minimum AICc model	GTRG	Figure S11 et Figure S12			
Concaténation	Minimum AIC model	GTRG	RaxML GTR CAT (CIPRES)	3409	1344 (39%)	1136 (33%)
6 gènes	Minimum AICc model	GTRIG	Figure 7 et Figure S14			

2.2.4.2 Concaténation des gènes et évaluation des partitions

Les séquences conservées à partir de cette première analyse ont ensuite été concaténées en ligne de commande en utilisant le script `nexusmerge.pl` mis au point par A. Cruaud. Plusieurs schémas de partition différents ont été envisagés :

- à partir des deux seuls gènes mitochondriaux concaténés (*COI* et *CytB*) avec 2 partitions associées à chacun des gènes (*COI*, *CytB*) et avec 6 partitions (une partition pour chaque position dans le codon pour chaque gène) proposées par l'analyse PartitionFinder2 (version 2.1.1) (Lanfear *et al.*, 2017) ;
- à partir des quatre gènes nucléaires concaténés (*EF1alpha-F2*, *Rpl27a*, *RpS4*, *Wingless*) avec 4 partitions associées à chacun des gènes et avec 7 partitions proposées par l'analyse PartitionFinder2 (*EF1alpha-F2* Position1, *EF1alpha-P2*, *EF1alpha-P3* + *Wingless P1*, *Rpl27a*, *RpS4*, *Wingless P2*, *Wingless P3*) ;
- les 6 gènes concaténés (*COI*, *CytB*, *EF1alpha-F2*, *Rpl27a*, *RpS4*, *Wingless*) avec 12 partitions (cf. Tableau 5) proposées par l'analyse PartitionFinder2 ;
- sur les conseils d'A. Cruaud, les 6 gènes concaténés avec 5 partitions choisies comme suit : une pour les ADN mitochondriaux (*COI*, *CytB*), puis une pour chacun des ADN nucléaires (*EF1alpha-F2*, *Rpl27a*, *RpS4*, *Wingless*).

Quelle que soit la partition, les arbres phylogénétiques ont été inférés en utilisant le programme RAxML basé sur le Maximum de Vraisemblance. Les reconstructions RaxML sont réalisées sous un modèle « Generalised Time Reversible » CAT (GTRCAT) avec 100 répétitions de bootstrap.

La comparaison des premiers schémas de partition (« gènes mitochondriaux » seuls et « gènes nucléaires » seuls) avec les topologies de chacun des gènes (cf. § 2.2.4.1 et 2.2.4.2) s'est avérée nécessaire afin de « nettoyer » encore un peu le jeu de séquences d'éventuelles contaminations et d'analyser les sous-structurations visibles uniquement sur les gènes mitochondriaux.

Les autres partitions sur les 6 gènes concaténés nous ont permis d'analyser plus précisément l'histoire évolutive au sein du genre *Trichogramma*. Mais des études ayant montré le risque du sur-paramétrage et recommandant de fusionner certains blocs de données afin d'optimiser la reconstruction phylogénétique (Li *et al.*, 2008), nous avons privilégié dans ce rapport cette dernière analyse à 5 partitions (cf. § 4.2).

Tableau 5 : Résultats de la recherche de la meilleure partition sur la concaténation des 6 gènes (*COI*, *CytB*, *Ef1alpha-F2*, *RpL27a*, *RpS4*, *Wingless*) par PartitionFinder2 utilisés pour la construction de la phylogénie Figure S14.

Partition	Gène et position	Position alignement		Espacement entre sites
		Début	Fin	
1	COI position1	1	600	3
2	COI position 2	2	600	3
3	COI position 3	3	600	3
4	CytB position 1	601	1065	3
	RpL27a	1606	2032	1
5	CytB position 2	602	1065	3
6	CytB position 3	603	1065	3
7	F2-EF1alpha position 1	1066	1605	3
	Wingless position 3	2978	3409	3
8	F2-EF1alpha position 2	1067	1605	3
9	F2-EF1alpha position 3	1068	1605	3
10	RpS4	2033	2975	1
11	Wingless position 1	2976	3409	3
12	Wingless position 2	2977	3409	3

2.2.5 Analyses complémentaires

2.2.5.1 Réseaux d'haplotypes

Des réseaux d'haplotypes ont été réalisés à partir des gènes concaténés sur chacun des 3 complexes d'espèces étudiés (« brassicae-euproctidis », « cacoeciae-embryophagum », « daumalae-evanescens ») à l'aide de PopARTs version 1.7 (Bandelt *et al.*, 1999) avec la méthode Median Joining (paramètre par défaut : epsilon = 0). Pour chaque complexe, le réseau est établi à partir de plus d'une vingtaine d'individus pour lesquels les 6 gènes séquencés étaient disponibles.

2.2.5.2 Cartographie intra-clusters

Dans le cas du complexe « brassicae-euproctidis », la sous-structuration moléculaire a été confrontée aux informations concernant l'origine géographique de souches à l'aide du logiciel QGIS version 2.16.2.

2.3 Détection des endosymbiotes bactériens du genre *Wolbachia*

Ces travaux ont été réalisés sur la période 2014-2018. Les activités de détection moléculaire *Wolbachia* (2.3.2 Détection moléculaire de *Wolbachia*) et une partie des expérimentations complémentaires (2.3.3. Expérimentations complémentaires) ont été réalisées par moi-même. La détection de *Wolbachia* par QPCR, leur caractérisation par MLST (2.3.3.1 Expérimentations complémentaires) et la détection d'autres symbiotes (données non présentées) ont été réalisées en 2017-2018 par Laurence Mouton, Hélène Henri et ses collaborateurs au sein du Laboratoire de Biométrie et Biologie Evolutive (LBBE, Lyon) dans le cadre du projet TRIPTIC. Le traitement des souches thélytoques par antibiotique a été réalisé en grande partie par Géraldine Groussier (Institut Sophia-Agrobiotech – Equipe RDLB) en 2018.

2.3.1 Matériel biologique

La recherche des endosymbiotes *Wolbachia* a été réalisée :

- Sur au moins une souche (de 1 à 48) par cluster moléculaire de *Trichogramma* identifié dans la phylogénie multilocus afin d'améliorer nos connaissances sur le statut d'infection des différents clusters ;
- Sur des souches thélytoques du complexe cacoeciae - embryophagum (n=48) et des espèces *T. oleae* (n=2), *T. cordubensis* (n=8) et *T. misB* (n=1), afin de vérifier si ce caractère est génétique ou induit par *Wolbachia* (n=48) ;

- Sur toutes les souches utilisées (n=28) pour les expériences de croisements (cf. § 2.4) afin, éventuellement, de restaurer une reproduction sexuée, d'évaluer l'échec de certains croisements par des phénomènes d'incompatibilité cytoplasmique.

2.3.2 Détection moléculaire de *Wolbachia*

L'ADN génomique d'un trichogramme est extrait suivant la méthode PrepGEM Insect (Zygem) dans 15µl de tampon d'extraction pendant 3h à 75°C puis 5min à 95°C pour désactiver la protéinase. L'extrait est ensuite conservé à -20°C.

Une portion du gène *ftsZ* (*Filamenting temperature-sensitive mutant Z*) de 400pb codant pour une protéine impliquée lors de la division cellulaire des bactéries, a été amplifiée en utilisant 1,2 µl d'extraction d'ADN dans 15µl de volume de réaction final composé de 0.84µl de Taq Qiagen (5U/µl), 1,5 mM MgCl₂ et 0,5 µM d'amorces forward *FtsZ-F* (TTGCAGAGCTTGGACTTGAA) et reverse *FtsZ-R* (CATATCTCCGCCACCAGTAA), 200µM de dNTP. Les conditions de PCR étaient les suivantes : 94°C pendant 5 minutes, 35 cycles à 94°C pendant 30 secondes, 58°C pendant 1 minute, 72°C pendant 1 minute, avec une élongation finale à 72°C pendant 10 minutes. La présence de l'amplicon a été analysée avec le kit "QIAxcel DNA Fast Analysis" (QIAGEN S.A.S) sur l'analyseur Qiaxcel en présence de témoins positifs et négatifs respectivement un *Psytalia lounsburyi* infecté et non infecté par *Wolbachia* et un témoin négatif avec de l'eau.

2.3.3 Expérimentations complémentaires

2.3.3.1 PCR quantitative et caractérisation MLST

Afin de déterminer plus précisément les statuts d'infection à *Wolbachia*, deux autres approches ont été adoptées par le laboratoire LBBE dans le cadre du projet ANR TRIPTIC :

- d'une part, la détection d'infections par qPCR, une approche potentiellement plus précise ;
- d'autre part, une caractérisation par « Multi Locus Sequence Typing » (MLST - (Baldo *et al.*, 2006) basée sur l'amplification de 6 marqueurs : *ftsZ*, *gatB*, *coxA*, *hcpA*, *fbpA*, et *wsp*.

Les espèces ou complexes d'espèces qui ont été ciblés sont notamment les 16 clusters suivants : I « daumalae-evanescens », II « principium », III « cordubensis », IV « chilonis », V « misB », VI « brassicae-euproctidis », VII « misA », VIII « semblidis », XI

« bourarachae », XII « cacoeciae/embryophaguum », XIII « swC », XIV « pretiosum », 2 « achaeae », 3 « swA », 5 « swD », 9 « oleae » (cf. Tableau 1 - voir §7).

Dix clusters n'ont pas fait l'objet d'une recherche de *Wolbachia* par manque de matériel biologique (IX « swE », X « dendrolimi », 1- « ostrinia », 4 « nerudai », 6 et 7 « exigum », 8 « swG ») ou par manque de temps (a « swI », b « swK », c « swJ »).

2.3.3.2 Traitements par antibiothérapie

Afin d'évaluer le rôle de *Wolbachia* sur l'induction de thélytoquie, les souches thélytoques infectées ou non par *Wolbachia* ont été traitées par antibiothérapie. Les souches ont été maintenues en élevage pendant au moins 3 générations selon les conditions habituelles (utilisation de l'hôte de substitution *E. kuehniella* ; température de 22 ± 2 ° C ; photopériode : Nuit/Jour=8h/16h) avec toutefois un apport de miel contenant une dose de 0,5% de rifampicine (en % de poids sec de nourriture). Deux critères sont ensuite surveillés au cours du temps, la survie de la lignée et l'apparition de mâles.

2.4 Tests de croisement

Ces travaux ont été réalisés sur la période 2016 - 2017. Ces activités expérimentales ont été réalisées la plupart du temps en binôme m'impliquant ainsi qu'un stagiaire : Hélène Wera (Master 1, Université de Nice Sophia-Antipolis - blocs 2 à 6), Alexis Estève (BTS1, Lycée Jules Ferry - bloc 7), Valentin Beck (BTS1, Lycée Jules Ferry - bloc 9) et Camille Duraj (BTS1, Lycée Jules Ferry - bloc 10).

2.4.1 Matériel biologique utilisé

Parmi les souches vivantes disponibles au sein du CRB EP-Coll, le choix des souches utilisées pour déterminer les compatibilités reproductives a été fait sur la base du « barcode » *COI* court (271 positions) afin de couvrir le mieux possible la diversité moléculaire :

- (i) inter-espèces, en choisissant des souches dans l'ensemble des clusters moléculaires ;
- (ii) intra-espèces, en choisissant des souches identifiées dans les sous-clusters *COI* notamment « brassicae-euproctidis », « daumalae-evanescens », « misB » et « principium ».

Les souches, leurs origines et informations associées sont listées dans le Tableau 6.

Tableau 6 : Souches utilisées pour les expérimentations de croisement

Les haplotypes COI et l'identification moléculaire ont été obtenus à partir des analyses de barcoding §3- Barcoding (voir tableau 1 en §7 ainsi que les figures 3 et 4)

Complexe	Espèce	Souche	Origine de la souche	Source	Barcode COI
2 - achaeae	achaeae	T_achaeae_biotop	Espagne, ?	BIOLINE Agrosiences	Hap_083
XI - bourarachae	bourarachae	ISA6646	France, 2016	Projet TripTIC	Hap_128
VI - brassicae-euproctidis	brassicae	pr002	France, 2015	Projet TripTIC	Hap_027
		pr010	France, 2015	Projet TripTIC	Hap_005 / Hap_027
		Tbra	?France, ?	Collections historiques	Hap_005
		pug247	France, 2015	Projet TripTIC	Hap_027 / Hap_039
		isa3080	France, 2015	Projet TripTIC	Hap_036
		pr001	France, 2015	Projet TripTIC	Hap_036
		pr006	France, 2015	Projet TripTIC	Hap_036
	euproctidis	Egypteupro	Egypte, ?	Collections historiques	Hap_068
IV - chilonis	chilonis	Taiwan	Taiwan, ?	Collections historiques	Hap_091
I - daumalae-evanescens	daumalae	Tdaum	France, ?	Collections historiques	Hap_002 ; Hap_008
		evanescens	isa1797	France, 2015	Projet TripTIC
		am002	France, 2015	Projet TripTIC	Hap_021
		mur009	France, 2015	Projet TripTIC	Hap_011
		af017	France, 2015	Projet TripTIC	Hap_021
		tsm004	France, 2015	Projet TripTIC	Hap_044
VII - misA	misA	al22	France, 1990	Collections historiques	Hap_004
V - misB	misB	ISA14313	France, 2015	Projet TripTIC	Hap_155
	ex-misC	ISA15550	France, 2015	Projet TripTIC	Hap_132
II - principium	principium	ISA1729	France, 2015	Projet TripTIC	Hap_113
		81a	France, 1975	Collections historiques	Hap_001
VIII - semblidis	semblidis	FPV025	France, 2015	Projet TripTIC	Hap_017
		FPV034_A	France, 2015	Projet TripTIC	Hap_112
		PR007	France, 2015	Projet TripTIC	Hap_017
		ACJYR0030	France, 2015	Projet TripTIC	Hap_017
XIII - swC	swC	FLO201	France, 2016	Projet TripTIC	Hap_121
5 - swD	swD	JAP0018	Japon, 2016	Projet TripTIC	Hap_137

2.4.2 Réalisation des croisements

Le mode opératoire précis est accessible dans le document standardisé MO-RDLB-015. Pour une question de faisabilité, les expériences de croisement ont été menées par bloc de 6 souches maximum. Neuf blocs ont été nécessaires pour déterminer les compatibilités reproductives entre 28 souches sexuées. Pour chaque bloc, le plan expérimental impliquait systématiquement (i) des tests sur des femelles vierges pour détecter un éventuel mode de reproduction thélytoque, (ii) des croisements parentaux ($1\text{♀}_i \times 1\text{♂}_i$, i indiquant une souche particulière), (iii) des croisements inter-souches bidirectionnels ($1\text{♀}_i \times 1\text{♂}_j$ et $1\text{♀}_j \times 1\text{♂}_i$, i et j

indiquant deux souches différentes). Les femelles vierges et les couples ont été maintenus tout au long de leurs vies dans des tubes contenant des œufs hôtes d'*E. kuehniella*. Les conditions abiotiques étaient de 22 ± 2 ° C, une humidité relative quotidienne comprise entre 50% et 70% et une photopériode nuit / jour de 8h / 16h. La descendance a ensuite été sexée et, pour chaque croisement parental ou inter-souches, deux paramètres ont été mesurés pour évaluer le niveau de compatibilité de la reproduction :

- RC1 : la proportion de croisements qui ont fourni au moins une fille
- RC2 : la proportion moyenne de filles dans les croisements où au moins une fille a été observée.

Sur la base de ces résultats, quatre niveaux d'(in)compatibilités de reproduction ont été définis :

- Incompatibilité totale : RC1 = 0%
- Faible compatibilité : $0 < RC1 \leq 30\%$
- Compatibilité moyenne : $30\% < RC1$ et $RC2 \leq RC_{seuil}$, le RC_{seuil} étant défini comme la plus faible valeur RC2 observée dans les croisements parentaux (1♀i x 1♂i).
- Forte compatibilité : $RC_{seuil} < RC2$ et $30\% < RC1$

D'une façon générale, $RC2_{seuil}$ est fonction des souches parentales utilisées et des conditions d'expérimentations. $RC2_{seuil}$ varie donc de la façon suivante :

- dans le paragraphe 3-Barcoding COI, la valeur $RC2_{seuil}$ est respectivement 55% et 33% pour les complexes « brassicae euproctidis » et « daumalae-evanescens ».
- dans le paragraphe 4 -Multilocus, la valeur $RC2_{seuil}$ du complexe « V-misB » est de 84%
- dans le Supplément S1 : « croisements généraux », la valeur de $RC2_{seuil}$ est définie comme la valeur de RC2 la plus faible observée chez les croisements parentaux pris dans leur globalité c'est-à-dire 32% à l'exclusion des croisements parentaux *swD*, très médiocres ($RC1=20\%$, $RC2=0.01\%$).

Les résultats généraux liés à l'ensemble des croisements sont présentés dans le Supplément S1, S2, S3. Seuls les résultats de croisements pertinents par rapport aux questions posées sont détaillés dans les paragraphes suivants.

3 CARACTERISATION DE LA BIODIVERSITE DES TRICHOGRAMMES VIA LE BARCODING COI

3.1 Introduction

Comme mentionné dans l'introduction, la systématique du genre *Trichogramma* reste mal connue. Une meilleure compréhension de la diversité de *Trichogramma* peut être envisagée par le séquençage d'une partie de la sous-unité 1 de la Cytochrome oxydase (*COI*) (DNA barcoding - Hebert *et al.*, 2003). Rétrospectivement, les études de barcoding ont toutefois souligné que des précautions doivent être prises pour interpréter les résultats moléculaires uniques en termes d'artefacts techniques (erreurs PCR / séquençage, co-amplification des pseudogènes ou contaminants), biologiques (hétéroplasmie ou présence de symbiotes) ou évolutifs (l'introgession de l'ADNmt ou l' « incomplete lineage sorting ») qui peuvent rendre la délimitation des espèces impossible (Cruaud *et al.*, 2017; Frézal and Leblois, 2008; Shokralla *et al.*, 2015; Song *et al.*, 2008). La qualité de l'échantillonnage est également un aspect clé à considérer et doit couvrir du mieux possible les variations inter et intraspécifiques (Meyer and Paulay, 2005). L'approche du barcoding doit donc être couplée à d'autres sources d'informations (caractérisation morphologique, expériences de croisement, données écologiques, etc.) dans une démarche de taxonomie intégrative (Schlick-Steiner *et al.*, 2010; Will *et al.*, 2005). Des résultats convaincants ont ainsi été obtenus en utilisant de telles approches pour plusieurs groupes d'hyménoptères (Al Khatib *et al.*, 2014; Gebiola *et al.*, 2017; Mottern and Heraty, 2014; Rugman-Jones *et al.*, 2009).

Le but de cette étude est un premier pas vers cet objectif et la volonté de combler un manque évident. En effet, une interrogation récente sur GenBank (2017-11-27) n'a mis en évidence que 759 séquences *COI* de *Trichogramma*, la plupart d'entre elles n'ayant pas été attribuées à un nom d'espèce. Nous avons ainsi généré un grand nombre de barcode *COI* provenant du taxon *Trichogramma* (plus de 3100 individus) obtenus à partir de 600 souches de laboratoire, prélevés en France et dans d'autres pays européens ou récupérés auprès de GENBANK (voir Tableau 1 - voir §7). Un arbre phylogénétique et un réseau d'haplotypes ont été reconstruits pour étudier le groupement moléculaire et sa congruence avec les identifications morphologiques (lorsqu'elles étaient disponibles) a été prise en compte pour éviter les artefacts possibles énumérés ci-dessus. Nous avons finalement effectué des expériences de croisement pour étudier plus précisément la délimitation des espèces dans deux complexes d'espèces.

3.2 Résultats

Les résultats présentés ci-dessous dans le paragraphe § 3.2 et dont le § 2-Matériels et Méthodes sont décrits dans le paragraphe § 2.1, ont fait l'objet de la rédaction d'une publication scientifique (Warot *et al.*, in prep) ainsi que d'une communication dans un congrès francophone (Warot *et al.*, 2017).

3.2.1 Diversité haplotype

Le court « barcode COI » (271pb) a été obtenu pour plus de 3100 individus. L'analyse de l'ensemble des données composé de ces séquences a révélé 191 haplotypes. De ces 191 haplotypes, 12 provenaient de taxons externes au genre *Trichogramma* (autres Chalcidoidea, supposés Platygastriidae ou même l'hôte d'élevage *E. kuehniella*) tandis que les 179 restants identifiés comme appartenant au genre *Trichogramma* (Tableau 1 - voir §7). Sur la base des fréquences de notre ensemble de données, ces haplotypes ont été répartis comme suit (Tableau 7).

- Catégorie A (très rare - une seule copie) : 52 haplotypes (29% du total)
- Catégorie B (rare - un seul échantillon) : 51 haplotypes (28%)
- Catégorie C (fréquente - 2 à 10 échantillons) : 49 haplotypes (27%)
- Catégorie D (commune - > 10 échantillons) : 17 haplotypes (9%)

21 de ces haplotypes (12% du total) présentaient au moins une variation non-synonyme dans leurs séquences d'acides aminés par rapport à la séquence partagée par tous les autres (Tableaux 1 en §7 et 7). Ces 21 haplotypes se trouvaient tous répartis dans les trois premières catégories (A-C) (respectivement n = 9, n = 7 et n = 5).

En outre, dix séquences de Genbank (catégorie E) ont été ajoutées. Parmi celles-ci, trois séquences (30% du total) présentaient des variations non-synonymes dans les séquences d'acides aminés, y compris la séquence unique disponible pour *T. danausicida* (Nagaraja).

Tableau 7 : Distribution des haplotypes COI en fonction des déterminations moléculaires

Les catégories (A-E) sont détaillées dans le §2.1.3 du Matériels et Méthodes.

L'étiquette "varprot" indique le nombre d'haplotypes pour lesquels au moins une variation de la séquence d'acides aminés a été observée.

	Total	Distribution des haplotypes					varprot	Identification, sources
		A	B	C	D	E		
<u>Clusters</u>								
I - daumalae_evaneszens	25	6	9	5	5	0	3	BP ; JS ; Genbank
II - principium	4	0	1	3	0	0	0	BP
III - cordubensis	11	4	2	2	3	0	4	BP ; Genbank
IV - chilonis	9	2	1	5	0	1	0	voir Benvenuto et al. 2012
V - misB	10	1	3	6	0	0	2	–
VI - brassicae_euproctidis	26	7	7	9	3	0	8	BP ; JS ; Genbank
VII - misA	2	0	1	1	0	0	0	–
VIII - sembridis	4	0	2	2	0	0	0	BP
IX - swE	2	0	1	1	0	0	0	–
X - dendrolimi	5	0	2	0	0	3	0	BP ; Genbank
XI - bourarachae	9	3	3	2	1	0	1	BP
XII - cacaoeciae_embryophagum	37	12	12	7	3	3	4	BP ; Genbank
XIII - swC	3	0	0	3	0	0	0	–
XIV - pretiosum	3	0	1	1	1	0	0	BP ; Genbank
<u>Haplotypes isolés</u>								
1 - ostriniae	1	0	0	0	0	1	0	Genbank
2 - achaeae	1	0	0	1	0	1	0	JS ; Genbank
3 - swA	1	1	1	0	0	0	0	–
4 - nerudai	1	0	1	0	0	0	0	BP
5 - swD	1	0	1	0	0	0	0	–
6 - exiguum 2	1	0	1	0	0	0	0	BP
7 - exiguum 1	1	0	1	0	0	0	0	BP
8 - swG	1	0	0	1	0	0	0	–
9 - oleae	1	0	0	0	1	0	1	BP
10 - danausicida	1	0	0	0	0	1	1	Genbank
<u>Non affiliés</u>	17	16	1	0	0	0	1	

3.2.2 Structuration moléculaire

La structuration moléculaire a été étudiée en utilisant deux approches complémentaires :

- le réseau des haplotypes (Figure 3) utilisant la méthode Median-joining et englobant les 179 haplotypes d'ADN pour plus de 3100 individus,
- l'arbre Neighbor-Joining (Figure 4) reconstruit sur la base de la distance de Tamura-Nei en excluant les haplotypes de la catégorie A et les séquences d'acides aminés avec variations non synonymes (nommé "varprot" dans les Tableaux 1 et 7), soit 114 séquences.

À partir de ces deux approches, 14 groupes différents (numération latine : I à XIV sur les Figures 3 et 4) ont été mis en évidence, chacun de ces groupes étant défini comme un ensemble monophylétique d'au moins 2 haplotypes, généralement soutenu par une valeur élevée de bootstrap (> 80%) dans l'arbre Neighbor-Joining (Figure 4), les seules exceptions étant les groupes I, VI et XII en positions basales dans l'arbre. Ces trois groupes étaient toutefois bien différenciés sur le réseau d'haplotypes (Figure 3). En fait, les délimitations des clusters les plus compliquées ont également été observées pour ces clusters (des valeurs de bootstrap faibles sont observées à leur base) (Figure 4). A propos du cluster I, l'haplotype Hap_074 (mis en évidence par un * dans la Figure 3 et la Figure 4) a finalement été inclus tandis que l'haplotype Hap_103 (mis en évidence par **) a été exclu en raison de sa position éloignée dans le réseau des haplotypes (Figure 3). Un raisonnement similaire a été appliqué pour le cluster VI pour lequel l'haplotype Hap_124 (mis en évidence par ***) a finalement été inclus. Sur la base du réseau des haplotypes, plus de trente haplotypes ont été regroupés dans le cluster XII malgré une structuration interne isolant les haplotypes Hap_196, Hap_197 et Hap_198 (rectangle en pointillés sur la Figure 4). Ces trois haplotypes étaient en effet bien entrelacés avec les autres dans le réseau des haplotypes (ellipse en pointillés sur la Figure 3). Certains haplotypes n'ont été attribués à aucun de ces 14 groupes.

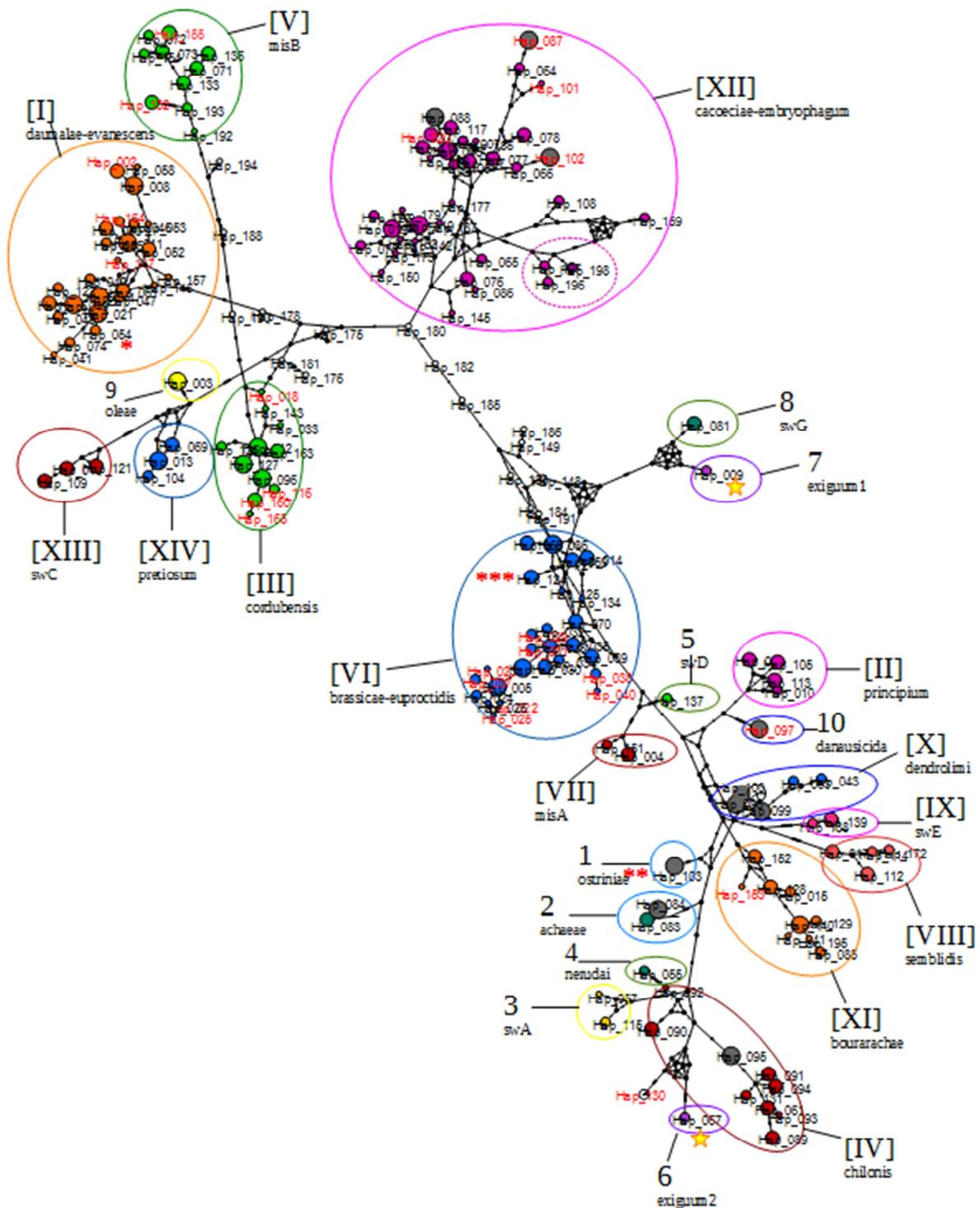


Figure 3 : Réseau d'haplotypes obtenu grâce à PopARTs

Le réseau a été déduit en utilisant la méthode Median Joining avec $\epsilon = 0$ (paramètre par défaut).

Chaque ellipse correspond à un cluster moléculaire (numération latine) ou à un haplotype isolé valide (numération arabe) tel que défini dans les résultats. À l'exception des cercles gris (séquences Genbank - Catégorie E), la taille du cercle correspond à sa catégorie d'occurrences (catégories A-D) (voir § 2- Matériels et Méthodes). Les petits cercles blancs correspondent à des haplotypes très rares (catégorie A) et non-affiliés (voir résultats). Les haplotypes en rouge présentent au moins une variation dans leurs séquences d'acides aminés. L'ellipse en pointillé, les astérisques rouges et les étoiles jaunes indiquent des haplotypes particuliers (voir résultats).

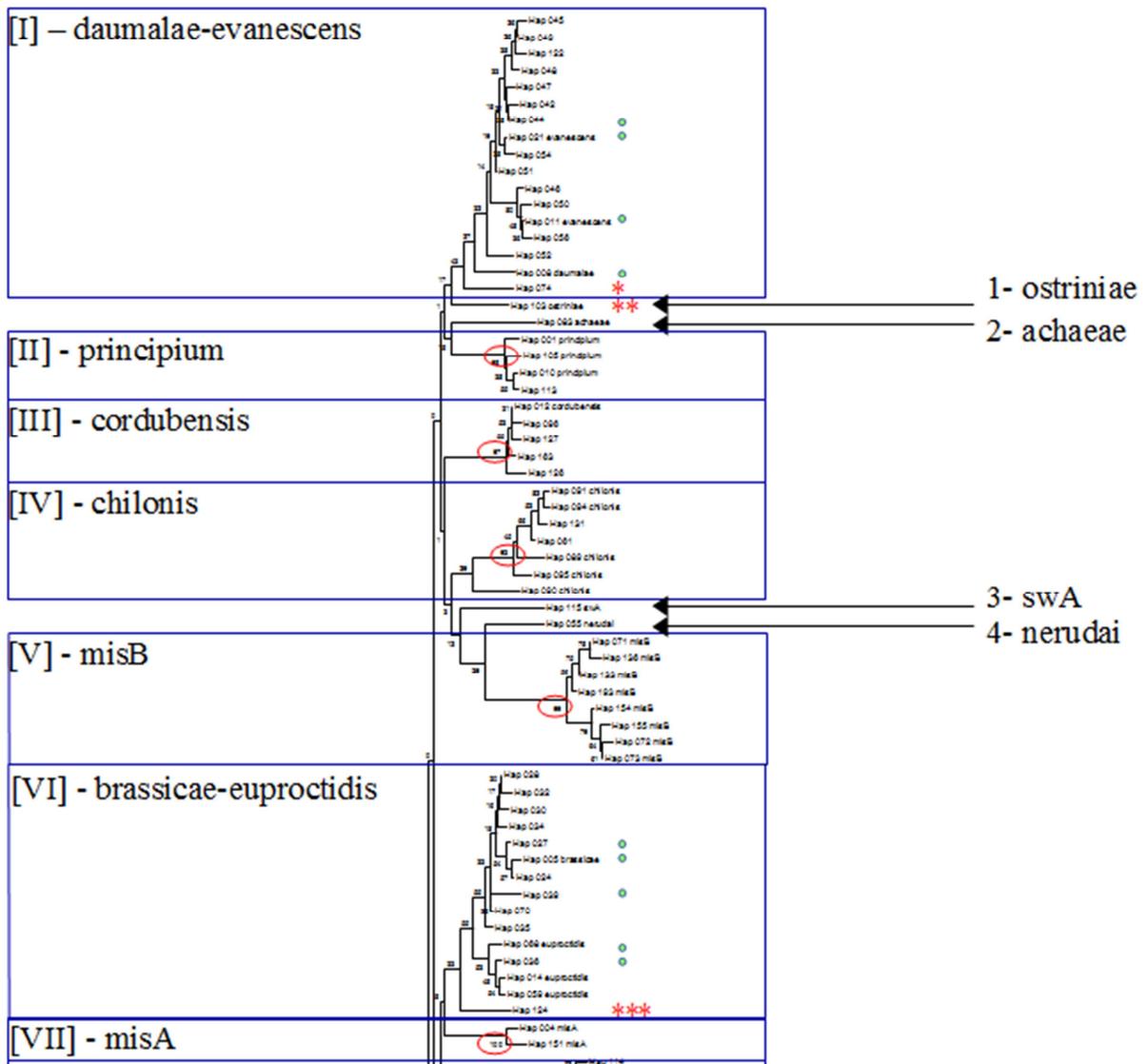


Figure 4 (début) : Arbre phylogénétique obtenu grâce à MEGA7.0

L'arbre phylogénétique a été construit à partir de 114 haplotypes c'est-à-dire des séquences uniques d'acides aminés et appartenant aux catégories d'occurrences B-E (voir §2-Matériels et Méthodes). L'histoire évolutive a été déduite en utilisant la méthode Neighbor-Joining à partir des distances évolutives calculées à l'aide de la méthode Tamura-Nei, 500 répétitions ont été utilisées pour le calcul du bootstrap. Au total, 271 positions ont été prises en compte dans l'analyse, toutes les positions contenant des lacunes et les données manquantes ont été éliminées.

Les rectangles et les flèches correspondent respectivement aux groupes moléculaires (numération latine) ou aux haplotypes isolés (numération arabe) tels que définis dans les résultats. Le rectangle en pointillés, les astérisques rouges et les étoiles jaunes indiquent des haplotypes particuliers (voir §3.2 Résultats) tandis que les cercles verts mettent en évidence des haplotypes impliqués dans des expériences de croisement.

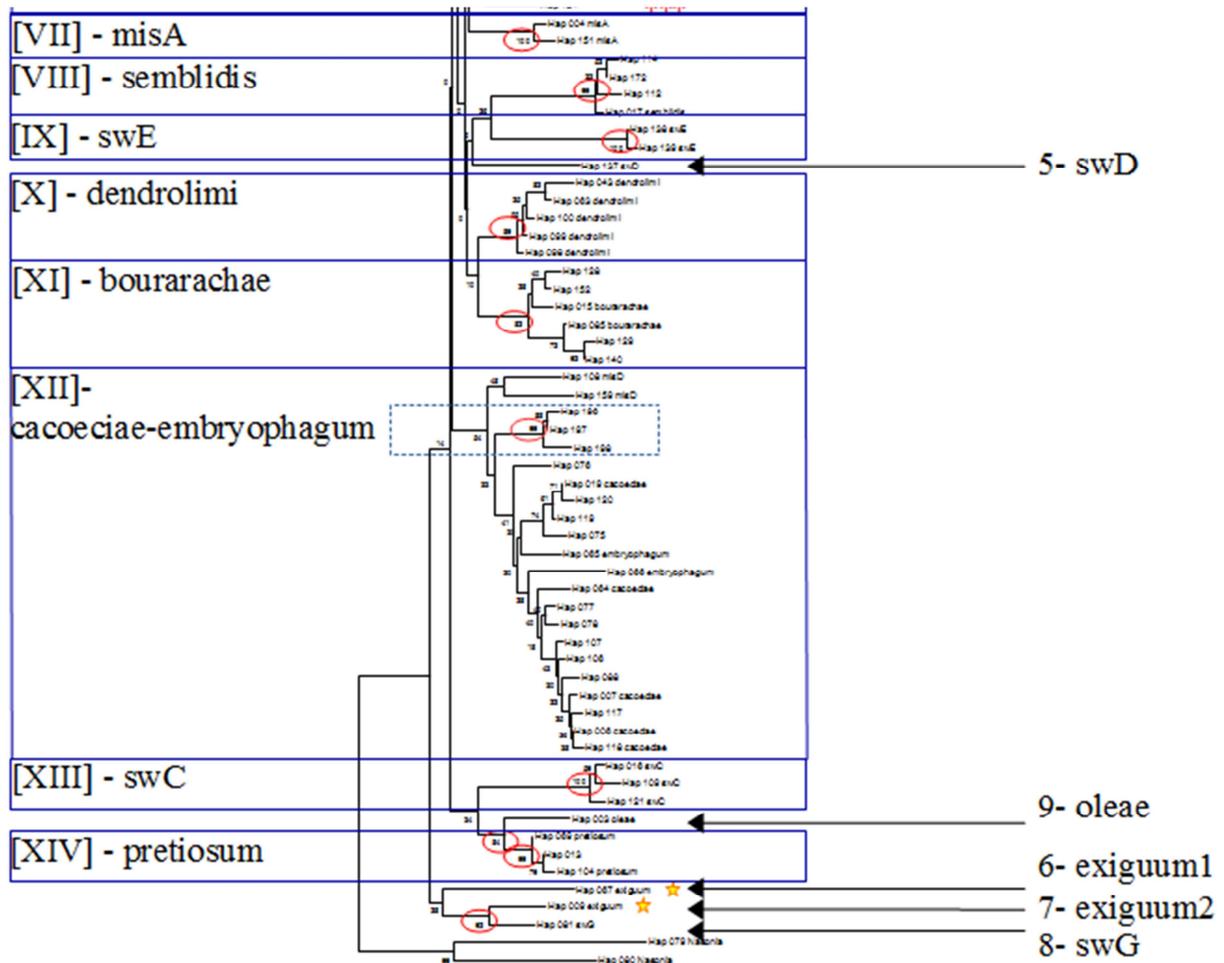


Figure 4 (fin) : Arbre phylogénétique obtenu grâce à MEGA7.0

L'arbre phylogénétique a été construit à partir de 114 haplotypes c'est-à-dire des séquences uniques d'acides aminés et appartenant aux catégories d'occurrences B-E (voir §2-Matériels et Méthodes). L'histoire évolutive a été déduite en utilisant la méthode Neighbor-Joining à partir des distances évolutives calculées à l'aide de la méthode Tamura-Nei, 500 répétitions ont été utilisées pour le calcul du bootstrap. Au total, 271 positions ont été prises en compte dans l'analyse, toutes les positions contenant des lacunes et les données manquantes ont été éliminées.

Les rectangles et les flèches correspondent respectivement aux groupes moléculaires (numération latine) ou aux haplotypes isolés (numération arabe) tels que définis dans les résultats. Le rectangle en pointillés, les astérisques rouges et les étoiles jaunes indiquent des haplotypes particuliers (voir §3.2 Résultats) tandis que les cercles verts mettent en évidence des haplotypes impliqués dans des expériences de croisement.

Parmi ceux-ci, neuf haplotypes semblaient valides (pas de variation de leur séquence d'acides aminés COI) et clairement isolés les uns des autres (Figures 3 et 4, numérotation arabe de 1 à 9). Cet ensemble de neuf haplotypes comprenait (i) l'accession GENBANK pour *T. ostriniae* Pang & Chen (Hap_103), (ii) cinq haplotypes rares (catégorie B : Hap_009 "exiguum 1", Hap_055 "nerudai", Hap_067 "exiguum 2", Hap_115 "swA" et Hap_137 "swD"), (iii) deux haplotypes assez fréquents (catégorie C : Hap_081 "swC", Hap_83 "achaeae"), (iv) un très fréquent (catégorie D : Hap_003 "oleae"). Un dixième haplotype (catégorie E, Hap_097, identifié comme *Trichogramma danausicida*) est apparu clairement isolé de tous les autres haplotypes du réseau (Figure 3) mais a présenté des variations au niveau des séquences d'acides aminés.

Enfin, 16 haplotypes très rares (catégorie A) et 1 rare (catégorie B) n'étaient pas affiliés (étiquetés comme « non affiliés » dans les Tableaux 1 et 7 - cercles blancs dans la Figure 3) et sont susceptibles d'être des artefacts.

A posteriori, la délimitation des clusters permet de s'intéresser au « gap de barcoding » c'est-à-dire la différence entre la distribution des distances intraspécifiques par rapport à celles interspécifiques (Figure 5). Ainsi, les médianes des distances évolutives sont de l'ordre de 0,01 en intraspécifique et 0,07. Le chevauchement entre les deux distributions est très limité puisque, par exemple, les trois premiers quartiles de la distribution intraspécifique sont non chevauchants avec les 3 derniers quartiles de la distribution interspécifique. Approximativement, la frontière entre les variabilités intra et interspécifiques semblent se situer autour de 0,04 - 0,05.

Par rapport à cette tendance générale, on observe certaines distances interspécifiques anormalement basses ($0.02 < d < 0.03$). Après vérification, elles impliquent toutefois systématiquement des haplotypes de la catégorie A identifiées comme de probables artefacts. A l'inverse, on observe les distances intraspécifiques les plus élevées pour les deux complexes « daumalae-evanescens » (en particulier, celle entre Hap008-Tdaum et Hap074) et « cacoeciae-embryophagum ». Concernant les données intraspécifiques hors complexes, les distances les plus élevées (ronds sur Figure 5) se rapportent au cluster « chilonis » à l'intérieur duquel des isolements reproducteurs ont déjà été observés lors d'une précédente étude (Benvenuto *et al.*, 2012). Enfin, on peut noter que, par comparaison des données impliquant ou non des variations dans leur séquence protéique, les distances impliquant au moins une séquence présentant des variations dans leur séquence protéique (« VARPROT » sur Figure 5) ont une tendance à être plus élevées.

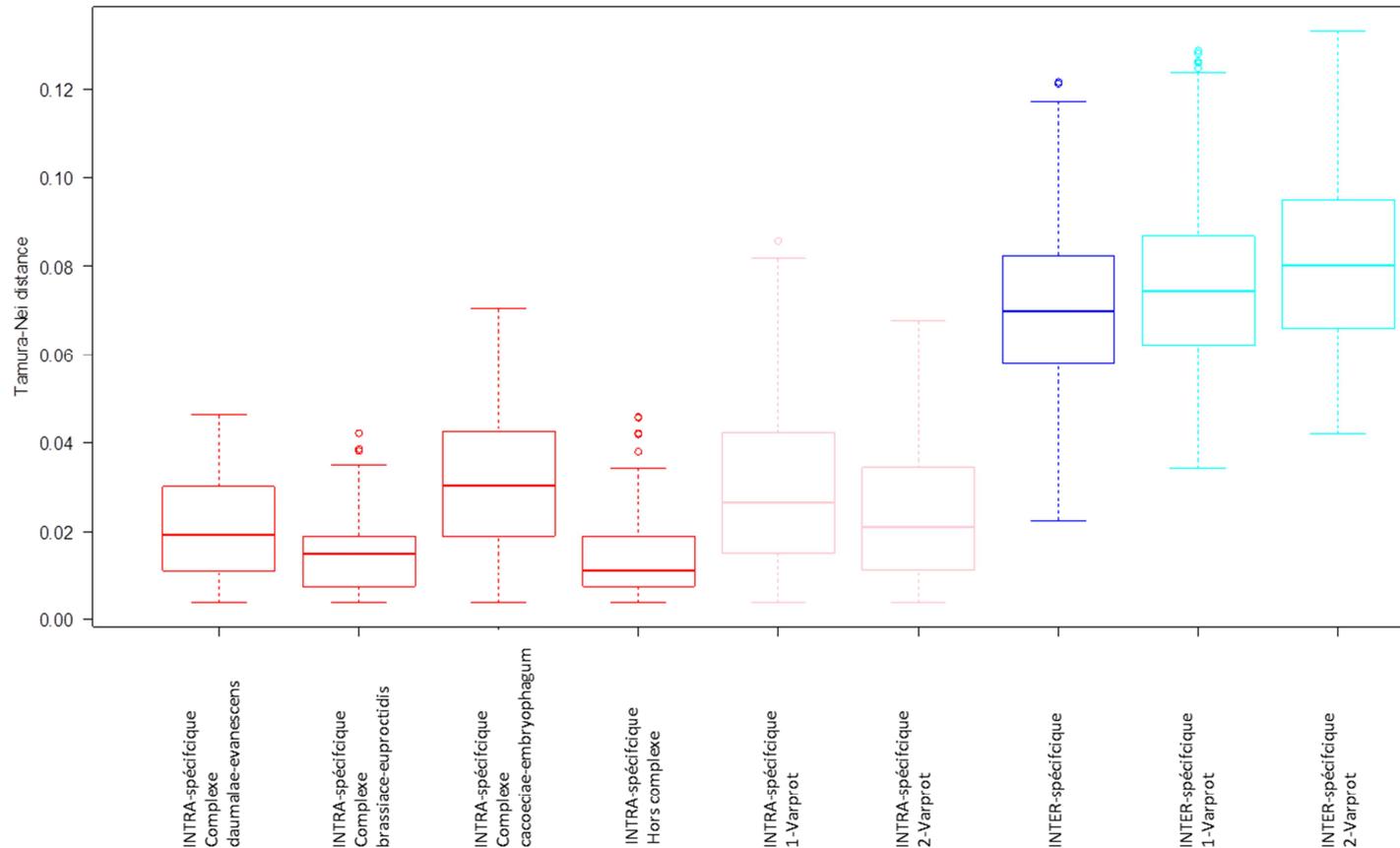


Figure 5 : Variabilité des distances évolutives des données intra et interspécifique

Variabilité des distances évolutives calculées sous le modèle de substitution Tamura Nei de plusieurs séries définies en Barcoding COI, catégorisées de gauche à droite sur le graphique, comme suit :

- en intraspécifique : les 3 complexes « daumalae-evanescens », « brassicae-euproctidis », et « cacoeciae-embryophagum », les données intraspécifiques sauf complexes et varprot, les données intraspécifiques impliquant une séquence avec variation d'acides aminés (1-Varprot), les données intraspécifiques impliquant deux séquences avec variation d'acides aminés (2-Varprot),
- en interspécifique : les données interspécifiques sauf Varprot, les données impliquant une séquence avec variation d'acides aminés (1-Varprot), les données interspécifiques impliquant deux séquences avec variation d'acides aminés (2-Varprot),

Le graphique représente une boîte avec la médiane, les quartiles supérieurs et inférieurs, et des moustaches avec les valeurs minimales et maximales. Les cercles représentent les valeurs aberrantes non prises en compte pour le calcul des quartiles.

3.2.3 Congruence avec les anciennes identifications

Le nom de l'espèce tel que déduit de l'examen morphologique (par B. Pintureau pour les collections historiques et par les auteurs qui ont déposé les séquences dans Genbank) a été porté sur les clusters et les haplotypes isolés. Parmi les 23 clusters / haplotypes isolés, 11 étaient associés à un nom d'espèce unique : *T. achaeae*, *T. bourarachae* Pintureau et Babault, *T. chilonis* Ishii, *T. cordubensis* Vargas et Cabello, *T. dendrolimi* Matsumura, *T. oleae* Voegelé et Pointel, *T. ostriniae*, *T. nerudai* Pintureau et Gerding, *T. pretiosum* Riley, *T. principium* Sugonjaev et Sorokina et *T. semblidis* (Aurivillius) (Figure 4). Étonnamment, deux haplotypes divergents (Hap_009 et Hap_067) ont été assignés à *T. exiguum* Pinto & Platner (voir les étoiles jaunes sur les Figures 3 et 4). Les trois plus grands groupes (I, VI et XII) comprenaient deux noms d'espèces différents et ont donc été considérés comme des complexes d'espèces : respectivement "daumalae-evanescens", "brassicae-euproctidis" et "cacoeciae-embryophagum". Enfin, sept clusters (misA, misB, swC et swE) ou haplotypes isolés (swA, swD, swG) n'ont été attribués à aucun nom d'espèce disponible (Tableau 7).

3.2.4 Expérimentations de croisements

A posteriori, les résultats d'(in)compatibilités reproductives diffèrent fortement entre les deux groupes, "brassicae-euproctidis" et "daumalae-evanescens".

Dans le complexe « brassicae-euproctidis », aucune des femelles vierges (n = 78, toutes souches incluses) n'a pu produire de filles et les croisements parentaux ont été très efficaces avec un RC1 moyen de 88% (entre 75% et 100% - n = 93, toutes souches incluses) et un RC2 moyen de 63% ([55% ; 80%] - n = 93, toutes souches incluses). Comme résumé dans la Figure 6A, des niveaux variables de (in)compatibilités ont été observés dans des croisements inter-souches avec un RC1 moyen de 51% ([0% ; 100%] - n = 189, tous croisements inclus) et un RC2 moyen de 41 % ([1% ; 93%] - n = 132). En particulier, la souche "Egypteupro" semblait être entièrement (flèches noires) ou fortement (flèches rouges) isolée reproductivement de la plupart (5/6) des autres souches. Cependant, son croisement était possible au moins avec la souche "ISA3080" qui était elle-même compatible avec la souche "PUG247". D'autres incompatibilités unidirectionnelles complètes ou fortes (respectivement flèches noires et rouges) ont également été observées dans ce complexe même entre deux souches ("ISA3080" et "PR001") partageant le même barcode COI (Hap_036) (Figure 6A).

Dans le complexe « daumalae-evanescens », aucune des femelles vierges n'a pu produire de filles (n = 23, toutes souches incluses). Les performances des croisements parentaux étaient plus faibles que celles des autres complexes, avec un RC1 moyen de 84% ([80% ; 88%] - n = 29) et un RC2 moyen de 45% ([33% ; 57%] - n = 29). Pris dans leur ensemble, les

croisements entre souches présentent en général des performances similaires avec un RC1 moyen de 92% ([67% ; 100%] - n = 60) et un RC2 moyen de seulement 47% ([24% ; 54%] - n = 60). Comme le montre la Figure 6B, les deux plus faibles compatibilités reproductives (flèches orange) ont été observées pour deux paires de souches différentes. Une expérience complémentaire impliquant un mélange de fils et les filles (G1) a montré la présence de petites-filles dans la génération suivante. Aucune fille vierge n'a pu produire de petites filles (n=84, toutes souches incluses) excluant donc toute forme de thélytoquie.

Fig 6A

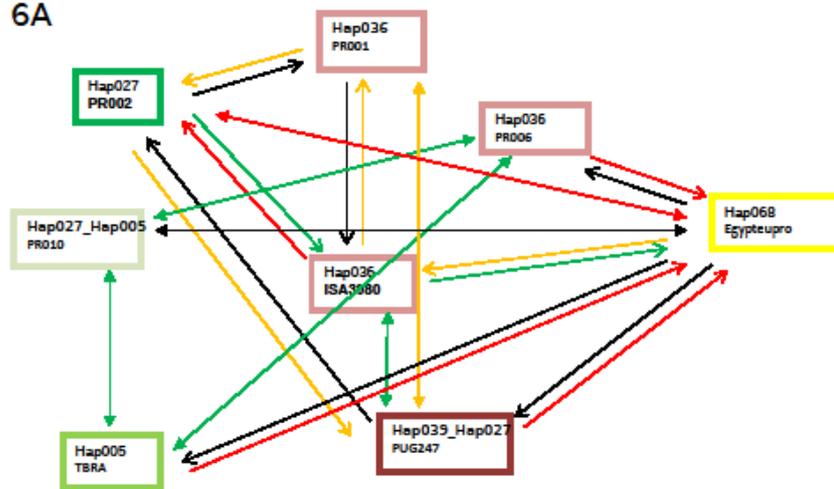


Fig 6B

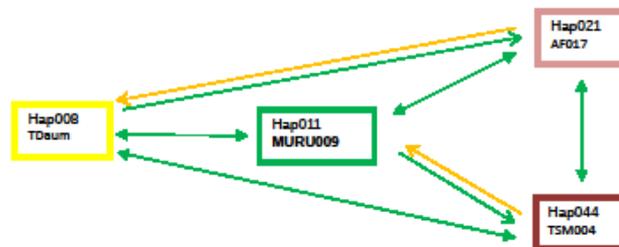


Figure 6 : Résumé des (in)compatibilités de reproduction dans deux complexes d'espèces, "brassicae - euproctidis" (A) et "daumalae - evanescens" (B).

Pour chaque complexe, la couleur de la flèche se réfère au niveau de la (in)compatibilité de reproduction comme suit : Noir = Incompatibilité totale ; Rouge = Compatibilité faible ; Orange = Incompatibilité moyenne ; Vert = Compatibilité élevée.

Des flèches doubles ont été utilisées lorsque les niveaux d'(in)compatibilité de reproduction étaient équivalents quelle que soit la direction des croisements ($1\text{♀}_i \times 1\text{♂}_j$ et $1\text{♀}_j \times 1\text{♂}_i$, i et j indiquant deux souches différentes).

3.3 Discussion

En investissant massivement sur le séquençage d'ADN de Sanger de plus de 3100 individus provenant de sources diverses, ce travail a probablement généré le plus grand jeu de barcode COI disponible jusqu'à présent sur les espèces du genre *Trichogramma*. En se concentrant sur un petit fragment de la COI, qui pourrait être facilement et de manière fiable, accessible aux différents utilisateurs finaux (scientifiques mais aussi producteurs d'agents de lutte biologique ou organismes de régulation), nous avons montré qu'il était possible de caractériser en routine des souches avec une précision qui relève du niveau de l'espèce ou de clusters d'espèces à l'intérieur desquels nous n'avons pas observé de souches complètement isolées hormis des changements de mode de reproduction (*T.pretiosum* par exemple). Ceci reste toutefois à confirmer en élargissant l'aire géographique d'échantillonnage afin de prendre en compte au mieux la variabilité intraspécifique et plus généralement la couverture taxonomique des *Trichogramma* (espèces rares spécialisées).

La seule difficulté technique rencontrée au cours de cette étude était l'apparition d'haplotypes correspondant vraisemblablement à des pseudogènes ou à des artefacts techniques (Buhay, 2009; Song *et al.*, 2008). Cependant, d'après (i) leur séquence d'acides aminés, (ii) leur fréquence relative et (iii) leur positionnement au sein du réseau d'haplotypes, nous pensons avoir pu distinguer les « vrais » haplotypes COI et les haplotypes « COI-like » obtenus à partir de pseudogènes ou d'artefacts techniques.

Notre estimation de la proportion d'haplotypes « COI-like » varie entre 22% (lorsque seuls les haplotypes « non affiliés » et « varprot » étaient pris en compte) et 61% (lorsque tous les haplotypes et haplotypes « varprot » des catégories A et B ont été inclus). Une proportion aussi élevée peut sembler alarmante, mais il faut garder à l'esprit que les haplotypes associés sont le plus souvent rares ou très rares et correspondent à une petite proportion d'individus. Sur l'ensemble de données, 1,6% des individus appartiennent à la catégorie A, et 4,9% à la catégorie B.

Un intérêt majeur des données générées est un premier appariement explicite entre les informations moléculaires et morphologiques de dix espèces paléarctiques occidentales (indigènes ou établies) : *T. achaeae*, *T. bourarachae*, *T. cacoeciae* Marchal, *T. cordubensis*, *T. daumalae*. Dugast et Voegelé, *T. dendrolimi*, *T. embryophagum* Hartig, *T. oleae*, *T. principium* et *T. semblidis*. Cet appariement doit encore être utilisé avec prudence puisque seulement sept des clusters / haplotypes isolés ont été documentés par au moins deux sources indépendantes (Tableau 7). De plus, quatre conflits plus ou moins aigus ont été observés.

Parmi ceux-ci figurait l'observation de deux souches hautement divergentes d'un *T. exiguum* putatif pour lesquelles la description disponible devrait être utile (Pinto *et al.*, 1983). D'autres conflits s'expliquent probablement par des complexes d'espèces tels qu'ils peuvent être observés chez les hyménoptères (Al Khatib *et al.*, 2014; Gebiola *et al.*, 2017; Mottern and Heraty, 2014; Rugman-Jones *et al.*, 2009). En particulier, la présence de souches identifiées comme *T. cacoeciae* et *T. embryophagum* au sein d'un même groupe est logique car les femelles semblent être morphologiquement proches, ces deux supposées "espèces" semblant principalement différer par leurs déterminismes du sexe, *T. cacoeciae* et *T. embryophagum* étant respectivement décrits à l'origine comme thélytoque et arrhénotoque (Pintureau, 1997; Vavre *et al.*, 2004). De même, les proximités moléculaires d'une part, de *T. oleae* et de *T. pretiosum* et, d'autre part, de *T. brassicae* Bezdenko et de *T. euproctidis* (Girault) correspondent bien à leurs affiliations respectives aux groupes "pretiosum" et "evanescens" définis par B. Pintureau sur la base de l'examen des caractères morphologiques (Pintureau, 2008). Le plus surprenant est la présence d'un cluster contenant *T. daumalae* et *T. evanescens*. En effet, sur la base de caractères morphologiques, ils étaient auparavant affiliés à deux groupes différents, respectivement « pretiosum » et « evanescens » (Pintureau, 2008). Dans ce cas particulier, nous ne pouvons cependant pas exclure que notre souche de référence pour *T. daumalae* soit erronée ; aussi, des investigations morphologiques sont actuellement menées pour vérifier cette possibilité. Quels que soient les véritables statuts taxonomiques des souches actuellement englobées dans les complexes « brassicae-euproctidis » et « daumalae-evanescens », les expériences de croisement ont mis en évidence des patrons contrastés de compatibilités reproductives. En particulier, des niveaux très variables de (in)compatibilités reproductives ont été observés dans le complexe "brassicae-euproctidis" comme déjà décrit par Pintureau et Babault (1981). Ce schéma ne peut pas être facilement expliqué ni par des divergences moléculaires et / ou par des isolements géographiques comme observé par exemple pour *T. chilonis* (Benvenuto *et al.*, 2012) et, ni par des infections à *Wolbachia* (cf. « Matériels et Méthodes » - § 2.3), l'induction d'incompatibilité cytoplasmique induite étant peu probable. Cette variabilité des compatibilités reproductives évoque donc plutôt les profils observés dans des croisements intraspécifiques (Pinto *et al.*, 1991), et dans le complexe « minutum-platneri » de *Trichogramma* (Pinto *et al.*, 1991; Stouthamer *et al.*, 2000). Enfin, cette étude a mis en évidence au moins quatre groupes (misA, misB, swC et swE) qui constituent probablement quatre espèces. Malheureusement, une caractérisation moléculaire complémentaire utilisant ITS2 - un marqueur moléculaire très variable utilisé dans plusieurs études (Almeida and Stouthamer, 2015; Ávila-Rodríguez *et al.*, 2013; Liu *et al.*, 2017;

Polaszek *et al.*, 2012; Poorjavad *et al.*, 2012; Sumer *et al.*, 2013) - n'a pas mis en évidence de similarités suffisantes avec des souches identifiées documentées dans Genbank. Leurs identifications devront donc s'appuyer sur la future caractérisation moléculaire d'espèces similaires et déjà identifiées et / ou sur une expertise renouvelée sur les caractères morphologiques.

Pris dans son ensemble, cette étude constitue une première contribution à la taxonomie intégrative des espèces du genre *Trichogramma* dans la zone Ouest-Paléarctique. Ces informations peuvent toutefois d'ores et déjà être utiles à diverses fins (recherche et développement en biocontrôle, recherche fondamentale ou questions de régulation). L'objectif suivant est d'étendre l'analyse moléculaire à d'autres marqueurs (en particulier nucléaires) pour mieux démêler les complexes d'espèces et fournir une phylogénie plus résolue (Al khatib *et al.*, 2016). Une fois obtenue, une telle phylogénie serait pertinente pour cartographier les caractères morphologiques traditionnellement utilisés pour la taxonomie mais également certaines caractéristiques phénotypiques. Pour être aussi fructueux et exhaustif que possible, ce travail s'appuie bien entendu sur le soutien des communautés scientifiques traitant du genre *Trichogramma* dans la région Ouest-Paléarctique afin de fournir un matériel biologique pertinent et une expertise basée sur la morphologie.

4 VERS UNE PHYLOGENIE MULTILOCUS DU GENRE *TRICHOGRAMMA*

4.1 Introduction

Comme évoqué précédemment, le genre *Trichogramma* regroupe plus de 200 espèces dont plus de 40 en Europe. Nous avons vu dans le Chapitre précédent comment le séquençage d'une portion de la *COI* (§ 3 « Barcoding ») peut permettre de caractériser précisément et rapidement la biodiversité. Ainsi 24 « entités » ont pu être identifiées sur la base d'un fragment court de *COI*.

D'une façon plus générale, l'utilisation d'une portion même longue d'un seul gène d'ADN en particulier mitochondrial, reste toutefois évidemment insuffisante pour retracer l'histoire évolutive des espèces. Plusieurs études montrent en effet une discordance entre l'histoire évolutive d'une espèce et la phylogénie reconstruite à partir du seul gène mitochondrial *COI* utilisé en barcoding (Dupuis *et al.*, 2012). En fait, seule une démarche de phylogénie multilocus peut permettre de documenter à la fois les relations phylogénétiques anciennes et plus récentes en choisissant des gènes évoluant de façon contrastée (Cruaud *et al.*, 2010; Danforth *et al.*, 2013). Dans ce contexte, nous avons sélectionné six marqueurs moléculaires, *COI*, *CytB*, *Wingless*, *EF1a*, *Rpl27a* et *RpS4* (cf. § 2.2.2 choix des marqueurs) qui permettent chacun de résoudre les relations phylogénétiques à des niveaux différents. Certains, comme la *COI* et le *CytB* permettent de déterminer les relations au niveau des clades terminaux et donc de décrire une évolution récente. D'autres permettent de récupérer des informations au niveau des nœuds profonds des ancêtres communs notamment *EF1a-F2* et *Rpl27a*. *Wingless* et *RpS4* permettent de résoudre des relations « intermédiaires ». Ces six marqueurs ont déjà été utilisés conjointement ou en partie avec succès dans des analyses phylogénétiques sur (i) des espèces du genre *Eupelmus* (Hymenoptera : Eupelmidae) (Al khatib *et al.* 2015 ; Al khatib *et al.* 2016) et (ii) des Agaonidae (Hymenoptera : Chalcidoidea), pollinisateurs des *Ficus* (Cruaud-2009). Ces résultats sont discutés en relation avec trois thématiques à savoir (i) la classification des *Trichogramma* en sections et groupes à partir de caractères morphologiques, (ii) la révision apparemment nécessaire de certains clusters et (iii) l'infection par *Wolbachia* au sein de ce genre.

4.2 Résultats

Les résultats présentés dans ce chapitre n'ont pour l'instant pas fait l'objet de communications au niveau international. Un article est actuellement en cours de rédaction.

4.2.1 Analyse séparée des marqueurs

Sur 321 individus analysés, 4 correspondent à des out-groups (2 *Eupelmus urozonus*, 1 *Anastatus temporalis*, 1 *Anastatus sidereus*) et 317 appartiennent au genre *Trichogramma*. Les succès d'amplification sur chacun des 6 gènes *COI*, *CytB*, *F2-EF1alpha*, *Rpl27a*, *RpS4*, *Wingless* sont respectivement de 91%, 82%, 87%, 69%, 73%, et 79%. Les échecs sont principalement dus, à de l'amplification non spécifique avec présence de bandes multiples non séquençables, notamment pour le gène *Rpl27a* pour lequel il serait nécessaire de modifier l'amorce pour une meilleure efficacité.

De manière générale, deux points principaux ressortent lorsque l'on compare les différents marqueurs. **D'une part, les phylogénies obtenues à partir de *F2-EF1alpha* et *Wingless*** (Figures supplémentaires S4 et S5) **s'avèrent peu informatives relativement aux autres marqueurs**. Les pourcentages de sites variables (respectivement 22% et 35%) et de sites informatifs (20% et 32%) de ces deux marqueurs sont les plus faibles (Tableau 4). Plus précisément sur l'arbre *F2-EF1alpha* (Figure S4), la majorité des individus des groupes VI-brassicæ-euproctidis (polyphylétique) et des individus des groupes I-evanescens/daumalæ, III-cordubensis, , IV-chilonis et d-not_affiliated, sont regroupés dans un cluster (BP = 32) comprenant lui-même un cluster monophylétique (BP=53) du groupe *XII-cacoecia/embryophagum*. Sur l'arbre *Wingless* (Figure S5), des individus des groupes I-evanescens/daumalæ, VI-brassicæ-euproctidis, II-cacoecia-embryophagum, d-notaffiliated se regroupent au niveau basal d'une même branche (BP<25). Egalement sur *Wingless* (Figure S5), les groupes VII-misA, V-misB, 3-swA et 4-nerudai sont regroupés dans un cluster (BP=60) sans structuration à l'intérieur de ce cluster. **D'autre part, les analyses gène par gène n'ont pas montré de conflits majeurs (impliquant plusieurs individus) entre les différentes topologies**. En particulier, aucun des 14 clusters (I-XIV) identifiés en « barcoding *COI* » (cf. § 3) ne s'avère structuré en groupes polyphylétiques sur d'autres marqueurs. Les 8 haplotypes isolés (2 à 9) par *COI* sont en général retrouvés isolés, selon la capacité de discrimination du marqueur.

Il est à noter également que des groupes supplémentaires (nommés a-swI, b-swK, c-swJ, d-swB) bien supportés ont été obtenus à partir d'échantillonnages récents sur les phylogénies *COI*, *CytB*, *RpS4*, (Figures supplémentaires S6, S7, S8).

4.2.2 Analyse multilocus

Comme présenté dans le chapitre 2 – Matériels et Méthodes, différentes approches ont été déployées pour inférer les relations phylogénétiques entre nos espèces.

4.2.2.1 Comparaison des schémas de partitions :

Concernant les deux gènes mitochondriaux (*COI* et *Cytb*), 2 schémas de partition ont été comparés, l'un avec deux partitions (Figure S9) et l'autre avec 6 partitions définies par PartitionFinder2 (Figure S10). Globalement, la structure des branches terminales est conservée mais des différences jugées mineures ont été observées. Ainsi, on peut noter que l'arbre obtenu à partir 6 partitions (Figure S10) :

- diminue le « bruit » au moins dans quatre phylogroupes : (i) le groupement des haplotypes 196, 197 et 198 du cluster « XII cacoeciae-embryophagum » ; (ii) les haplotypes 159 et 108 du cluster « XII cacoeciae-embryophagum » ; (iii) le cluster « V misB » ; (iv) le cluster « IX swE » ;
- supporte moins bien certains groupes : III cordubensis, II daumalae-evanescens, c-swJ ;
- positionne moins bien le groupe VI brassicae-euproctidis à la racine des *Trichogramma* ;
- éloigne l'individu 33142 de l'espèce c-swJ auquel les autres individus du même échantillon sont affiliés.

Sur cette base, l'arbre obtenu à partir du schéma avec deux partitions (Figure S9) a été privilégié par la suite.

Concernant les quatre gènes nucléaires, 2 schémas de partition ont également été comparés, l'un avec quatre partitions (Figure S11), l'autre avec 7 partitions définies par PartitionFinder2 (Figure S12) ont été comparés. Globalement, la structure des branches terminales est conservée et la majorité des groupes présente une structure similaire entre les deux représentations. On peut toutefois noter que, dans l'arbre édité avec 7 partitions (Figure S12) :

- le groupe IX swE est relativement moins bien supporté
- à l'inverse, 3 autres groupes semblent mieux supportés : XII-cacoeciae-embryophagum, XIII-swC et c-swJ.

Sur cette base, l'arbre obtenu à partir du schéma avec deux partitions (Figure S8) a été privilégié par la suite.

Nous avons donc ensuite comparé la topologie obtenue avec les marqueurs mitochondriaux (Figure S9) et celle obtenue avec les marqueurs nucléaires (Figure S11). Globalement, les topologies sont concordantes et seul un conflit jugé « mineur » a été observé concernant les haplotypes Hap159 et Hap108. Plus précisément, ces deux haplotypes sont proches du cluster « XII *cacoeciae/embryophagum* » mais sont positionnés dans un groupe non monophylétique sur les marqueurs nucléaires (voir * sur Figure S9 et sur Figure S11).

Compte tenu de ces résultats, nous avons ensuite procédé à la concaténation de l'ensemble des gènes (*COI*, *CytB*, *EF1alpha-F2*, *Rpl27a*, *RpS4*, *Wingless*). Deux schémas de partition ont été comparés :

- Celui à dire d'expert proposé par A. Cruaud prenant en compte cinq partitions, une pour les deux gènes mitochondriaux et une pour chacun des gènes nucléaires (cf. Figure 7) ;
- Celui retenu par PartitionFinder2 prenant en compte 12 partitions (cf. partitions Tableau 5) (cf. Figure S14).

Les deux topologies se sont révélées concordantes et celle à cinq partitions a été retenue pour la suite des analyses.

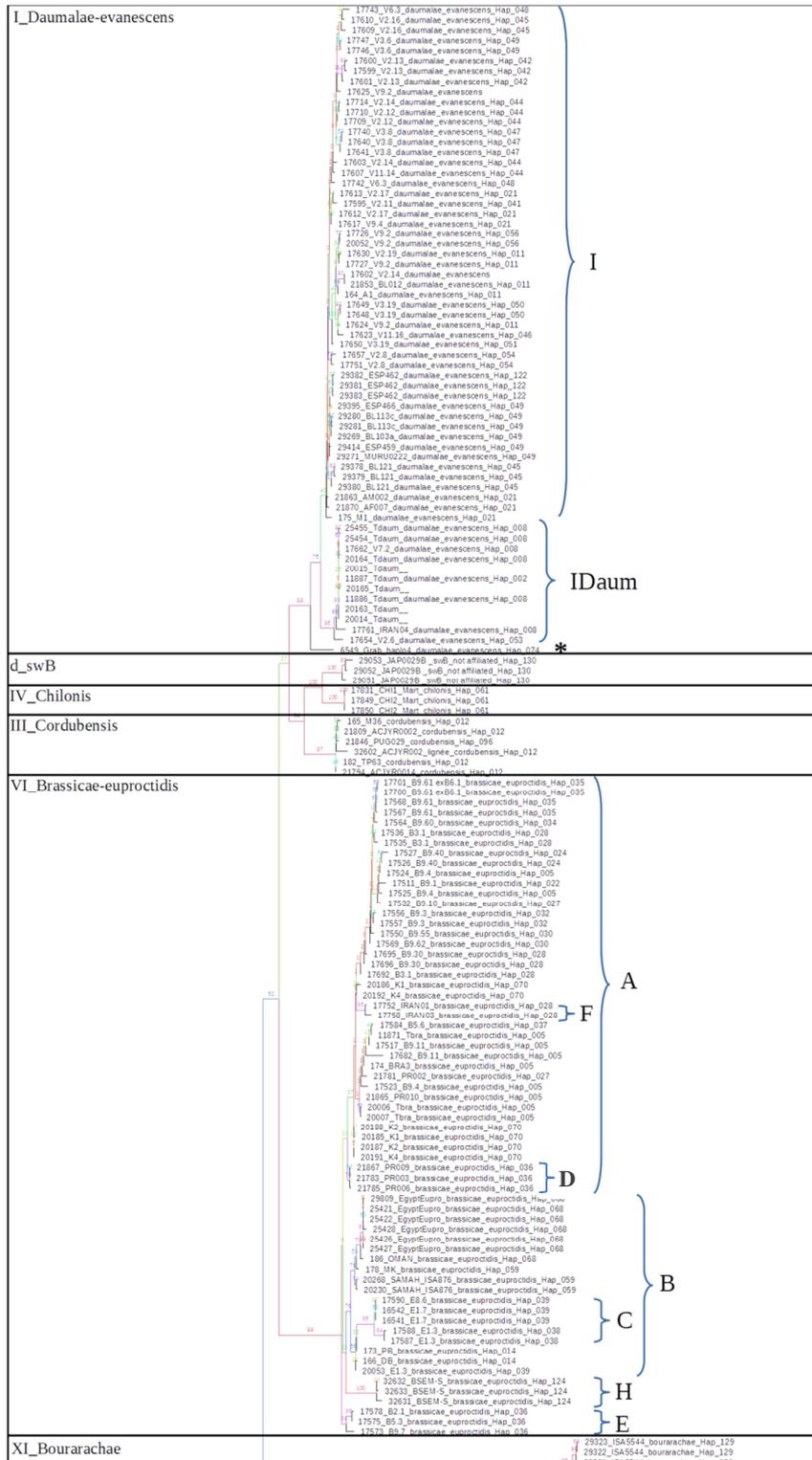


Figure 7 (début) : Arbre ML sur la concaténation de 6 gènes *COI*, *CytB*, *EF1alpha-F2*, *RpL27a*, *RpS4*, *Wingless* avec 5 partitions

Arbre ML basé sur les données de *Trichogramma* obtenu à partir de l'alignement ClustalW de l'ensemble de données *COI*, *CytB*, *EF1alpha-F2*, *RpL27a*, *RpS4*, *Wingless* combinées sous le schéma de 5 partitions. Les arbres ML correspondant par gènes sont donnés sur les Figures S4, S5, S6, S7, S8, S13. Les supports BP (100 répliqués) des clades sont indiqués aux nœuds en pourcentage. Les groupes renvoient aux clades discutés dans le §4-Multilocus voir aussi §3-Barcoding Figure 4.

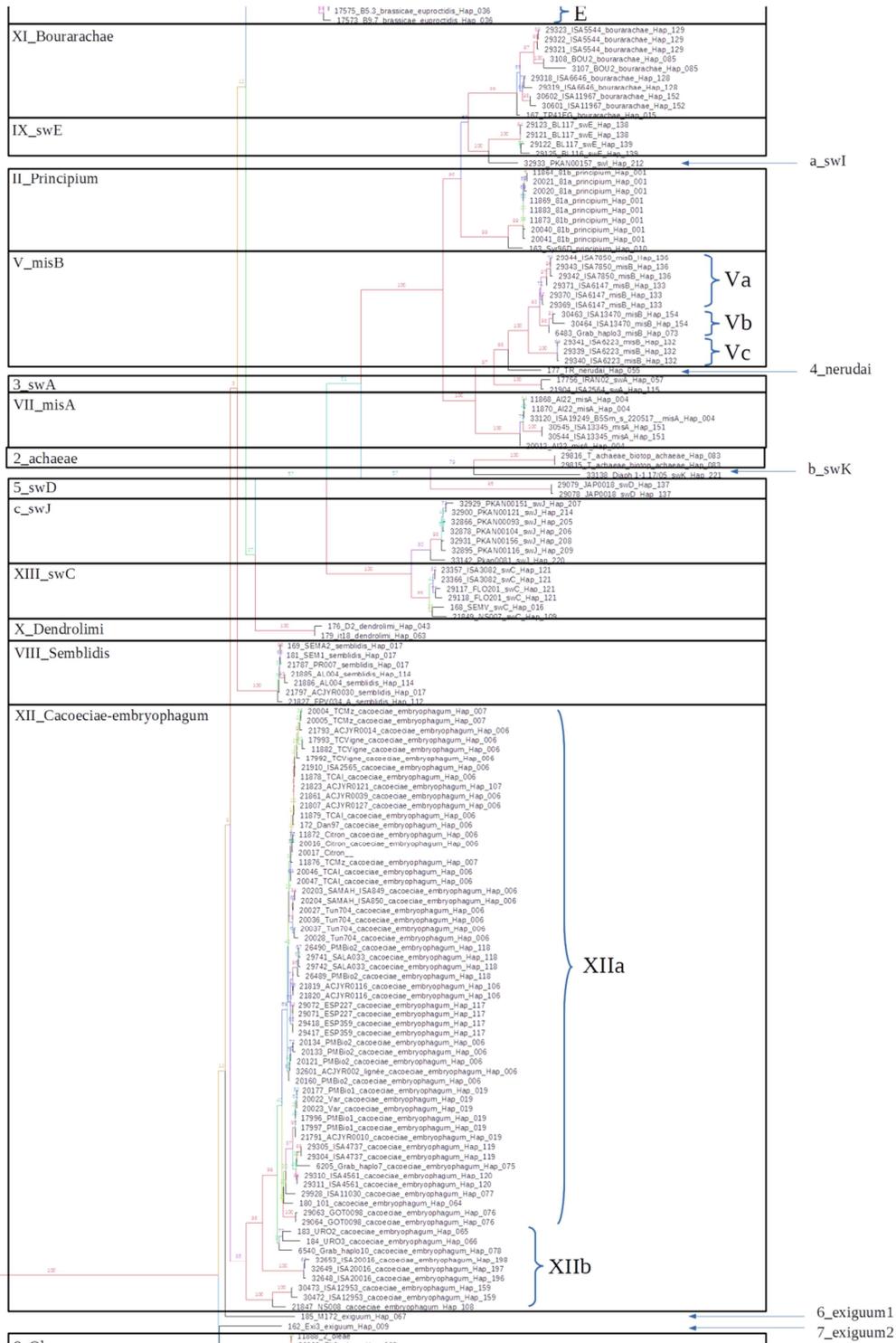


Figure 7 (suite) : Arbre ML sur la concaténation de 6 gènes *COI*, *CytB*, *EF1alpha-F2*, *RpL27a*, *RpS4*, *Wingless* avec 5 partitions

Arbre ML basé sur les données de *Trichogramma* obtenu à partir de l'alignement ClustalW de l'ensemble de données *COI*, *CytB*, *EF1alpha-F2*, *RpL27a*, *RpS4*, *Wingless* combinées sous le schéma de 5 partitions. Les arbres ML correspondant par gènes sont donnés sur les Figures S4, S5, S6, S7, S8, S13. Les supports BP (100 répliqués) des clades sont indiqués aux nœuds en pourcentage. Les groupes renvoient aux clades discutés dans le §4-Multilocus voir aussi §3-Barcoding Figure 4.

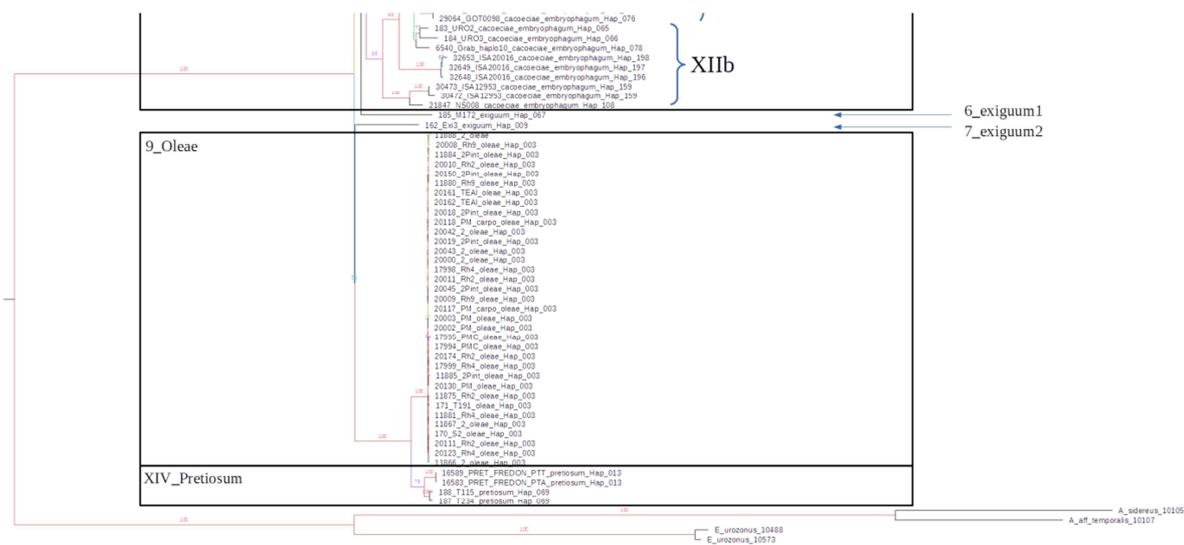


Figure 7 (fin) : Arbre ML sur la concaténation de 6 gènes *COI*, *CytB*, *EF1alpha-F2*, *RpL27a*, *RpS4*, *Wingless* avec 5 partitions

Arbre ML basé sur les données de *Trichogramma* obtenu à partir de l'alignement ClustalW de l'ensemble de données *COI*, *CytB*, *EF1alpha-F2*, *RpL27a*, *RpS4*, *Wingless* combinées sous le schéma de 5 partitions. Les arbres ML correspondant par gènes sont donnés sur les Figures S4, S5, S6, S7, S8, S13. Les supports BP (100 répliqués) des clades sont indiqués aux nœuds en pourcentage. Les groupes renvoient aux clades discutés dans le §4-Multilocus voir aussi §3-Barcoding Figure 4.

4.2.2.2 Structuration :

La longueur de l'alignement des données concaténées est de 3409pb avec des longueurs de partition suivantes : 1065pb pour l'ADN mitochondrial (*COI* et *CytB*), 540pb pour *F2-Ef1alpha*, 427pb pour *Rpl27a*, 943pb pour *RpS4*, et 434pb pour *Wingless*. Sur les 321 séquences concaténées analysées, 3% des individus sont représentés sur un seul gène, 3% sur deux gènes, 11% sur 3 gènes, 20% sur 4 gènes, 22% sur 5 gènes et 41% sur l'ensemble des 6 gènes.

La phylogénie présentée en Figure 7 présente une structuration similaire à celle obtenue avec *COI* seulement, avec généralement, comme on pouvait l'espérer, de meilleurs soutiens de bootstrap (BP). Les 14 clusters initialement notés de I à XIV (Figure 4) ainsi que les 8 clusters ou séquences isolées notés 2 à 9 (Figure 4) sont ainsi retrouvés auxquels s'ajoutent 4 nouveaux clusters ou séquences isolées (notées a, b, c et d dans la Figure 7) qui correspondent à des échantillons récents (cf. 4.2.2.2 de cette section). Les BP des branches basales restent cependant peu soutenues (BP<65) tandis que les branches terminales sont généralement très bien supportées (BP>94%). Parmi les clusters identifiés, des sous-structurations sont visibles à l'intérieur de plusieurs d'entre eux et présentées ci-dessous par ordre décroissant, selon nous, en termes d'intérêt scientifique.

4.2.2.2.1 Cas d'étude « cacoeciae-embryophagum »

Comme indiqué sur la **Figure 7** (5 partitions), le cluster monophylétique XII *cacoeciae-embryophagum* est sous-structuré avec un cluster XIIIa en branche terminale (BP=40) –et un ensemble basal appelé XIIIb incluant quelques clusters (topologie confirmée sur analyse avec 12 partitions - voir Figure S14 en Supplément). Le réseau d'haplotypes (Figure 8) obtenu à partir de ces données confirme également la plus grande proximité des haplotypes du cluster XIIIa entre eux, à l'exception de deux haplotypes plus isolés, l'haplotype 065 et l'haplotype 066 correspondant respectivement aux souches URO2 et URO3 récoltées tous deux en Iran. Du point de vue du déterminisme du sexe, le cluster XIIIa rassemble la totalité des individus appartenant à des souches thélytoques. Cette thélytoquie semble induite par le génotype de l'individu puisque la recherche de *Wolbachia* (*ftsZ*) réalisée sur 146 individus ne montre pas d'infections.

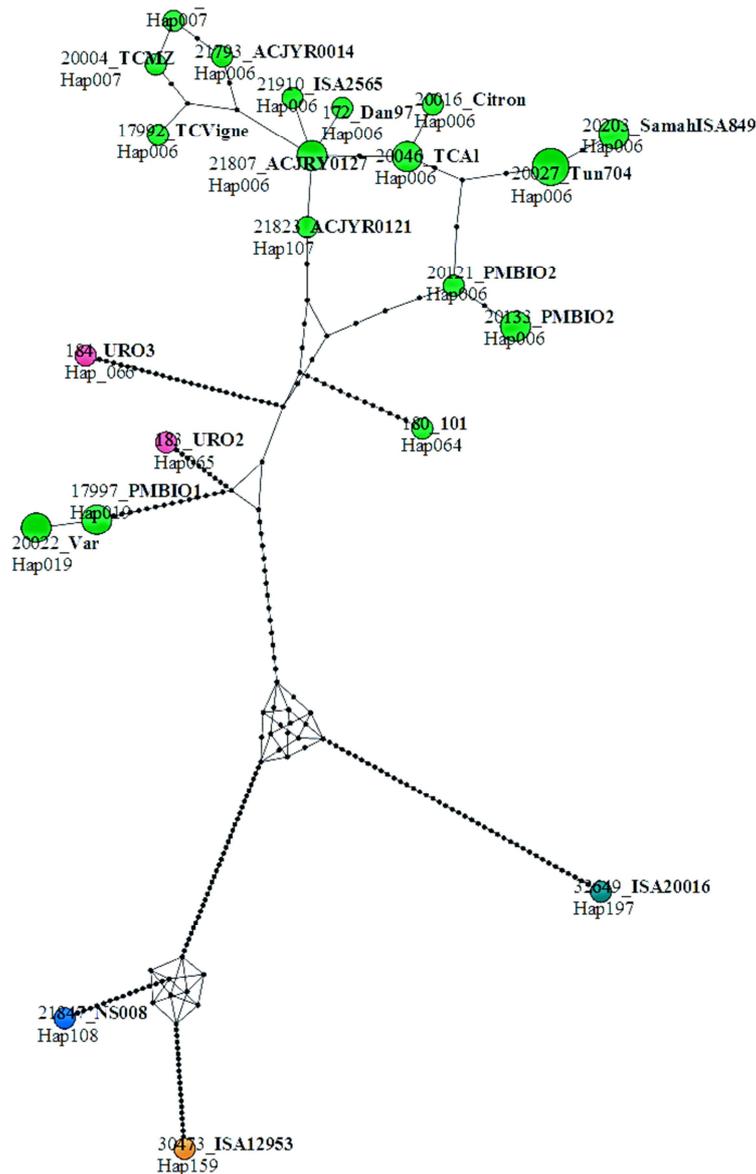


Figure 8 : Réseau d'haplotypes du complexe « cacoeciae-embryophagum » à partir de PopARTs
 Le réseau a été déduit en utilisant la méthode Median Joining avec epsilon = 0 (paramètre par défaut).
 La taille du cercle correspond à la fréquence de l'haplotype. Les distances moléculaires sont représentées par des ronds noirs sur les branches du réseau.

Les individus du groupe XIIb sont structurés en sous-clusters, très éloignés les uns des autres dans le réseau d'haplotypes et qui présentent par ailleurs des modes de reproduction et des statuts d'infection à *Wolbachia* voire des écologies contrastés (cf. Tableau 8). Ainsi :

- le sous-groupe (i) représenté par les Hap065, Hap066 et Hap078 inclut deux souches (URO2 : représentant les haplotypes Hap065 et URO3 : Hap066) anciennement identifiées comme *T.embryophagum* par B. Pintureau sur des caractères morphologiques. Toutes deux sont infectées par *Wolbachia* avec toutefois des modes de reproduction différents, sexués pour l'une (URO2), et asexués (URO3) pour l'autre ;

- le sous-groupe (ii) représenté seulement par l'haplotype Hap108 correspond à une souche (NS008) thélytoque, cette thélytoquie étant induite par *Wolbachia* puisque la production de mâle est rétablie après antibiothérapie ;
- le sous-groupe (iii) représenté uniquement par Hap159 correspond à des individus (ISA12953) apparemment infectés par *Wolbachia* mais dont le mode de reproduction n'a pas pu être déterminé (pas de souches vivantes) ;
- le sous-groupe (iv) représenté par les haplotypes Hap196, Hap197 et Hap198. Ce cluster regroupe seulement des souches sexuées algériennes récoltées sur *Thaumetopoea pityocampa*, la processionnaire du pin. Ces souches sont *a priori* non infectées par *Wolbachia* (analyses complémentaires en cours au LBBE).

L'ensemble de ces résultats nous incitent à remettre en cause les délimitations d'espèces dans ce complexe (cf. Discussion).

Tableau 8 : Infection à *Wolbachia* du sous cluster XIIb à l'intérieur du complexe cacoeciae-embryophaguum. Infection à *Wolbachia* prenant en compte l'ensemble des individus appartenant à ce cluster sur la base de leur détermination COI.

Souches	Haplotypes COI	Origine (Pays ; Donateur)	Recherche <i>Wolbachia</i>		Sexe	Sexe après traitement antibiotique	Plantes/insectes Hôtes
			Présence <i>Wolbachia</i>	Absence <i>Wolbachia</i>			
NS008	Hap_108	France ; ANR TRIPTIC	8	0	F	F + M	Poirier / (œufs sentinelles <i>E.khueniella</i>)
URO2	Hap_065	Iran ; Collection historique	1	0	F + M	/	Pommier / <i>Laspeyresia pomonella</i>
URO3	Hap_066	Iran ; Collection historique	1	0	F	/	Pommier / <i>Laspeyresia pomonella</i>
ISA12953	Hap_159	France ; ANR TRIPTIC	2 ?	0	?	/	<i>Erica arborea</i> / / (œufs sentinelles <i>E.khueniella</i>)
P_naila (31souches)	Hap_196/197/198	Algérie ; Autres sources	2 ?	90	F + M	/	Pin / <i>Thaumetopoea pityocampa</i>

4.2.2.2.2 Cas d'étude « brassicae/euproctidis »

Les analyses sur les 6 gènes concaténés (Figure 7) révèlent une structuration fortement soutenue (BP : 98-99%) à l'intérieur de ce cluster monophylétique regroupant initialement des individus identifiés comme *T. brassicae* et *T. euproctidis*.

On distingue notamment 2 groupes principaux :

- le groupe A incluant la souche de référence pour *T. brassicae* (souche : Tbra – Haplotype 005) qui apparaît largement distribuée : Europe (de l'Ouest vers l'Est : France, Croatie, Moldavie), Turquie. Dans cette analyse, le groupe A inclut également un sous-groupe F contenant uniquement des individus en provenance d'Iran.

- Le groupe B incluant la souche de référence pour *T. euproctidis* (souche : Egypteupro – Haplotype 068) a une distribution globalement plus méridionale : Sud de la France, Portugal, Afrique du Nord (de l'Ouest vers l'Est : Maroc, Tunisie, Egypte), Péninsule Arabique. Ce groupe inclut également un sous-groupe C correspondant à des individus observés en France. Il n'est pas exclu que ce sous-groupe soit un artefact dans la mesure où, d'une part, l'haplotype 039 se retrouve dans (individus 17590, 16541 et 16542) et hors (20053) du cluster C et, d'autre part, une même souche (E1.3) présente également des individus dans et hors de ce sous-groupe.

Trois groupes mineurs sont également mis en évidence, à savoir les groupes D (retrouvé en Europe et Iran), le groupe E (France) et le groupe H (une seule localité en France). A noter que ce dernier groupe H présente la particularité d'être infecté par *Wolbachia*, contrairement à tous les autres.

Cette structuration se retrouve également globalement sur le réseau d'haplotypes correspondant (Figure 9).

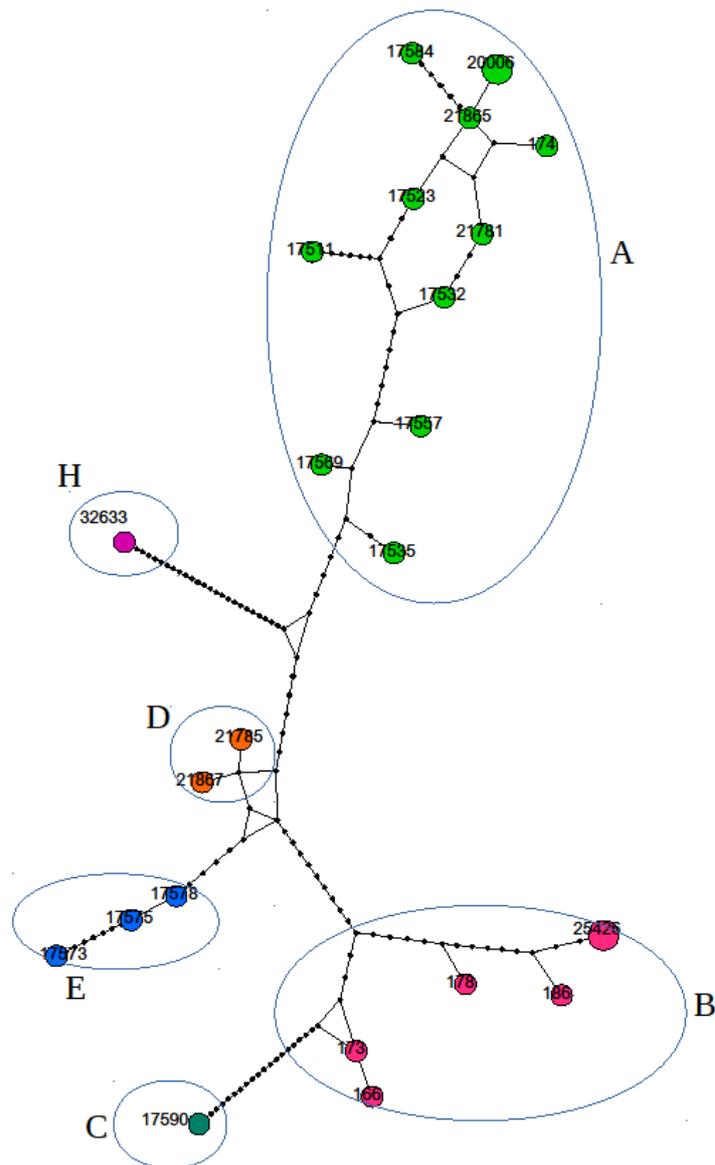


Figure 9 : Réseau d'haplotypes du complexe « brassicae-euproctidis » à partir de PopARTs

Le réseau a été déduit en utilisant la méthode Median Joining avec $\epsilon = 0$ (paramètre par défaut).

Chaque ellipse correspond à un sous-cluster moléculaire tel que défini dans les résultats. La taille du cercle correspond à sa fréquence de l'haplotype. Les distances moléculaires sont représentées par des ronds noirs sur les branches du réseau.

Faute de souches vivantes disponibles, les croisements déjà présentés dans le Chapitre 3 n'explorent que partiellement les compatibilités de reproduction au sein de ce cluster VI-brassicae-euproctidis puisqu'*a posteriori* seuls des représentants des groupes A, B et D ont pu être testés. Si l'ensemble apparaît globalement comme interfécond, les incompatibilités partielles ou totales sont quand même observées entre les groupes A et B.

En l'état, nos résultats ne permettent pas de trancher concernant la nature de la structuration au sein de ce complexe, notamment entre variabilité intraspécifique avec possible isolement par la distance et spéciation récente.

4.2.2.2.3 Cas d'étude « cluster I-daumalae-evanescens »

L'analyse phylogénétique multilocus du phylogroupe I-daumalae-evanescens (Figure 7) a mis en évidence la structuration interne suivante :

- le sous-groupe « I » incluant des individus identifiés, sur des bases morphologiques (genitalia et antennes des mâles, Figure 1), comme *T. evanescens* par B. Pintureau ou J. Seguret ;
- le sous-cluster « Idaum » incluant certains individus identifiés comme *T. daumalae* par B. Pintureau et d'autres comme *T. evanescens* par J. Séguret ;
- une séquence isolée (* sur Figure 7) d'un individu issu du terrain (pas de souches conservées) et non identifié sur des bases morphologiques.

Comme précédemment indiqué (cf. Chapitre 3.2.4), les souches testées des deux sous-clusters I et Idaum s'avèrent compatibles en donnant des filles mais également des petites-filles à la génération suivante.

Nos résultats penchent donc en faveur de l'hypothèse que ce phylogroupe ne constitue qu'une seule espèce avec toutefois une variabilité moléculaire intraspécifique importante ($0.004 < d < 0.046$, cf. Figure 5).

4.2.2.2.4 Cas d'étude cluster V-misB :

Le cluster V monophylétique (Figure 7) est structuré avec 2 sous-clusters (BP>99%), un premier [Va-Vb] lui-même restructuré en deux (Va et Vb) et un deuxième Vc. Les individus issus du groupe Vc présentent systématiquement des variations de leurs séquences d'acides aminés par rapport aux individus du groupe [Va-Vb] sur leurs séquences mitochondriales *COI* et *CytB* (groupe Vc avec un BP100 sur Figure S9), les séquences d'acides aminés des autres gènes (*Wingless* et *F2-Eflalpha*) codants étant conservées.

Les croisements réalisés entre ISA14313 (haplotype 155 – sous-cluster Va-Vb) et ISA15550 (haplotype 132 – sous-cluster Vc) se sont par contre avérés particulièrement compatibles avec un RC1 moyen de 100% et un RC2 moyen de 84% (n = 16, deux sens de croisement inclus), les femelles vierges témoin n'étant pas capables de produire de filles (n = 11, toutes souches

confondues). Les filles issues de la génération 1 ont été capables de produire à leur tour des filles.

En conclusion, nous considérons que les souches / haplotypes de ce cluster ne constituent qu'une seule espèce et que les séquences mitochondriales divergentes (Vc) sont des pseudogènes. Des situations similaires ont été observées dans notre laboratoire pour d'autres taxons (Ris, comm. pers.). Une perspective technique à court terme serait de définir de nouvelles amorces pour amplifier le gène correct.

4.2.2.3 Infections à *Wolbachia*

Comme indiqué dans le Tableau 9, le genre *Wolbachia* a été retrouvé sans ambiguïté dans 8 clusters : « II – principium » ; « III - cordubensis », « V - misB », « VI - brassicae-euproctidis », « XII – cacoeciae-embryophagum », XIV « pretiosum », « 9 - oleae » et « b - swK ». La présence de *Wolbachia* n'a pas été observée par PCR simple, voire pour 9 d'entre eux par qPCR, dans les 9 clusters suivants : « IV – chilonis », « VII – misA », « VIII – semblidis », « IX – swE », « XIII – swC », « 2 – achaeae », « 5 – swD », « a –swI » et « c – swJ ». L'infection à *Wolbachia* est encore incertaine pour les 3 clusters, « XI – bourarachae », « 3 – swA » et « b- swK » pour lesquels le signal PCR FtsZ est difficile à interpréter (analyses complémentaires en cours par qPCR voire du MLST).

Plus précisément (Tableau 9), si l'infection semble généralisée à l'intérieur de certains clusters (II – principium » ; « III - cordubensis », « XIV - pretiosum », « 9 - oleae »), elle semble rare dans d'autres. En particulier, seule une souche très particulière est infectée à l'intérieur du cluster « XII cacoeciae-embryophagum » de même qu'à l'intérieur du cluster « brassicae-euproctidis » (cf. § précédent). A l'intérieur du complexe « II daumalae-evanescens », la recherche de *Wolbachia* par qPCR a permis de révéler une infection très rare (1 individu sur 81).

Certains individus infectés par *Wolbachia* ont fait l'objet d'une caractérisation par MLST directement sur produits PCR (Supplément S15). Les résultats mettent en évidence une variabilité notable au sein du clade B, un des deux super-groupes de *Wolbachia* les plus représentés dans les affections des arthropodes.

Tableau 9 : Répartition des infections *Wolbachia* en fonction des espèces à partir des résultats obtenus en PCR FtsZ, en qPCR et en MLST.

Espèces /cluster	Total Individus PCR FtsZ	Total Individus qPCR	Total Individus MLST	Pourcentage W-	Pourcentage W+	Pourcentage W(?)	Infection Wolbachia
achaeae	14	4		100	0	0	Aucune
bourarachae	6	4	0	90	0	10 (FtsZ)	Aucune ou partielle
brassicae_euproctidis	100	32	0	99	1 (qPCR)	0	Très faible
<i>brassicae_euproctidis_Hap_124</i>	53		3	2 (FtsZ)	98	0	Elevée
cacoeciae_embryophagum	238	24	1	92	4	4 (FtsZ)	Partielle (voir Tableau 8)
chilonis	13	8		100	0	0	Aucune
cordubensis	50	4	1	4 (FtsZ)	96	0	Elevée
daumalae_evanescens	53	28		99	1 (qPCR)	0	Très faible
misA	11	4		100	0	0	Aucune
misB	23	8		52	45	3	Partielle
oleae	28		3	13	87	0	Partielle
pretiosum	55	5	2	24	76	0	Partielle
principium	24	4		89	11(qPCR)	0	Partielle
semblidis	12	12		100	0	0	Aucune
swA	5	4		56	0	44 (FtsZ)	Indéterminée
swC	7			100	0	0	Aucune
swD	5	4		100	0	0	Aucune
swE	3			100	0	0	Aucune
swI	17			100	0	0	Aucune
swJ	16			100	0	0	Aucune
swK	2			0	0	100	Indéterminée

4.3 Discussion

4.3.1 Confrontation entre structuration moléculaire et classification moléculaire

La phylogénie multilocus réalisée à partir de 2 gènes mitochondriaux et 4 gènes nucléaires a permis d'obtenir un arbre avec de fortes valeurs de bootstrap au niveau des taxons terminaux. Les clusters et haplotypes initialement définis en *COI* ont ainsi été confirmés auxquels s'ajoutent des nouveaux groupes non identifiés sur des bases morphologiques. Au niveau plus basal, certaines relations phylogénétiques demeurent toutefois encore imprécises. Cette première topologie peut toutefois être déjà confrontée à la classification en « sections » et « groupes » proposée par B. Pintureau sur la base de caractères morphologiques (Pintureau, 2008). Notre échantillonnage comprend essentiellement des espèces des sections « exiguum » (*T. brassicae*, *T. cacoeciae*, *T. cordubensis*, *T. daumalae*, *T. dendrolimi*, *T. embryophagum*, *T. euproctidis*, *T. evanescens*, *T. oleae*, *T. pretiosum* et *T. semblidis*) et « parkeri et thalense » (*T. bourarachae* et *T. principium* voire le groupe c que l'on suppose être *T. gicai* (Groussier G., comm. pers.)). Si les espèces de la section « parkeri et thalense » semblent former avec d'autres clusters non identifiés, un ensemble monophylétique, cela ne semble pas être le cas pour les représentants de la section « exiguum ». Si l'on se place à l'échelle des groupes définis à l'intérieur de la section « exiguum », le groupe « evanescens » (représenté par nos complexes « daumalae-evanescens » et « brassicae-euproctidis ») ne semble pas monophylétique du fait de la présence intercalée de *T. cordubensis* (groupe « minutum »). Le groupe « pretiosum » (représenté par *T. cacoeciae*, *T. daumalae*, *T. embryophagum*, *T. oleae* et *T. semblidis*) semble par contre être regroupé en position basale à l'exception de *T. daumalae* dont le statut a déjà été discuté (cf. Discussion du Chapitre 3 et § 4.2.2.2.3 ce Chapitre).

4.3.2 Précisions concernant les complexes d'espèces majeurs

Grâce à cette approche multilocus, des précisions ont pu être obtenues concernant plusieurs groupes identifiés initialement avec *COI*. *A posteriori*, les cas les plus intéressants sont les complexes « cacoeciae-embryophagum » et « brassicae-euproctidis ».

Concernant le complexe « cacoeciae-embryophagum », la structuration observée sur la phylogénie multilocus, la variabilité des modes de reproduction et l'hétérogénéité des statuts

d'infections à *Wolbachia* laissent supposer l'existence de plusieurs espèces au sein de ce complexe. Le groupe XIIa (Figure 7) qui inclut, d'une part, des souches identifiées par différentes sources comme *T. cacoeciae* (cf. Tableau 1 en §7) et d'autre part, l'ensemble des souches thélytoques non infectées par *Wolbachia*. Il correspond bien à l'espèce *T. cacoeciae* telle que décrit dans la littérature (Pintureau, 2008). Cette espèce est largement distribuée avec une certaine variabilité moléculaire dont l'origine (isolement par la distance, spécialisation écologique) reste à étudier plus précisément. Concernant les autres groupes du cluster, les compatibilités de reproduction n'ont malheureusement pas pu être testées faute de souches vivantes disponibles. Le nombre d'espèces reste donc difficile à définir. Nous supposons toutefois que l'espèce *T. embryophagum* correspond au sous-groupe (iv) représenté par les haplotypes Hap196, Hap197 et Hap198, sur la base de l'hôte de développement, *Thaumetopoea pytyocampa* (Pintureau, 2008; Tsankov *et al.*, 1998). Des études morphologiques à partir des mâles obtenus sont actuellement en cours (Groussier G., comm. pers.).

Concernant le complexe « brassicae-euproctidis », une structuration moléculaire a été mise en évidence avec, en particulier, la présence de deux sous-clusters partiellement compatibles occupant deux aires géographiques partiellement chevauchants :

- D'une part, un sous-cluster présent en Europe voire Moyen-Orient incluant la souche de référence *T. brassicae* ;
- D'autre part, un sous-cluster présent en Afrique du Nord, Au Moyen-Orient et en Europe du Sud incluant notre souche de référence *T. euproctidis*.

Actuellement, il est difficile de faire la part des choses entre un initial isolement par la distance et ou un processus de divergence voire spéciation récent. Une étude morphologique est actuellement en cours au laboratoire (Groussier G, comm. pers.).

4.3.3 Infections à *Wolbachia*

Suite aux différentes approches utilisées (PCR, qPCR et MLST), moins de la moitié des 20 espèces identifiées semblent infectées par *Wolbachia*, soit de façon généralisée (quasiment toutes les populations/individus), soit de façon plus ponctuelle (quelques souches/individus) ce qui pourraient correspondre à deux scénarios différents d'acquisition, respectivement une infection ancienne à l'échelle évolutive qui se serait maintenue/généralisée et des acquisitions beaucoup plus récentes et locales. Dans tous ces cas, nous pensons qu'il s'agit de mono-

infection ou en tout cas qu'un seul variant est majoritaire. A l'échelle du genre *Trichogramma*, nous mettons en évidence une certaine diversité moléculaire des variants *Wolbachia* qui semblent toutefois tous appartenir au groupe B et induire de la thélytoquie. Ces patrons d'infections peuvent être comparés avec ceux observés chez les hôtes potentiels, les Lépidoptères, ou chez d'autres hyménoptères parasitoïdes. D'après (Paniagua Voirol *et al.*, 2018), deux symbiotes bactériens, *Wolbachia* et des *Spiroplasma*, ont été retrouvés chez les lépidoptères et (Ahmed *et al.*, 2015) admettent que 80% des lépidoptères sont infectés par *Wolbachia*. En Europe, ces auteurs estiment cependant le taux d'infection par *Wolbachia* à seulement 45% sur 46 espèces étudiées, avec une prévalence d'infection à l'intérieur d'une population qui est d'environ 1/3. Sachant que, à l'échelle évolutive, les transferts horizontaux peuvent expliquer l'acquisition d'endosymbiotes par une espèce (Vavre *et al.*, 1999), ces taux d'infection ne semblent pas incohérents avec ceux obtenus avec notre échantillonnage dans lequel au minimum 45% d'espèces de trichogrammes sont infectées par ce symbiote bactérien de manière totale ou partielle suivant les espèces. Si l'on considère toutefois que le mode de vie « parasitoïde » est plus susceptible de favoriser l'acquisition d'endosymbiotes variés (Vavre *et al.*, 1999), les taux d'infection, la diversité moléculaire bactérienne ainsi que la diversité des effets induits par *Wolbachia* apparaissent plutôt faibles aux regards d'autres taxons (Stouthamer *et al.*, 1999). En conclusion, les dynamiques d'acquisition-perte de *Wolbachia* restent encore mal comprises, un manque notable concernant l'infection à *Wolbachia* chez les hôtes réels des souches étudiées.

5 CONCLUSION ET PERSPECTIVES

5.1 Bilan synthétique des résultats et apports par rapport à l'état de l'art

Sur la base d'analyses moléculaires et d'expériences de croisement, ce travail a permis d'identifier vingt-quatre « entités », différenciées les unes des autres d'un point de vue moléculaire et isolées reproductivement les unes des autres. Parmi elles, douze ont été associées à une seule espèce déjà décrite (*T. achaeae*, *T. bourarachae*, *T. chilonis*, *T. cordubensis*, *T. danausicida*, *T. dendrolimi*, *T. nerudai*, *T. oleae*, *T. ostrinae*, *T. pretiosum*, *T. principium*, *T. semblidis*). Deux de ces « entités » semblent correspondre à des complexes d'espèces regroupant au moins deux espèces difficilement séparables sur des critères moléculaires. Le premier regroupe des souches identifiées initialement comme *T. cacoecliae* et *T. embryophagum* ainsi que des souches nouvellement acquises moléculairement proches. En l'état, l'interprétation que nous proposons concernant ce complexe d'espèces est que les souches asexuées correspondent bien à *T. cacoecliae* mais que les souches sexuées (toutes initialement attribuées à *T. embryophagum*) pourraient se répartir en deux voire trois espèces. Le deuxième complexe d'espèces est composé notamment de souches initialement identifiées comme *T. brassicae* et de *T. euproctidis*. *A posteriori*, les données sur les distances moléculaires faibles et les isolements reproducteurs incomplets dans ce complexe, nous laissent envisager que ce complexe puisse être une seule espèce avec cependant une forte variabilité intraspécifique possiblement expliquée par (i) des isolements par la distance, (ii) des spécialisations écologiques voire (iii) des processus de spéciation en cours. Deux autres situations particulières ont été mises en évidence concernant, d'une part, *T. exiguum* associé à deux entités paraphylétiques et, d'autre part, *T. daumalae* et *T. evanescens* décrites comme bien différentes sur des bases morphologiques mais formant pourtant ici un ensemble monophylétique et interfécond. Enfin, sept de ces entités n'ont, en l'état, pas pu être associées à un nom d'espèce sans que l'on sache s'il s'agit d'espèces déjà décrites ou d'espèces nouvelles pour la science.

5.2 Complémentarité des approches moléculaires

Concernant le travail de caractérisation moléculaire, les deux approches (Barcoding et phylogénie multilocus) se sont montrées complémentaires.

Dans un premier temps, l'utilisation du marqueur *COI* s'est avérée d'une utilité suffisante et d'un coût (temps, argent) suffisamment limité pour explorer la diversité moléculaire au sein d'un grand nombre de souche. Hormis quelques cas avérés d'amplifications de pseudogènes (cas de l'entité « Vc » du complexe « V-misB » cf. Figure 7), la principale difficulté est la production de séquences approximatives (erreur de séquençage ou d'analyse), celle-ci ne concernant toutefois au final que peu d'individus. *A posteriori*, le diagnostic sur *COI* s'est révélé, la plupart du temps, concordant avec celui obtenu en phylogénie multilocus. Cette concordance suggère, à l'échelle de notre couverture taxonomique actuelle, l'absence d'introgession ou d'hybridation (Machado and Hey, 2003).

En comparaison avec le barcoding *COI*, la phylogénie multilocus présente évidemment un coût financier nettement supérieur (x 6-7 / individu) et nécessite à la fois un temps de traitement plus long ainsi qu'une analyse plus complexe. A ce titre et contrairement au barcoding *COI*, l'appropriation de ce type d'approche sera probablement plus difficile pour des acteurs intéressés par le diagnostic mais non spécialisés (ex : sociétés productrices d'auxiliaires). Parallèlement à des objectifs de diagnostic, la phylogénie multilocus présente aussi l'avantage de clarifier l'histoire évolutive des trichogrammes et de caractériser ainsi plus finement ce modèle d'étude pour des études en écologie évolutive (ex : Vercken *et al.*, 2015), ou plus généralement pour classer proprement ce taxon dans le cadre de la lutte biologique. Dans notre cas, la phylogénie multilocus a ainsi nettement amélioré la résolution des complexes, ainsi que les supports des branches terminales des différents clusters. Malgré tout, l'analyse multilocus obtenue présente encore des valeurs de bootstrap décevantes au niveau des nœuds profonds. Dans l'étude de (Al khatib *et al.*, 2016), quasiment les mêmes marqueurs (gène BUB3 utilisé en plus) et les mêmes méthodes ont été utilisées pour résoudre les relations phylogénétiques au sein du complexe d'espèce « *Eupelmus urozonus* » composé *a posteriori* de 18 espèces. L'analyse réalisée avait alors présenté des soutiens de branches nettement supérieurs à ceux observés dans notre étude avec pourtant des qualités d'information (taux de sites variables et informatifs) comparables. Comparativement au cas d'étude *Eupelmus urozonus*, ce manque de résolution chez *Trichogramma* pourrait être

interprété par des processus de spéciation plus récents et/ou rapides au sein du genre *Trichogramma*.

Une perspective à court terme de notre travail sera d'utiliser des techniques de NGS telle que le Rad-sequencing⁹ (Davey and Blaxter, 2010) en collaboration notamment avec le CBGP. De telles techniques permettront en effet d'utiliser un très grand nombre de marqueurs répartis dans tout le génome avec la possibilité d'explorer simultanément la variabilité inter et intraspécifique avec une puissance de résolution supérieure à celle que l'on peut espérer avec des marqueurs classiques (par exemple : séquençage Sanger pour la phylogénie et génotypage par microsatellites pour la génétique des populations). Cette avancée technologique nous permettrait (i) d'obtenir une meilleure résolution phylogénétique des trichogrammes et (ii) d'acquérir un outil de contrôle qualité rapide et fiable des souches de notre CRB.

5.3 Limite de la couverture taxonomique

Malgré le nombre important d'échantillons (> 600), la zone couverte (Sud de la France, pays frontaliers et plus distants) et la diversité des habitats (milieux cultivés et naturels), la couverture du genre *Trichogramma* reste encore faible avec moins de 25% des espèces de trichogrammes décrites en région Ouest-Paléarctique représentées et moins de 10 % des espèces mondiales. Cette couverture partielle limite évidemment notre capacité à tester précisément l'efficacité de nos méthodes de caractérisation moléculaire (Barcoding COI et phylogénie multilocus) même si nous avons prouvé leurs capacités de discrimination sur des espèces considérées comme proches (*T. oleae* et *T. pretiosum*) voire « sœurs » (*T. cacoeciae* et *T. embryophagum*). A l'exception peut-être des complexes d'espèces, cette couverture partielle existe d'ailleurs probablement aussi à l'échelle intraspécifique. En fait, la principale limite concerne vraisemblablement du choix de l'espèce *Ephesttia kuehniella* comme hôtes sentinelles. Si cet hôte est facilement disponible et présent l'avantage de ne pas éclore après exposition sur le terrain, il a été constaté des limites à son usage, certaines espèces de trichogrammes délaissant cet hôte ou s'y développant avec difficulté (Groussier, comm. pers). Ainsi, certaines espèces comme *T. gicai* et *T. embryophagum* ont récemment été capturées sur des pontes naturelles dans des aires géographiques préalablement échantillonnées avec des œufs-sentinelles d'*E. kuehniella*. Par ailleurs, l'élevage de ces deux espèces en laboratoire

⁹ RADSeq Restriction site-Associated DNA Sequencing : nouvelle génération de séquençage dont le principe global repose sur le séquençage massif de courtes régions encadrant le site de restriction d'une endonucléase.

s'est avéré difficile voire impossible sur cet hôte et leur maintien ne s'est avéré possible qu'en utilisant une autre espèce, *Spilosoma lutea*. Ce genre d'observations montre d'ailleurs qu'il existe probablement des phénomènes de spécialisation écologique chez les trichogrammes qui ne sont pas encore compris. Que ce soit pour augmenter la couverture taxonomique du genre *Trichogramma*, pour échantillonner davantage la variabilité intraspécifique ou pour mieux comprendre la spécialisation écologique, la poursuite d'échantillonnage s'avère donc encore une priorité.

5.4 La caractérisation morphologique, le « chaînon manquant »

Historiquement, la caractérisation morphologique est la première source d'information pour essayer de classer les espèces et donc le point de départ pour mener des travaux de taxonomie intégrative. Concernant les trichogrammes, l'accès à ce type de données s'avère particulièrement compliqué compte tenu (i) du faible nombre et/ou de la faible disponibilité de spécialistes, (ii) du petit nombre de caractères considérés comme « diagnostic » et leur éventuelle subjectivité ; (iii) des difficultés d'accès aux « types¹⁰ » du fait de leurs localisations dispersées (Nagarkatti and Nagaraja, 1977) ou (iv) des éventuels freins/risques liés à des tentatives d'extraction d'ADN sur ces « types ». Comme évoqué à plusieurs reprises dans ce document, cette activité de caractérisation morphologique demeure toutefois indispensable pour lever certaines ambiguïtés majeures (par exemple : nombre d'espèces différentes actuellement regroupées sous l'appellation *Trichogramma embryophagum* ; monophylie des souches *T. daumalae* et *T. evanescens* pourtant réparties dans des sections morphologiques sur des bases morphologiques) ou plus mineures (paraphylie de deux souches identifiées initialement comme *T. exiguum*). Dans le cadre du projet ANR TRIPTIC et au sein d'Ep-Coll, des études morphologiques sont donc menées dans le but de réviser les clés d'identification de *Trichogramma*.

¹⁰Type : en systématique, le type est l'élément de référence attaché à un nom scientifique à partir duquel une espèce vivante ou ayant vécu, a été décrite.

5.5 Vers un couplage « Marqueurs – Traits »

Comme évoqué précédemment, on peut être, à moyen terme (3-6 ans) raisonnablement optimiste concernant les possibilités (i) d'améliorer la couverture taxonomique du genre *Trichogramma* ; (ii) d'améliorer la cohérence entre caractérisations moléculaires et morphologiques ; (iii) de développer des méthodes NGS avec une grande densité de marqueurs. Une ambition serait donc maintenant de rapprocher les informations obtenues au niveau moléculaire à celles obtenues au niveau phénotypique. En effet, que ce soit à l'échelle interspécifique ou intraspécifique, de nombreuses études ont mis en évidence des différences sur des caractères comportementaux (Wajnberg *et al.*, 2000) ou phénotypiques (Benvenuto *et al.*, 2012). Actuellement des travaux sont en cours (Thèse de Michela ION SCOTTA, UMR ISA) sur la diversité intraspécifique chez *T.cacoeciae* (cluster IIa) sur des aspects de « thermal biology », c'est-à-dire l'étude de variabilités intraspécifiques de trait d'histoire de vie (capacités de locomotion), et d'éventuelles adaptations locales (capacités de diapause¹¹, capacités d'overwintering¹²) en réponse à la résistance au froid. Parallèlement, des travaux sont menés pour étudier la variabilité de traits de personnalité chez *T. evanescens* (Thèse de Silène LARTIGUE, Collab. Univ Bourgogne & UMR ISA). Enfin, des collègues de l'Institut Sophia Agrobiotech s'intéressent à la réponse comportementale des trichogrammes en réponse à des composés chimiques ou à d'autres facteurs (photopériode). Les traits les plus « phénotypables » seraient évidemment les plus favorables pour une tentative de couplage « marqueurs – traits ». D'un point de vue appliqué, de telles études pourraient être pertinentes à l'amélioration génétique des trichogrammes. En lutte biologique, les traits phénotypiques recherchés chez un agent de lutte biologique concernent à la fois les performances en conditions de production industrielle, comme par exemple, le taux de parasitisme, la durée de développement, la fécondité réalisée, trois facteurs qui déterminent le coût de production et de spécificité, mais également les performances au champ qui font appel à d'autres caractères, par exemple, les capacités de locomotion (vitesse, saut, endurance), l'adéquation aux facteurs abiotiques (température par exemple). Sur certains de ces caractères, des méthodes de « sélection assistée par marqueurs¹³ » pourraient être envisagées ce qui représenterait une « rupture » par rapport aux tentatives actuelles d'amélioration génétique des trichogrammes et plus généralement des arthropodes utilisés comme auxiliaires de lutte biologique.

¹¹ Diapause : phase génétiquement déterminée dans le développement d'un organisme où il diminue l'intensité de ses activités métaboliques en attendant des conditions plus clémentes.

¹² Overwintering : processus qui permet à un organisme de passer la saison hivernale.

¹³ SAM : méthode de sélection d'individus à partir de traits phénotypiques jugés performants dont la sélection est réalisée à l'aide de marqueurs moléculaires pertinents associés. Cette méthode a pour but de sélectionner rapidement et à moindre coût des individus ou des souches plus performants.

6 BIBLIOGRAPHIE

- Ahmed, M.Z., Araujo-Jnr, E.V., Welch, J.J., and Kawahara, A.Y. (2015). *Wolbachia* in butterflies and moths: geographic structure in infection frequency. *Front. Zool.* 12, 16.
- Al Khatib, F., Fusu, L., Cruaud, A., Gibson, G., Borowiec, N., Rasplus, J.-Y., Ris, N., and Delvare, G. (2014). An integrative approach to species discrimination in the *Eupelmus urozonus* complex (Hymenoptera, Eupelmidae), with the description of 11 new species from the Western Palaearctic. *Syst. Entomol.* 39, 806–862.
- Al khatib, F., Cruaud, A., Fusu, L., Genson, G., Rasplus, J.-Y., Ris, N., and Delvare, G. (2016). Multilocus phylogeny and ecological differentiation of the “*Eupelmus urozonus* species group” (Hymenoptera, Eupelmidae) in the West-Palaearctic. *BMC Evifiol. Biol.* 16.
- Almeida, R.P., and Stouthamer, R. (2015). ITS-2 sequences-based identification of *Trichogramma* species in South America. *Braz. J. Biol.* 75, 974–982.
- Ávila-Rodríguez, V., Alvarado-Gómez, O.G., González-Hernández, A., and Nava-Camberos, U. (2013). Differentiation and Phylogeny of Trichogrammatidae (Hymenoptera: Chacidoidea) from Mexico Based on ITS2 and 18S Molecular Markers of rDNAr and COII of the mDNA. *Southwest. Entomol.* 38, 299–312.
- Baldo, L., Hotopp, J.C.D., Jolley, K.A., Bordenstein, S.R., Biber, S.A., Choudhury, R.R., Hayashi, C., Maiden, M.C.J., Tettelin, H., and Werren, J.H. (2006). Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Appl. Environ. Microbiol.* 72, 7098–7110.
- Bandelt, H.J., Forster, P., and Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37–48.
- Benvenuto, C., Tabone, E., Vercken, E., Sorbier, N., Colombel, E., Warot, S., Fauvergue, X., and Ris, N. (2012). Intraspecific variability in the parasitoid wasp *Trichogramma chilonis*: can we predict the outcome of hybridization? *Evol. Appl.* 5, 498–510.
- Brady, S.G., Schultz, T.R., Fisher, B.L., and Ward, P.S. (2006). Evaluating alternative hypotheses for the early evolution and diversification of ants. *Proc. Natl. Acad. Sci.* 103, 18172–18177.
- Brower, A.V.Z., and DeSalle, R. (1998). Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: the utility of wingless as a source of characters for phylogenetic inference. *Insect Mol. Biol.* 7, 73–82.
- Buhay, J.E. (2009). “COI-like” Sequences Are Becoming Problematic in Molecular Systematic and DNA Barcoding Studies. *J. Crustac. Biol.* 29, 96–110.
- Cônsoli, F.L., José R.P. Parra, and Roberto A. Zucchi (2010). Egg Parasitoids in Agroecosystems with Emphasis on *Trichogramma*. In *Progress of Biological Control*, p. 479.
- Cruaud, A., Jabbour-Zahab, R., Genson, G., Cruaud, C., Couloux, A., Kjellberg, F., van Noort, S., and Rasplus, J.-Y. (2010). Laying the foundations for a new classification of Agaonidae (Hymenoptera: Chalcidoidea), a multilocus phylogenetic approach. *Cladistics* 26, 359–387.

- Cruaud, A., Rønsted, N., Chantarasuwan, B., Chou, L.S., Clement, W.L., Couloux, A., Cousins, B., Genson, G., Harrison, R.D., Hanson, P.E. (2012). An Extreme Case of Plant–Insect Codiversification: Figs and Fig-Pollinating Wasps. *Syst. Biol.* *61*, 1029–1047.
- Cruaud, A., Underhill, J.G., Huguin, M., Genson, G., Jabbour-Zahab, R., Tolley, K.A., Rasplus, J.-Y., and van Noort, S. (2013). A Multilocus Phylogeny of the World Sycoecinae Fig Wasps (Chalcidoidea: Pteromalidae). *Plos One* *8*, UNSP e79291.
- Cruaud, P., Rasplus, J.-Y., Rodriguez, L.J., and Cruaud, A. (2017). High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy. *Sci. Rep.* *7*, 41948.
- Danforth, B.N., Brady, S.G., Sipes, S.D., and Pearson, A. (2004). Single-copy nuclear genes recover Cretaceous-age divergences in bees. *Syst. Biol.* *53*, 309–326.
- Danforth, B.N., Sipes, S., Fang, J., and Brady, S.G. (2006). The history of early bee diversification based on five genes plus morphology. *Proc. Natl. Acad. Sci.* *103*, 15118–15123.
- Danforth, B.N., Cardinal, S., Praz, C., Almeida, E.A.B., and Michez, D. (2013). The Impact of Molecular Data on Our Understanding of Bee Phylogeny and Evolution. In *Annual Review of Entomology*, Vol 58, M.R. Berenbaum, ed. (Palo Alto: Annual Reviews), pp. 57-+.
- Davey, J.W., and Blaxter, M.L. (2010). RADSeq: next-generation population genetics. *Brief. Funct. Genomics* *9*, 416–423.
- Dayrat, B. (2005). Towards integrative taxonomy. *Biol. J. Linn. Soc.* *85*, 407–415.
- Desneux, N., Decourtye, A., and Delpuech, J.-M. (2006). The Sublethal Effects of Pesticides on Beneficial Arthropods. *Annu. Rev. Entomol.* *52*, 81–106.
- Dupuis, J.R., Roe, A.D., and Sperling, F.A.H. (2012). Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Mol. Ecol.* *21*, 4422–4436.
- Eilenberg, J., Hajek, A., and Lomer, C. (2001). Suggestions for unifying the terminology in biological control. *BioControl* *46*, 387–400.
- Ercan, F.S., and Öztem, S. (2013). Mitochondrial and ribosomal DNA sequence analysis for discrimination of *Trichogramma euproctidis* Girault and *Trichogramma brassicae* Bezdenko (Hymenoptera: Trichogrammatidae). *8*.
- Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* *3*, 294–299.
- Frézal, L., and Leblois, R. (2008). Four years of DNA barcoding: Current advances and prospects. *Infect. Genet. Evol.* *8*, 727–736.
- Fulton, T.L., Letts, B., and Shapiro, B. (2012). Multiple losses of flight and recent speciation in steamer ducks. *Proc. R. Soc. B-Biol. Sci.* *279*, 2339–2346.
- Gebiola, M., Gómez-Zurita, J., Monti, M.M., Navone, P., and Bernardo, U. (2012). Integration of molecular, ecological, morphological and endosymbiont data for species

delimitation within the *Pnigalio soemius* complex (Hymenoptera: Eulophidae). *Mol. Ecol.* 21, 1190–1208.

Gebiola, M., Monti, M.M., Johnson, R.C., Woolley, J.B., Hunter, M.S., Giorgini, M., and Pedata, P.A. (2017). A revision of the *Encarsia pergandiella* species complex (Hymenoptera: Aphelinidae) shows cryptic diversity in parasitoids of whitefly pests: Revision of the *Encarsia pergandiella* complex. *Syst. Entomol.* 42, 31–59.

Hance, T., van Baaren, J., Vernon, P., and Boivin, G. (2007). Impact of Extreme Temperatures on Parasitoids in a Climate Change Perspective. *Annu. Rev. Entomol.* 52, 107–126.

Hassan, S.A. (1993). The mass rearing and utilization of *Trichogramma* to control lepidopterous pests: Achievements and outlook. *Pestic. Sci.* 37, 387–391.

Hebert, P.D.N., Cywinska, A., Ball, S.L., and deWaard, J.R. (2003). Biological identifications through DNA barcodes. *Proc. R. Soc. B Biol. Sci.* 270, 313–321.

Heraty, J.M., Woolley, J.B., Hopper, K.R., Hawks, D.L., Kim, J.-W., and Buffington, M. (2007). Molecular phylogenetics and reproductive incompatibility in a complex of cryptic species of aphid parasitoids. *Mol. Phylogenet. Evol.* 45, 480–493.

Jermiin, L., and Crozier, R. (1994). The Cytochrome-B Region in the Mitochondrial DNA of the Ant *Tetraponera rufoniger*: Sequence Divergence in Hymenoptera May Be Associated with Nucleotide Content. *J. Mol. Evol.* 38, 282–294.

Kazmer, D. (1991). Isoelectric-Focusing Procedures for the Analysis of Allozymic Variation. *Ann. Entomol. Soc. Am.* 84, 332–339.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinforma. Oxf. Engl.* 28, 1647–1649.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33, 1870–1874.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., and Calcott, B. (2017). PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Mol. Biol. Evol.* 34, 772–773.

Lenteren, J.C.V., and Bueno, V.H.P. (2003). Augmentative biological control of arthropods in Latin America. *BioControl* 48, 123–139.

Li, C., Lu, G., Ortí, G., and Buckley, T. (2008). Optimal Data Partitioning and a Test Case for Ray-Finned Fishes (Actinopterygii) Based on Ten Nuclear Loci. *Syst. Biol.* 57, 519–539.

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452.

- Liu, Y.-D., Hou, M., and SONG, K. (2017). Molecular identification of *Trichogramma* species from south and South-East Asia and natural *Wolbachia* infection. *Entomol. Fenn.* 28, 57–66.
- Lohse, K., Sharanowski, B., Blaxter, M., Nicholls, J.A., and Stone, G.N. (2011). Developing EPIC markers for chalcidoid Hymenoptera from EST and genomic data. *Mol. Ecol. Resour.* 11, 521–529.
- Machado, C.A., and Hey, J. (2003). The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proc. R. Soc. B Biol. Sci.* 270, 1193–1202.
- Marchand, A., Sellier, N., Warot, S., Ionscotta, M., Ris, N., and Groussier-Bout, G. (2017). Formalisation d'un centre de Ressources Biologiques dédié aux parasitoïdes oophages : CRB Ep-Coll. Cah. Tech. INRA 49–58.
- Mayr, E. (1999). *Systematics and the origin of species, from the viewpoint of a zoologist* (Cambridge, Mass: Harvard University Press).
- Meyer, C.P., and Paulay, G. (2005). DNA Barcoding: Error Rates Based on Comprehensive Sampling. *PLoS Biol.* 3, e422.
- Miller, M.A., Pfeiffer, W., and Schwartz, T. (2012). The CIPRES Science Gateway: Enabling High-impact Science for Phylogenetics Researchers with Limited Resources. In *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the EXtreme to the Campus and Beyond*, (New York, NY, USA: ACM), pp. 39:1–39:8.
- Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B., and Worm, B. (2011). How Many Species Are There on Earth and in the Ocean? *PLOS Biol.* 9, e1001127.
- Mottern, J.L., and Heraty, J.M. (2014). Revision of the *Cales noacki* species complex (Hymenoptera, Chalcidoidea, Aphelinidae). *Syst. Entomol.* 39, 354–379.
- Muru, D., Groussier, G., Marchand, A., Rasplus, J.-Y., Cruaud, A., Warot, S., Calcagno, V., Vercken, E., and Ris, N. (in prep). A nation-wide collective sampling of *Trichogramma* species with data on plants and habitat.
- Nagarkatti, S., and Nagaraja, H. (1977). Biosystematics of *Trichogramma* and *Trichogrammatoidea* Species. *Annu. Rev. Entomol.* 22, 157–176.
- Nylander, J.A.. (2004). MrAIC.pl. Program distributed by the author (Uppsala University: Evolutionary Biology Centre).
- Orrego, C., and Agudelo-Silva, F. (1993). Genetic Variation in the Parasitoid Wasp *Trichogramma* (Hymenoptera: Trichogrammatidae) Revealed by DNA Amplification of a Section of the Nuclear Ribosomal Repeat. *Fla. Entomol.* 76, 519–524.
- Paniagua Voirol, L.R., Frago, E., Kaltenpoth, M., Hilker, M., and Fatouros, N.E. (2018). Bacterial Symbionts in Lepidoptera: Their Diversity, Transmission, and Impact on the Host. *Front. Microbiol.* 9.

- Parra, J.R.P., and Zucchi, R.A. (2004). *Trichogramma* in Brazil: feasibility of use after twenty years of research. *Neotrop. Entomol.* 33, 271–281.
- del Pino, M., Hernandez-Suarez, E., Cabello, T., Rugman-Jones, P., Stouthamer, R., and Polaszek, A. (2013). *Trichogramma canariensis* (Insecta: Hymenoptera: Trichogrammatidae) a parasitoid of eggs of the twin-spot moth *Chrysodeixis chalcites* (Lepidoptera: Noctuidae) in the Canary Islands. *Arthropod Syst. Phylogeny* 71, 169–179.
- Pinto, J.D., Koopmanschap, A.B., Platner, G.R., and Stouthamer, R. (2002). The North American *Trichogramma* (Hymenoptera: Trichogrammatidae) Parasitizing Certain Tortricidae (Lepidoptera) on Apple and Pear, with ITS2 DNA Characterizations and Description of a New Species. *Biol. Control* 23, 134–142
- Pinto, J.D., Oatman, E.R., and Platner, G.R. (1983). The identity of 2 closely related and frequently encountered species of new world *Trichogramma* (Hymenoptera, Trichogrammatidae). *Proc. Entomol. Soc. Wash.* 588–593.
- Pinto, J.D., Platner, G., and Sassaman, C. (1993). Electrophoretic Study of 2 Closely-Related Species of North-American *Trichogramma* - *T. pretiosum* and *T. deion* (hymenoptera, Trichogrammatidae). *Ann. Entomol. Soc. Am.* 86, 702–709.
- Pinto, J.D., Stouthamer, R., Platner, G.R., and Oatman, E.R. (1991). Variation in Reproductive Compatibility in *Trichogramma* and Its Taxonomic Significance (Hymenoptera: Trichogrammatidae). *Ann. Entomol. Soc. Am.* 84, 37–46.
- Pintureau, B. (1997). Systematic and genetical problems revised in two closely related species of *Trichogramma*, *T. embryophagum* and *T. cacoeciae* (Hym., Trichogrammatidae). *Misc. Zool.* 20.2, 8.
- Pintureau, B. (2008). Les espèces européennes de Trichogrammes (France : InLibroVeritas-Editions).
- Pintureau, B. (2009). La lutte biologique et les trichogrammes : application au contrôle de la pyrale du maïs : essai (Paris : Le Manuscrit).
- Pintureau, B., and Babault, M. (1981). Caractérisation enzymatique de *Trichogramma evanescens* et de *T. maidis* [Hym.: Trichogrammatidae]; Étude des hybrides. *Entomophaga* 26, 11–22.
- Pintureau, B., Grenier, S., Boléat, B., Lassablière, F., Heddi, A., and Khatchadourian, C. (2000). Dynamics of *Wolbachia* Populations in Transfected Lines of *Trichogramma*. *J. Invertebr. Pathol.* 76, 20–25.
- Pintureau, B., and Keita, F. (1989). New Data on the Esterases of Some *Trichogramma* (hym-Trichogrammatidae). *Biochem. Syst. Ecol.* 17, 603–608.
- Polaszek, A., Rugman-Jones, P.F., Stouthamer, R., Hernandez-Suarez, E., Cabello, T., and del Pino Pérez, M. (2012). Molecular and morphological diagnoses of five species of *Trichogramma*: biological control agents of *Chrysodeixis chalcites* (Lepidoptera: Noctuidae) and *Tuta absoluta* (Lepidoptera: Gelechiidae) in the Canary Islands. *BioControl* 57, 21–35.

- Poorjavad, N., Goldansaz, S.H., Machtelinckx, T., Tirry, L., Stouthamer, R., and van Leeuwen, T. (2012). Iranian *Trichogramma*: ITS2 DNA characterization and natural *Wolbachia* infection. *BioControl* 57, 361–374.
- Ronquist, F., and Liljeblad, J. (2001). Evolution of the gall wasp – host plant association. *Evolution* 55, 2503.
- Rousset, F., Bouchon, D., Pintureau, B., Juchault, P., and Solignac, M. (1992). *Wolbachia* Endosymbionts Responsible for Various Alterations of Sexuality in Arthropods. *Proc. R. Soc. B Biol. Sci.* 250, 91–98.
- Rugman-Jones, P.F., Wharton, R., Noort, T. van, and Stouthamer, R. (2009). Molecular differentiation of the *Psytalia concolor* (Szépligeti) species complex (Hymenoptera: Braconidae) associated with olive fly, *Bactrocera oleae* (Rossi) (Diptera: Tephritidae), in Africa. *Biol. Control* 49, 17–26.
- Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, E., and Crozier, R.H. (2010). Integrative Taxonomy: A Multisource Approach to Exploring Biodiversity. *Annu. Rev. Entomol.* 55, 421–438.
- Shokralla, S., Porter, T.M., Gibson, J.F., Dobosz, R., Janzen, D.H., Hallwachs, W., Golding, G.B., and Hajibabaei, M. (2015). Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Sci. Rep.* 5.
- Smith, S.M. (1996). Biological Control with *Trichogramma*: Advances, Successes, and Potential of Their Use. *Annu. Rev. Entomol.* 41, 375–406.
- Song, H., Buhay, J.E., Whiting, M.F., and Crandall, K.A. (2008). Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc. Natl. Acad. Sci.* 105, 13486–13491.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Stouthamer, R. (1993). The use of sexual versus asexual wasps in biological control. *Entomophaga* 38, 3–6.
- Stouthamer, R., Breeuwer, J. a. J., and Hurst, G.D.D. (1999). *Wolbachia pipientis*: Microbial manipulator of arthropod reproduction. *Annu. Rev. Microbiol.* 53, 71–102.
- Stouthamer, R., Hu, J., Kan, F.J.P.M. van, Platner, G.R., and Pinto, J.D. (1999). The utility of internally transcribed spacer 2 DNA sequences of the nuclear ribosomal gene for distinguishing sibling species of *Trichogramma*. *BioControl* 43, 421–440
- Stouthamer, R., Jochemsen, P., Platner, G.R., and Pinto, J.D. (2000). Crossing Incompatibility Between *Trichogramma minutum* and *T. platneri* (Hymenoptera: Trichogrammatidae): Implications for Application in Biological Control. *Environ. Entomol.* 29, 832–837.
- Sumer, F., Oztemiz, S., S, O., and Tuncbilek, A. (2013). Mitochondrial and ribosomal DNA sequence analysis for discrimination of *Trichogramma euproctidis* Girault and *Trichogramma brassicae* Bezdenko (Hymenoptera:Trichogrammatidae). *Turk. J. Entomol.* 37.

Tsankov, G., Schmidt, G.H., and Mirchev, P. (1998). Studies on the egg parasitism in *Thaumetopoea pityocampa* over a period of four years (1991-1994) at Marikostino/Bulgaria. *Anz. Schadlingskunde Pflanzenschutz Umweltschutz* 71, 1–7.

Vanlerberghe-Masutti, F. (1994). Marqueurs moléculaires et génétique des populations d'insectes parasitoïdes. *Vet. Res.* 25, 599–600

Vavre, F., Fleury, F., Lepetit, D., Fouillet, P., and Bouletreau, M. (1999). Phylogenetic evidence for horizontal transmission of *Wolbachia* in host-parasitoid associations. *Mol. Biol. Evol.* 16, 1711–1723.

Vavre, F., de Jong, J.H., and Stouthamer, R. (2004). Cytogenetic mechanism and genetic consequences of thelytoky in the wasp *Trichogramma cacoeciae*. *Heredity* 93, 592–596.

Vercken, E., Fauvergue, X., Ris, N., Crochard, D., and Mailleret, L. (2015). Temporal autocorrelation in host density increases establishment success of parasitoids in an experimental system. *Ecol. Evol.* 5, 2684–2693.

Wajnberg, E., Fauvergue, X., and Pons, O. (2000). Patch leaving decision rules and the Marginal Value Theorem: an experimental analysis and a simulation model. *Behav. Ecol.* 11, 577–586.

Warot, S., Cruaud, A., Wera, H., Rasplus, J.-Y., and Ris, N. (2017). Délimitations d'espèces ouest-paléarctiques du genre *Trichogramma*.

Warot, S., Cruaud, A., Groussier, G., Malausa, T., Martinez-Rodriguez, P., Pintureau, B., Seguret, J., and Ris, N. Insights into the molecular diversity and species delineation in the genus *Trichogramma* with a focus on West Palaearctic.

Will, K.W., Mishler, B.D., Wheeler, Q.D., and Savolainen, V. (2005). The Perils of DNA Barcoding and the Need for Integrative Taxonomy. *Syst. Biol.* 54, 844–851.

Wilson, E.O., and Holldobler, B. (2005). The rise of the ants: A phylogenetic and ecological explanation. *Proc. Natl. Acad. Sci.* 102, 7411–7414.

Yamakawa, A.Y., and Imai, H. (2013). PCR-RFLP typing reveals a new invasion of Taiwanese *Meretrix* (Bivalvia: Veneridae) to Japan. *Aquat. Invasions* 8, 407–415.

7 TABLEAUX VOLUMINEUX

Compte tenu de leurs tailles, ces deux tableaux ont été placés en fin du mémoire avant les suppléments.

Tableau 1 : Détails des haplotypes COI utilisés dans cette étude

Les catégories (A-E) sont détaillées dans le § 2-Matériels et Méthodes.

L'étiquette "varprot" indique le nombre d'haplotypes pour lesquels au moins une variation de la séquence d'acides aminés a été observée.

Les références individuelles et les souches de référence indiquent respectivement l'individu (numérotation interne) et la souche à partir de laquelle l'ADN a été obtenu.

id	Souches de référence	Origine des souches	Source	Expertise morphologique	Identification morphologique	Identification moléculaire	GENBANK Accession	Catégories	Séquences acides aminés
Hap_001	81a	France, 1975	Historic collections	BP	<i>T. near principium</i>	II - principium	MG932137	C	
Hap_002	Tdaum	France, ?	Historic collections	BP	<i>T. daumalae</i>	I - daumalae-evanescens	MG932138	C	varprot
Hap_003	S2	Yougoslavia, 1972	Historic collections	BP	<i>T. oleae</i>	9 - oleae	MG932139	D	
Hap_004	Al22	France, 1990	Historic collections	–	–	VII - misA	MG932140	C	
Hap_005	BRA3	Moldavia, 1973	Historic collections	BP	<i>T. brassicae</i>	VI - brassicae-euproctidis	MG932141	D	
Hap_006	TCAL	France, 1987-1989	Historic collections	BP	<i>T. cacoeciae</i>	XII - cacoeciae-embryophagum	MG932142	D	
Hap_007	TCMz	France, 1987-1989	Historic collections	BP	<i>T. cacoeciae</i>	XII - cacoeciae-embryophagum	MG932143	B	
Hap_008	Tdaum	France, ?	Historic collections	BP	<i>T. daumalae</i>	I - daumalae-evanescens	MG932144	D	
Hap_009	Exi3	USA, 1995	Historic collections	BP	<i>T. exiguum</i>	7 - exiguum 1	MG932145	B	
Hap_010	Syr96D	Uruguay, 1993	Historic collections	BP	<i>T. principium</i>	II - principium	MG932146	B	
Hap_011	A1	France, 1989	Historic collections	BP	<i>T. evanescens</i>	I - daumalae-evanescens	MG932147	D	
Hap_012	TP63	Egypt, 2004	Historic collections	BP	<i>T. cordubensis</i>	III - cordubensis	MG932148	D	
Hap_013	PRET_FREDON_PTA	France (Martinique), 2014	Other source	–	–	XIV - pretiosum	MG932149	D	
Hap_014	DB	Portugal, 1991	Historic collections	BP	<i>T. euproctidis</i>	VI - brassicae-euproctidis	MG932150	C	
Hap_015	TP41EG	Egypt, 2002	Historic collections	BP	<i>T. bourarachae</i>	XI - bourarachae	MG932151	B	
Hap_016	SEMV	France, 1998	Historic collections	BP	<i>T. semblidis</i>	XIII - swC	MG932152	C	
Hap_017	SEMA2	France, 2002	Historic collections	BP	<i>T. semblidis</i>	VIII - semblidis	MG932153	C	
Hap_018	ACJYR014-29	France, 2015	TRIPTIC project	–	–	III - cordubensis	MG932154	A	varprot
Hap_019	PMBio1	France, 2014	Historic collections	BP	<i>T. cacoeciae</i>	XII - cacoeciae-embryophagum	MG932155	D	
Hap_020	ISA15034	France, 2016	TRIPTIC project	–	–	XII - cacoeciae-embryophagum	MG932156	C	varprot
Hap_021	M1	France, 1993	Historic collections	BP	<i>T. evanescens</i>	I - daumalae-evanescens	MG932157	D	
Hap_022	B9.1	Moldavia, 2013	BIOLINE Agrosiences	JS	<i>T. brassicae</i>	VI - brassicae-euproctidis	MG932158	A	varprot

Hap_023	B9.6	Croatia, 2013	BIOLINE Agrosciences	JS	<i>T. brassicae</i>	VI - brassicae-euproctidis	MG932159	A	varprot
Hap_024	B9.40	Croatia, 2013	BIOLINE Agrosciences	JS	<i>T. brassicae</i>	VI - brassicae-euproctidis	MG932160	B	
Hap_025	B9.41	Croatia, 2013	BIOLINE Agrosciences	JS	<i>T. brassicae</i>	VI - brassicae-euproctidis	MG932161	A	varprot
Hap_026	B9.41	Croatia, 2013	BIOLINE Agrosciences	JS	<i>T. brassicae</i>	VI - brassicae-euproctidis	MG932162	A	
Hap_027	B9.10	France, 2013?	BIOLINE Agrosciences	JS	<i>T. brassicae</i>	VI - brassicae-euproctidis	MG932163	D	
Hap_028	B9.30	Moldavia, 2013	BIOLINE Agrosciences	JS	<i>T. brassicae</i>	VI - brassicae-euproctidis	MG932164	C	
Hap_029	B9.53	Croatia, 2013	BIOLINE Agrosciences	JS	<i>T. brassicae</i>	VI - brassicae-euproctidis	MG932165	B	varprot
Hap_030	B9.55	Croatia, 2013	BIOLINE Agrosciences	JS	<i>T. brassicae</i>	VI - brassicae-euproctidis	MG932166	C	
Hap_031	B9.57	Croatia, 2013	BIOLINE Agrosciences	JS	<i>T. brassicae</i>	VI - brassicae-euproctidis	MG932167	B	varprot
Hap_032	B9.3	Croatia, 2013	BIOLINE Agrosciences	JS	<i>T. brassicae</i>	VI - brassicae-euproctidis	MG932168	B	
Hap_033	ACJYR014-31	France, 2015	TRIPTIC project	–	–	III - cordubensis	MG932169	A	
Hap_034	B9.60	Croatia, 2013	BIOLINE Agrosciences	JS	<i>T. brassicae</i>	VI - brassicae-euproctidis	MG932170	B	
Hap_035	B6.1	Croatia, 2013	BIOLINE Agrosciences	JS	<i>T. brassicae</i>	VI - brassicae-euproctidis	MG932171	C	
Hap_036	B5.3	France, 2013?	BIOLINE Agrosciences	JS	<i>T. brassicae</i>	VI - brassicae-euproctidis	MG932172	D	
Hap_037	B5.6	France, 2013?	BIOLINE Agrosciences	JS	<i>T. brassicae</i>	VI - brassicae-euproctidis	MG932173	B	varprot
Hap_038	E1.3	France, 2013	BIOLINE Agrosciences	JS	<i>T. euproctidis</i>	VI - brassicae-euproctidis	MG932174	B	varprot
Hap_039	E1.7	France, 2013	BIOLINE Agrosciences	JS	<i>T. euproctidis</i>	VI - brassicae-euproctidis	MG932175	C	
Hap_040	E8.6	France, 2013?	BIOLINE Agrosciences	JS	<i>T. euproctidis</i>	VI - brassicae-euproctidis	MG932176	A	varprot
Hap_041	V2.11	Moldavia, 2013	BIOLINE Agrosciences	JS	<i>T. evanescens</i>	I - daumalae-evanescens	MG932177	A	
Hap_042	V2.13	Moldavia, 2013	BIOLINE Agrosciences	JS	<i>T. evanescens</i>	I - daumalae-evanescens	MG932178	B	
Hap_043	D2	China, 1992	Historic collections	BP	<i>T. dendrolimi</i>	X - dendrolimi	MG932179	B	
Hap_044	V2.12	Moldavia, 2013	BIOLINE Agrosciences	JS	<i>T. evanescens</i>	I - daumalae-evanescens	MG932180	D	
Hap_045	V2.16	Moldavia, 2013	BIOLINE Agrosciences	JS	<i>T. evanescens</i>	I - daumalae-evanescens	MG932181	C	
Hap_046	V11.16	France, 2013	BIOLINE Agrosciences	JS	<i>T. evanescens</i>	I - daumalae-evanescens	MG932182	B	
Hap_047	V3.8	France, ?	BIOLINE Agrosciences	JS	<i>T. evanescens</i>	I - daumalae-evanescens	MG932183	B	
Hap_048	V6.3	France, ?	BIOLINE Agrosciences	JS	<i>T. evanescens</i>	I - daumalae-evanescens	MG932184	B	
Hap_049	V3.6	France, ?	BIOLINE Agrosciences	JS	<i>T. evanescens</i>	I - daumalae-evanescens	MG932185	D	
Hap_050	V3.19	France, ?	BIOLINE Agrosciences	JS	<i>T. evanescens</i>	I - daumalae-evanescens	MG932186	C	
Hap_051	V3.19	France, ?	BIOLINE Agrosciences	JS	<i>T. evanescens</i>	I - daumalae-evanescens	MG932187	C	
Hap_052	V3.5	France, ?	BIOLINE Agrosciences	JS	<i>T. evanescens</i>	I - daumalae-evanescens	MG932188	C	
Hap_053	V2.6	Spain (Canarias), ?	BIOLINE Agrosciences	JS	<i>T. evanescens</i>	I - daumalae-evanescens	MG932189	A	
Hap_054	V2.8	Egypt, ?	BIOLINE Agrosciences	JS	<i>T. evanescens</i>	I - daumalae-evanescens	MG932190	B	
Hap_055	TR	Chili, 1994	Historic collections	BP	<i>T. nerudai</i>	4 - nerudai	MG932191	B	
Hap_056	V9.2	France, 2013	BIOLINE Agrosciences	JS	<i>T. evanescens</i>	I - daumalae-evanescens	MG932192	B	

Hap_057	IRAN02	Iran, 2014	Other source	–	–	3 - swA	MG932193	A	
Hap_058	IRAN04	Iran, 2014	Other source	–	–	I - daumalae-evanescens	MG932194	A	
Hap_059	MK	Morocco, 1991	Historic collections	BP	<i>T. euproctidis</i>	VI - brassicae-euproctidis	MG932195	C	
Hap_061	CHI1_Mart	France (Martinique), 2014	Other source	–	–	IV - chilonis	MG932196	C	
Hap_063	it18	Italy, ?	Historic collections	BP	<i>T. dendrolimi</i>	X - dendrolimi	MG932197	B	
Hap_064	101	France, 1976	Historic collections	BP	<i>T. cacoeciae</i>	XII - cacoeciae-embryophagum	MG932198	B	
Hap_065	URO2	Iran, 1996	Historic collections	BP	<i>T. embryophagum</i>	XII - cacoeciae-embryophagum	MG932199	B	
Hap_066	URO3	Iran, 1996	Historic collections	BP	<i>T. embryophagum</i>	XII - cacoeciae-embryophagum	MG932200	B	
Hap_067	M172	Uruguay, 1993	Historic collections	BP	<i>T. exiguum</i>	6 - exiguum 2	MG932201	B	
Hap_068	EgyptEupro	Egypt, ?	Historic collections	BP	<i>T. euproctidis</i>	VI - brassicae-euproctidis	MG932202	C	
Hap_069	T234	Uruguay, 1994	Historic collections	BP	<i>T. pretiosum</i>	XIV - pretiosum	MG932203	C	
Hap_070	K1	Turkey, 2013	Other source	–	–	VI - brassicae-euproctidis	MG932204	C	
Hap_071	Grab_haplo1	France, 2011	Other source	–	–	V - misB	MG932205	C	
Hap_072	Grab_haplo2	France, 2011	Other source	–	–	V - misB	MG932206	B	
Hap_073	TSM0016	France, 2015	TRIPTIC project	–	–	V - misB	MG932207	C	
Hap_074	Grab_haplo4	France, 2011	Other source	–	–	I - daumalae-evanescens	MG932208	B	
Hap_075	Grab_haplo7	France, 2011	Other source	–	–	XII - cacoeciae-embryophagum	MG932209	B	
Hap_076	GOT0098	France, 2016	TRIPTIC project	–	–	XII - cacoeciae-embryophagum	MG932210	C	
Hap_077	ISA11030	France, 2016	TRIPTIC project	–	–	XII - cacoeciae-embryophagum	MG932211	C	
Hap_078	Grab_haplo10	France, 2011	Other source	–	–	XII - cacoeciae-embryophagum	MG932212	C	
Hap_081	PeruD_haplo	Peru, 2011	Other source	–	–	8 - swG	MG932213	C	
Hap_083	commercial product	–	BIOLINE Agrosiences	JS	<i>T. achaeae</i>	2 - achaeae	MG932214	C	
Hap_084	–	–	Genbank	–	–	2 - achaeae	DQ177918	E	
Hap_085	BOU2	Morocco, 1992	Historic collections	BP	<i>T. bourarachae</i>	XI - bourarachae	MG932215	B	
Hap_086	COR10 sp1	France (Corsica), 2016	Other source	–	–	XII - cacoeciae-embryophagum	MG932216	A	
Hap_087	–	–	Genbank	–	–	XII - cacoeciae-embryophagum	KM242285	E	varprot
Hap_088	–	–	Genbank	–	–	XII - cacoeciae-embryophagum	DQ177917	E	
Hap_089	reunion	France (La Réunion), 1998	Historic collections	BP	<i>T. chilonis</i>	IV - chilonis	MG932217	C	
Hap_090	ChiP	Pakistan, 2003	Historic collections	BP	<i>T. chilonis</i>	IV - chilonis	MG932218	C	
Hap_091	ChiCa	China, before 1996	Historic collections	BP	<i>T. chilonis</i>	IV - chilonis	MG932219	C	
Hap_092	ChiV	Viet-Nam, 2000	Historic collections	BP	<i>T. chilonis</i>	IV - chilonis	MG932220	A	
Hap_093	ChiJ	Japan, 1999	Historic collections	BP	<i>T. chilonis</i>	IV - chilonis	MG932221	A	

Hap_094	ChiH	USA (Hawai), 2006	Historic collections	BP	<i>T. chilonis</i>	IV - chilonis	MG932222	C	
Hap_095	–	–	Genbank	–	–	IV - chilonis	DQ177915	E	
Hap_096	CAS134	France, 2016	TRIPTIC project	–	–	III - cordubensis	MG932223	D	
Hap_097	–	–	Genbank	–	–	10 - danausicida	JF776380	E	varprot
Hap_098	–	–	Genbank	–	–	X - dendrolimi	KC411494	E	
Hap_099	–	–	Genbank	–	–	X - dendrolimi	KC411496	E	
Hap_100	–	–	Genbank	–	–	X - dendrolimi	DQ177913	E	
Hap_101	ACJYR0132	France, 2015	TRIPTIC project	–	–	XII - cacoeciae-embryophagum	MG932224	A	varprot
Hap_102	–	–	Genbank	–	–	XII - cacoeciae-embryophagum	KM105169	E	varprot
Hap_103	–	–	Genbank	–	–	1 - ostriniae	DQ177914	E	
Hap_104	PREB	Brazil, 1995	Historic collections	BP	<i>T. pretiosum</i>	XIV - pretiosum	MG932225	B	
Hap_105	PRIB	Bulgaria, ?	Historic collections	BP	<i>T. principium</i>	II - principium	MG932226	C	
Hap_106	ACJYR0116	France, 2015	TRIPTIC project	–	–	XII - cacoeciae-embryophagum	MG932227	B	
Hap_107	ACJYR0121	France, 2015	TRIPTIC project	–	–	XII - cacoeciae-embryophagum	MG932228	C	
Hap_108	NS008	France, 2015	TRIPTIC project	–	–	XII - cacoeciae-embryophagum	MG932229	B	
Hap_109	NS007	France, 2015	TRIPTIC project	–	–	XIII - swC	MG932230	C	
Hap_112	FPV034_A	France, 2015	TRIPTIC project	–	–	VIII - semblidis	MG932231	C	
Hap_113	ISA1729	France, 2015	TRIPTIC project	–	–	II - principium	MG932232	C	
Hap_114	AL004	France, 2015	TRIPTIC project	–	–	VIII - semblidis	MG932233	B	
Hap_115	ISA2564	Kazakhstan, ?	TRIPTIC project	–	–	3 - swA	MG932234	B	
Hap_116	PUG191	France, 2015	TRIPTIC project	–	–	III - cordubensis	MG932235	B	varprot
Hap_117	ESP227	Spain, 2016	TRIPTIC project	–	–	XII - cacoeciae-embryophagum	MG932236	C	
Hap_118	PMBIO2	France, 2014	Historic collections	BP	<i>T. cacoeciae</i>	XII - cacoeciae-embryophagum	MG932237	C	
Hap_119	ISA4737	France, 2016	TRIPTIC project	–	–	XII - cacoeciae-embryophagum	MG932238	D	
Hap_120	ISA4561	France, 2016	TRIPTIC project	–	–	XII - cacoeciae-embryophagum	MG932239	B	
Hap_121	ISA3082	France, 2015	TRIPTIC project	–	–	XIII - swC	MG932240	C	
Hap_122	ESP462	Spain, 2016	TRIPTIC project	–	–	I - daumalae-evanescens	MG932241	B	
Hap_124	BSEM-S	France, 2017	TRIPTIC project	–	–	VI - brassicae-euproctidis	MG932242	C	
Hap_125	ISA5169	France, 2015	TRIPTIC project	–	–	VI - brassicae-euproctidis	MG932243	A	
Hap_126	FLO241c	France, 2016	TRIPTIC project	–	–	III - cordubensis	MG932244	B	
Hap_127	ISA4375	France, 2016	TRIPTIC project	–	–	III - cordubensis	MG932245	D	
Hap_128	ISA6646	France, 2016	TRIPTIC project	–	–	XI - bourarachae	MG932246	C	
Hap_129	ISA5544	France, 2016	TRIPTIC project	–	–	XI - bourarachae	MG932247	B	
Hap_130	JAP0029B	Japan, 2016	TRIPTIC project	–	–	not affiliated	MG932248	B	varprot

Hap_131	JAP0027a	Japan, 2016	TRIPTIC project	-	-	IV - chilonis	MG932249	B	
Hap_132	ISA6223	France, 2016	TRIPTIC project	-	-	V - misB	MG932250	C	varprot
Hap_133	ISA6147	France, 2016	TRIPTIC project	-	-	V - misB	MG932251	C	
Hap_134	ISA5169	France, 2015	TRIPTIC project	-	-	VI - brassicae-euproctidis	MG932252	A	
Hap_135	COR7	France (Corsica), 2016	TRIPTIC project	-	-	XII - cacoeciae-embryophagum	MG932253	A	
Hap_136	ISA7850	France, 2016	TRIPTIC project	-	-	V - misB	MG932254	C	
Hap_137	JAP0018	Japan, 2016	TRIPTIC project	-	-	5 - swD	MG932255	B	
Hap_138	BL117	France, 2016	TRIPTIC project	-	-	IX - swE	MG932256	B	
Hap_139	BL116	France, 2016	TRIPTIC project	-	-	IX - swE	MG932257	C	
Hap_140	ISA10438	France, 2016	TRIPTIC project	-	-	XI - bourarachae	MG932258	D	
Hap_141	ISA10438_R1.3	France, 2016	TRIPTIC project	-	-	XI - bourarachae	MG932259	A	
Hap_142	ISA6151_R2.6	France, 2016	TRIPTIC project	-	-	XII - cacoeciae-embryophagum	MG932260	A	
Hap_143	ISA11333	France, 2016	TRIPTIC project	-	-	III - cordubensis	MG932261	A	
Hap_145	GOT0098	France, 2016	TRIPTIC project	-	-	XII - cacoeciae-embryophagum	MG932262	A	
Hap_146	MURU0025	France, 2016	TRIPTIC project	-	-	I - daumalae-evanescens	MG932263	A	
Hap_147	ISA4379	France, 2016	TRIPTIC project	-	-	I - daumalae-evanescens	MG932264	A	varprot
Hap_148	ISA13075	France, 2016	TRIPTIC project	-	-	not affiliated	MG932265	A	
Hap_149	ISA13075	France, 2016	TRIPTIC project	-	-	not affiliated	MG932266	A	
Hap_150	ISA12949	France, 2016	TRIPTIC project	-	-	XII - cacoeciae-embryophagum	MG932267	A	
Hap_151	ISA13345	France, 2016	TRIPTIC project	-	-	VII - misA	MG932268	B	
Hap_152	ISA11967	France, 2016	TRIPTIC project	-	-	XI - bourarachae	MG932269	C	
Hap_153	ISA11965	France, 2016	TRIPTIC project	-	-	XI - bourarachae	MG932270	A	varprot
Hap_154	ISA13470	France, 2016	TRIPTIC project	-	-	V - misB	MG932271	B	
Hap_155	ISA12366	France, 2016	TRIPTIC project	-	-	V - misB	MG932272	C	varprot
Hap_157	ISA6926	France, 2016	TRIPTIC project	-	-	I - daumalae-evanescens	MG932273	A	
Hap_158	ISA11783	France, 2016	TRIPTIC project	-	-	XII - cacoeciae-embryophagum	MG932274	A	
Hap_159	ISA12953	France, 2016	TRIPTIC project	-	-	XII - cacoeciae-embryophagum	MG932275	B	
Hap_160	ISA13034	France, 2016	TRIPTIC project	-	-	III - cordubensis	MG932276	C	varprot
Hap_162	ISA15036	France, 2016	TRIPTIC project	-	-	XII - cacoeciae-embryophagum	MG932277	A	
Hap_163	ISA15214	France, 2016	TRIPTIC project	-	-	III - cordubensis	MG932278	C	
Hap_164	ISA8699	France, 2016	TRIPTIC project	-	-	I - daumalae-evanescens	MG932279	B	varprot
Hap_165	ISA10054	France, 2016	TRIPTIC project	-	-	III - cordubensis	MG932280	A	varprot
Hap_172	BL110	France, 2016	TRIPTIC project	-	-	VIII - semblidis	MG932281	B	
Hap_173	FLO218	France, 2016	TRIPTIC project	-	-	XII - cacoeciae-embryophagum	MG932282	A	

Hap_175	ISA11333_7	France, 2016	TRIPTIC project	-	-	not affiliated	MG932283	A
Hap_176	ISA11333_14	France, 2016	TRIPTIC project	-	-	not affiliated	MG932284	A
Hap_177	ISA7851_9	France, 2016	TRIPTIC project	-	-	XII - cacoeciae-embryophagum	MG932285	A
Hap_178	ISA7851_11	France, 2016	TRIPTIC project	-	-	not affiliated	MG932286	A
Hap_179	ISA11365_2	France, 2016	TRIPTIC project	-	-	XII - cacoeciae-embryophagum	MG932287	A
Hap_180	ISA11365_5	France, 2016	TRIPTIC project	-	-	not affiliated	MG932288	A
Hap_181	ISA11365_6	France, 2016	TRIPTIC project	-	-	not affiliated	MG932289	A
Hap_182	ISA11365_8	France, 2016	TRIPTIC project	-	-	not affiliated	MG932290	A
Hap_183	ISA13075.3	France, 2016	TRIPTIC project	-	-	not affiliated	MG932291	A
Hap_184	ISA13075.5	France, 2016	TRIPTIC project	-	-	not affiliated	MG932292	A
Hap_185	ISA13075.9	France, 2016	TRIPTIC project	-	-	not affiliated	MG932293	A
Hap_186	ISA13075.13	France, 2016	TRIPTIC project	-	-	not affiliated	MG932294	A
Hap_187	ISA10468.2	France, 2016	TRIPTIC project	-	-	XII - cacoeciae-embryophagum	MG932295	A
Hap_188	ISA6414.11	France, 2016	TRIPTIC project	-	-	not affiliated	MG932296	A
Hap_190	ISA6414	France, 2016	TRIPTIC project	-	-	not affiliated	MG932297	A
Hap_191	ISA15057.8	France, 2016	TRIPTIC project	-	-	not affiliated	MG932298	A
Hap_192	ISA17526	France, 2017	TRIPTIC project	-	-	V - misB	MG932299	A
Hap_193	ISA17526	France, 2017	TRIPTIC project	-	-	V - misB	MG932300	B
Hap_194	ISA17526	France, 2017	TRIPTIC project	-	-	not affiliated	MG932301	A
Hap_195	ISA18485	France, 2017	TRIPTIC project	-	-	XI - bourarachae	MG932302	A
Hap_196	Pch1	Algeria, 2017	Other source	-	-	XII - cacoeciae-embryophagum	MG932303	B
Hap_197	Pch1	Algeria, 2017	Other source	-	-	XII - cacoeciae-embryophagum	MG932304	B
Hap_198	Pch2	Algeria, 2017	Other source	-	-	XII - cacoeciae-embryophagum	MG932305	B

Tableau 2 : Souches utilisées pour l'analyse de phylogénie multilocus

Les haplotypes COI et l'identification moléculaire ont été obtenus à partir des analyses de barcoding (cf.-§3-Barcoding) (voir tableau 1 (voir §7) ainsi que les figures 3 et 4). Les résultats de la recherche de *Wolbachia* par PCR sont indiqués par W+ en cas de présence, W- en cas d'absence et W ? en cas d'incertitude.

Souches	Nb d'individus	Origine de la souche	Source	Barcode COI	Identification moléculaire	<i>Wolbachia</i>
2	6	Yougoslavie, 1972	Collections historiques	Hap_003	9 - oleae	
101	1	France, 1976	Collections historiques	Hap_064	XII - cacoeciae-embryophagum	
2Pint	5	Yougoslavie, 1972	Collections historiques	Hap_003	9 - oleae	
2Pint (contamination)	1	Yougoslavie, 1972	Collections historiques	Hap_002	I - daumalae-evanescens	
81a	4	France, 1975	Collections historiques	Hap_001	II - principium	
81b	4	France, 1975	Collections historiques	Hap_001	II - principium	
A_aff	outgroup			outgroup	outgroup	
A_sid	outgroup			outgroup	outgroup	
A1	1	France, 1989	Collections historiques	Hap_011	I - daumalae-evanescens	
ACJYR0002_cord	2	France, 2015	Projet TripTIC	Hap_012	III - cordubensis	
ACJYR0002_cac	1	France, 2015	Projet TripTIC	Hap_006	XII - cacoeciae-embryophagum	
ACJYR0010	1	France, 2015	Projet TripTIC	Hap_019	XII - cacoeciae-embryophagum	
ACJYR0014_cord	1	France, 2015	Projet TripTIC	Hap_012	III - cordubensis	
ACJYR0014_cac	1	France, 2015	Projet TripTIC	Hap_006	XII - cacoeciae-embryophagum	
ACJYR0030	1	France, 2015	Projet TripTIC	Hap_017	VIII - semblidis	
ACJYR0039	1	France, 2015	Projet TripTIC	Hap_006	XII - cacoeciae-embryophagum	
ACJYR0116	2	France, 2015	Projet TripTIC	Hap_106	XII - cacoeciae-embryophagum	
ACJYR0121	1	France, 2015	Projet TripTIC	Hap_107	XII - cacoeciae-embryophagum	
ACJYR0127	1	France, 2015	Projet TripTIC	Hap_006	XII - cacoeciae-embryophagum	
AF007	1	France, 2015	Projet TripTIC	Hap_021	I - daumalae-evanescens	
AL004	2	France, 2015	Projet TripTIC	Hap_114	VIII - semblidis	
A122	3	France, 1990	Collections historiques	Hap_004	VII - misA	
AM002	1	France, 2015	Projet TripTIC	Hap_021	I - daumalae-evanescens	
B2.1	1	Iran, 2013?	BIOLINE Agrosociences	Hap_036	VI - brassicae-euproctidis	
B3.1	3	Italie, 2013?	BIOLINE Agrosociences	Hap_028	VI - brassicae-euproctidis	
B5.3	1	France, 2013?	BIOLINE Agrosociences	Hap_036	VI - brassicae-euproctidis	

B5.6	1	France, 2013?	BIOLINE Agrosociences	Hap_037	VI - brassicae-euproctidis	
B9.1	1	Moldavia, 2013	BIOLINE Agrosociences	Hap_022	VI - brassicae-euproctidis	
B9.10	1	France, 2013?	BIOLINE Agrosociences	Hap_027	VI - brassicae-euproctidis	
B9.11	2	France, 2013	BIOLINE Agrosociences	Hap_005	VI - brassicae-euproctidis	
B9.3	2	Croatia, 2013	BIOLINE Agrosociences	Hap_032	VI - brassicae-euproctidis	
B9.30	2	Moldavia, 2013	BIOLINE Agrosociences	Hap_028	VI - brassicae-euproctidis	
B9.4	3	Croatia, 2013	BIOLINE Agrosociences	Hap_005	VI - brassicae-euproctidis	
B9.40	2	Croatia, 2013	BIOLINE Agrosociences	Hap_024	VI - brassicae-euproctidis	
B9.55	1	Croatia, 2013	BIOLINE Agrosociences	Hap_030	VI - brassicae-euproctidis	
B9.60	1	Croatia, 2013	BIOLINE Agrosociences	Hap_034	VI - brassicae-euproctidis	
B9.61	4	Croatia, 2013	BIOLINE Agrosociences	Hap_035	VI - brassicae-euproctidis	
B9.62	1	Croatia, 2013	BIOLINE Agrosociences	Hap_030	VI - brassicae-euproctidis	
B9.7	1	Hongrie, 2013	BIOLINE Agrosociences	Hap_036	VI - brassicae-euproctidis	
BL012	1	France, 2015	Projet TriPTIC	Hap_011	I - daumalae-evanescens	
BL103a	1	France, 2016	Projet TriPTIC	Hap_049	I - daumalae-evanescens	
BL113c	2	France, 2016	Projet TriPTIC	Hap_049	I - daumalae-evanescens	
BL116	1	France, 2016	Projet TriPTIC	Hap_139	IX - swE	W-
BL117	3	France, 2016	Projet TriPTIC	Hap_138 ; Hap_139	IX - swE	W-
BL121	3	France, 2016	Projet TriPTIC	Hap_045	I - daumalae-evanescens	
BOU2	2	Morocco, 1992	Collections historiques	Hap_085	XI - bourarachae	
BRA3	1	Moldavia, 1973	Collections historiques	Hap_005	VI - brassicae-euproctidis	
BSEM-S	3	France, 2017	Projet TriPTIC	Hap_124	VI - brassicae-euproctidis	W+
CHI1_Mart	1	France (Martinique), 2014	Autre source	Hap_061	IV - chilonis	
CHI2_Mart	2	France (Martinique), 2014	Autre source	Hap_061	IV - chilonis	
Citron	3	France, 2012	Collections historiques	Hap_006	XII - cacoeciae-embryophagum	
D2	1	China, 1992	Collections historiques	Hap_043	X - dendrolimi	
Dan97	1	Danemark, 1997	Collections historiques	Hap_006	XII - cacoeciae-embryophagum	
DB	1	Portugal, 1991	Collections historiques	Hap_014	VI - brassicae-euproctidis	
Diaph 1-1.17/05	1	Guyane, 2017	Autre source	Hap_221	b - swK	W-?
E_ur3	outgroup			outgroup	outgroup	

E_ur8	outgroup			outgroup	outgroup	
E1.3	2	France, 2013	BIOLINE Agrosociences	Hap_038	VI - brassicae-euproctidis	
E1.7	2	France, 2013	BIOLINE Agrosociences	Hap_039	VI - brassicae-euproctidis	
E8.6	1	France, 2013?	BIOLINE Agrosociences	Hap_039	VI - brassicae-euproctidis	
EgyptEupro	6	Egypte, ?	Collections historiques	Hap_068	VI - brassicae-euproctidis	
ESP227	2	Espagne, 2016	Projet TripTIC	Hap_117	XII - cacoeciae-embryophagum	
ESP359	2	Espagne, 2016	Projet TripTIC	Hap_117	XII - cacoeciae-embryophagum	
ESP459	1	Espagne, 2016	Projet TripTIC	Hap_049	I - daumalae-evanescens	
ESP462	3	Espagne, 2016	Projet TripTIC	Hap_122	I - daumalae-evanescens	
ESP466	1	Espagne, 2016	Projet TripTIC	Hap_049	I - daumalae-evanescens	
Exi3	1	USA, 1995	Collections historiques	Hap_009	7 - exiguum 1	
FLO201	2	France, 2016	Projet TripTIC	Hap_121	XIII - swC	
FPV034_A	1	France, 2015	Projet TripTIC	Hap_112	VIII - semblidis	
GOT0098	2	France, 2016	Projet TripTIC	Hap_076	XII - cacoeciae-embryophagum	
Grab_haplo10	1	France, 2011	Autre source	Hap_078	XII - cacoeciae-embryophagum	
Grab_haplo3	1	France, 2011	Autre source	Hap_073	V - misB	
Grab_haplo4	1	France, 2011	Autre source	Hap_074	I - daumalae-evanescens	
Grab_haplo7	1	France, 2011	Autre source	Hap_075	XII - cacoeciae-embryophagum	
IRAN01	1	Iran, 2012	Autre source	Hap_028	VI - brassicae-euproctidis	
IRAN02	1	Iran, 2014	Autre source	Hap_057	3 - swA	
IRAN03	1	Iran, 2012	Autre source	Hap_028	VI - brassicae-euproctidis	
IRAN04	1	Iran, 2014	Autre source	Hap_008	I - daumalae-evanescens	
ISA11030	1	France, 2016	Projet TripTIC	Hap_077	XII - cacoeciae-embryophagum	
ISA11967	2	France, 2016	Projet TripTIC	Hap_152	XI - bourarachae	
ISA12953	2	France, 2016	Projet TripTIC	Hap_159	XII - cacoeciae-embryophagum	W+?
ISA13345	2	France, 2016	Projet TripTIC	Hap_151	VII - misA	
ISA13470	2	France, 2016	Projet TripTIC	Hap_154	V - misB	
ISA19249_B5Sm_s_220517_	1	France, 2017	Autre source	Hap_004	VII - misA	
ISA20016_Pch2	3	Algérie, 2017	Autre source	Hap_196 ; Hap_197 ; Hap_198	XII - cacoeciae-embryophagum	
ISA2564	1	Kazakhstan, ?	Autre source	Hap_115	3 - swA	W+?

ISA2565	1	Russie, ?	Autre source	Hap_006	XII - cacaoeciae-embryophagum	W?
ISA3082	2	France, 2015	Projet TripTIC	Hap_121	XIII - swC	W-
ISA4561	2	France, 2016	Projet TripTIC	Hap_120	XII - cacaoeciae-embryophagum	
ISA4737	2	France, 2016	Projet TripTIC	Hap_119	XII - cacaoeciae-embryophagum	
ISA5544	3	France, 2016	Projet TripTIC	Hap_129	XI - bourarachae	W-
ISA6147	3	France, 2016	Projet TripTIC	Hap_133	V - misB	
ISA6223	3	France, 2016	Projet TripTIC	Hap_132	V - misB	
ISA6646	2	France, 2016	Projet TripTIC	Hap_128	XI - bourarachae	
ISA7850	3	France, 2016	Projet TripTIC	Hap_136	V - misB	
it18	1	Italy, ?	Collections historiques	Hap_063	X - dendrolimi	
JAP0018	2	Japan, 2016	Projet TripTIC	Hap_137	5 - swD	W-
JAP0029B	3	Japan, 2016	Projet TripTIC	Hap_130	not affiliated	W-
K1	2	Turquie, 2013	Autre source	Hap_070	VI - brassicae-euproctidis	
K2	2	Turquie, 2013	Autre source	Hap_070	VI - brassicae-euproctidis	
K4	2	Turquie, 2013	Autre source	Hap_070	VI - brassicae-euproctidis	
M1	1	France, 1993	Collections historiques	Hap_021	I - daumalae-evanescens	
M172	1	Uruguay, 1993	Collections historiques	Hap_067	6 - exiguum 2	
M36	1	France, 1982	Collections historiques	Hap_012	III - cordubensis	
MK	1	Morocco, 1991	Collections historiques	Hap_059	VI - brassicae-euproctidis	
MURU0222	1	France, 2016	Projet TripTIC	Hap_049	I - daumalae-evanescens	
NS007	1	France, 2015	Projet TripTIC	Hap_109	XIII - swC	
NS008	1	France, 2015	Projet TripTIC	Hap_108	XII - cacaoeciae-embryophagum	W+
OMAN	1	Oman, 1998	Collections historiques	Hap_068	VI - brassicae-euproctidis	
PB	1	Portugal, 1991	Collections historiques	Hap_014	VI - brassicae-euproctidis	
PKAN00093	1	France, 2017	Autre source	Hap_205	c - swJ	W-
PKAN00104	1	France, 2016	Autre source	Hap_206	c - swJ	W-
PKAN00116	1	France, 2016	Autre source	Hap_209	c - swJ	W-
PKAN00121	1	France, 2016	Autre source	Hap_214	c - swJ	W-
PKAN00151	1	France, 2017	Autre source	Hap_207	c - swJ	W-
PKAN00156	1	France, 2017	Autre source	Hap_208	c - swJ	W-

PKAN00157	1	France, 2017	Autre source	Hap_212	a - swI	W-
PKAN00081	1	France, 2016	Autre source	Hap_220	c - swJ	
PM	3	?	Collections historiques	Hap_003	9 - oleae	
PM_carpo	2	?	Collections historiques	Hap_003	9 - oleae	
PMBio1	3	France, 2014	Collections historiques	Hap_019	XII - cacoeciae-embryophagum	
PMBio2	6	France, 2014	Collections historiques	Hap_006 ; Hap_118	XII - cacoeciae-embryophagum	
PMC	2	France, 2014	Collections historiques	Hap_003	9 - oleae	
PR002	1	France, 2015	Projet TripTIC	Hap_027	VI - brassicae-euproctidis	
PR003	1	France, 2015	Projet TripTIC	Hap_036	VI - brassicae-euproctidis	
PR006	1	France, 2015	Projet TripTIC	Hap_036	VI - brassicae-euproctidis	
PR007	1	France, 2015	Projet TripTIC	Hap_017	VIII - semblidis	
PR009	1	France, 2015	Projet TripTIC	Hap_036	VI - brassicae-euproctidis	
PR010	1	France, 2015	Projet TripTIC	Hap_005	VI - brassicae-euproctidis	
PRET_FREDON_PTA	1	France (Martinique), 2014	Autre source	Hap_013	XIV - pretiosum	W-
PRET_FREDON_PTT	1	France (Martinique), 2014	Autre source	Hap_013	XIV - pretiosum	W+
PUG029	1	France, 2015	Projet TripTIC	Hap_096	III - cordubensis	
Rh2	5	France, 1982	Collections historiques	Hap_003	9 - oleae	
Rh4	4	France, 1982	Collections historiques	Hap_003	9 - oleae	
Rh9	3	France, 1982	Collections historiques	Hap_003	9 - oleae	
S2	1	Yougoslavia, 1972	Collections historiques	Hap_003	9 - oleae	
SALA033	2	France, 2016	Projet TripTIC	Hap_118	XII - cacoeciae-embryophagum	
SAMAH_ISA849	1	Tunisie, 2015	Autre source	Hap_006	XII - cacoeciae-embryophagum	
SAMAH_ISA850	1	Tunisie, 2015	Autre source	Hap_006	XII - cacoeciae-embryophagum	
SAMAH_ISA876	2	Tunisie, 2015	Autre source	Hap_059	VI - brassicae-euproctidis	
SEM1	1	France, 1998	Collections historiques	Hap_017	VIII - semblidis	
SEMA2	1	France, 2002	Collections historiques	Hap_017	VIII - semblidis	
SEMV	1	France, 1998	Collections historiques	Hap_016	XIII - swC	
Syr96D	1	Uruguay, 1993	Collections historiques	Hap_010	II - principium	
T_achaeae_biotop	2	?	BIOLINE Agrosociences	Hap_083	2 - achaeae	W-
T115	1	Uruguay, 1988	Collections historiques	Hap_069	XIV - pretiosum	

T191	1	Uruguay, 1993	Collections historiques	Hap_003	9 - oleae
T234	1	Uruguay, 1994	Collections historiques	Hap_069	XIV - pretiosum
Tbra	3	?	Collections historiques	Hap_005	VI - brassicae-euproctidis
TCAI	4	France, 1987-1989	Collections historiques	Hap_006	XII - cacoeciae-embryophagum
TCMz	3	France, 1987-1989	Collections historiques	Hap_007	XII - cacoeciae-embryophagum
TCVigne	3	France, 1987-1989	Collections historiques	Hap_006	XII - cacoeciae-embryophagum
Tdaum	9	France, ?	Collections historiques	Hap_002 ; Hap_008	I - daumalae-evanescens
TEAI	2	France, ?	Collections historiques	Hap_003	9 - oleae
TP41EG	1	Egypt, 2002	Collections historiques	Hap_015	XI - bourarachae
TP63	1	Egypt, 2004	Collections historiques	Hap_012	III - cordubensis
TR	1	Chili, 1994	Collections historiques	Hap_055	4 - nerudai
Tun704	4	Tunisie, ?	Collections historiques	Hap_006	XII - cacoeciae-embryophagum
URO2	1	Iran, 1996	Collections historiques	Hap_065	XII - cacoeciae-embryophagum W+
URO3	1	Iran, 1996	Collections historiques	Hap_066	XII - cacoeciae-embryophagum W+
V11.14	1	France, 2013	BIOLINE Agrosciences	Hap_044	I - daumalae-evanescens
V11.16	1	France, 2013	BIOLINE Agrosciences	Hap_046	I - daumalae-evanescens
V2.11	1	Moldavia, 2013	BIOLINE Agrosciences	Hap_041	I - daumalae-evanescens
V2.12	2	Moldavia, 2013	BIOLINE Agrosciences	Hap_044	I - daumalae-evanescens
V2.13	3	Moldavia, 2013	BIOLINE Agrosciences	Hap_042	I - daumalae-evanescens
V2.14	3	Moldavia, 2013	BIOLINE Agrosciences	Hap_044	I - daumalae-evanescens
V2.16	2	Moldavia, 2013	BIOLINE Agrosciences	Hap_045	I - daumalae-evanescens
V2.17	2	Moldavia, 2013	BIOLINE Agrosciences	Hap_021	I - daumalae-evanescens
V2.19	1	Moldavia, 2013	BIOLINE Agrosciences	Hap_011	I - daumalae-evanescens
V2.6	1	Spain (Canarias), ?	BIOLINE Agrosciences	Hap_053	I - daumalae-evanescens
V2.8	2	Egypt, ?	BIOLINE Agrosciences	Hap_054	I - daumalae-evanescens
V3.19	3	France, ?	BIOLINE Agrosciences	Hap_050 , Hap_051	I - daumalae-evanescens
V3.6	2	France, ?	BIOLINE Agrosciences	Hap_049	I - daumalae-evanescens
V3.8	3	France, ?	BIOLINE Agrosciences	Hap_047	I - daumalae-evanescens
V6.3	2	France, ?	BIOLINE Agrosciences	Hap_048	I - daumalae-evanescens
V7.2	1	France, ?	BIOLINE Agrosciences	Hap_008	I - daumalae-evanescens

V9.2	5	France, 2013	BIOLINE Agrosciences	Hap_011 ; Hap_056	I - daumalae-evanescens
V9.4	1	France, 2013	BIOLINE Agrosciences	Hap_021	I - daumalae-evanescens
Var	2	France, 2012?	Autre source	Hap_019	XII - cacoeciae-embryophagum

8 SUPPLEMENTS :

8.1 Suppléments S1 : Résultats généraux liés à l'ensemble des croisements

I- Généralités :

En complément des tests réalisés à l'intérieur des clusters « I-daumalae-evanescens » et « VI-brassicae-euproctidis » présentés dans le paragraphe 3 – Barcoding COI et du cluster V-misB § 4-Multilocus, les (in)compatibilités de reproduction testées entre les différents clusters présentent des résultats généralement concordants avec les délimitations d'espèces proposées par la phylogénie multilocus. Au total, 1375 tests de croisements ont été réalisés sur des souches disponibles et sexuées dont 7 espèces (ou complexes) ont été identifiées comme *T. « daumalae-evanescens »*, *T.principium*, *T.chilonis*, *T. « brassicae-euproctidis »*, *T.semblidis*, *T.bourarachae*, *T.achaeae*, et 4 espèces non identifiées *T.misA*, *T.misB*, *T.swC*, *T.swD*. Certaines espèces n'ont pas été étudiées puisqu'il s'agissait de souches non sexuées *T.cordubensis*, *T.caceociae-embryophagum*, *T.pretiosum*, *T.oleae* ou non disponibles au laboratoire *T.dendrolimi*, *T.ostrinae*, *T.nerudai*, *T.exiguum*, *T.danausicida*, *T.swA*, *T.swG*, *T.swI*, *T.swK*, *T.swJ*, *T.swB* et *T.swE*). Prises dans l'ensemble, toutes les femelles non fécondées, toutes espèces confondues (n=190) n'ont pas produit de filles. Les croisements parentaux ont été plus ou moins efficaces en fonction des espèces avec un succès de croisements (i) médiocre pour *T.swD* [RC1= 20%, RC2=0.01% - n=5], (ii) moyen pour *T.achaeae* [RC1=57%, RC2=0.35, n=21] et (ii) élevé [RC1 moyen de 93% (79-100%), RC2 moyen de 0.64% (35-86%) - n=217] pour les autres espèces *T.bourarachae*, *T.brassicae-euproctidis*, *T.chilonis*, *T.daumalae-evanescens*, *T.misA*, *T.misB*, *T.misC*, *T.principium*, *T.semblidis*, *T.swC*.

Concernant les tests de compatibilité, la répartition des croisements (n = 1324) à l'intérieur des classes déterminées dans le § Test de croisement 2.4.2 est représentée dans la Figure S2. Plus précisément, à l'intérieur de chaque catégorie, la répartition des croisements est la suivante :

- **les croisements compatibles** cf. Figure S2:

96,4% des croisements parentaux (n = 238) ;

69.0% (n = 116) des croisements intraspécifiques ;

6.7% (n = 44) des croisements interspécifiques qui concernent dans la majorité des cas le complexe « T.evanescens-daumalae » (T.eva x T.daum : n = 30) et le complexe « brassicae-euproctidis » (T.brass x T.brass : n= 10) vu dans la partie § 3-barcoding COI. Un dernier cas

de croisement interspécifique (n = 4) implique des souches *T.evanescens* et *T.brassicae* dont les résultats sont présentés dans § Interspécifique non conforme ci-dessous

- les croisements incompatibles cf. Figure S2 :

100% des femelles vierges (n = 191), conformément à ce qui est attendu ;

82.2% (n = 587) des croisements interspécifiques ;

6,9% (n = 14) des croisements intraspécifiques plus précisément du complexe « brassicae-euproctidis », ce complexe étant discuté dans le § 3-barcoding COI.

- les croisements moyennement compatibles cf. Figure S2 :

20,7% (n = 32) des croisements intraspécifiques dont principalement des croisements concernant le complexe « brassicae-euproctidis » (n = 22) discuté dans le § 3-barcoding COI. Les autres croisements intraspécifiques concernent les espèces *T.principium* (n = 5) et *T.evanescens* (n = 5) pourraient s'expliquer par des évitements comportementaux ou autres problèmes non identifiés ;

1.5% (n = 13) des croisements interspécifiques concernant des souches *T.misA* (1 souche) X *T.swD* (1 souche) (n = 7), *T.evanescens* (1 souche) X *T.brassicae* (1 souche) (n = 6) dont les résultats sont présentés dans le § Interspécifique non conforme ci-dessous.

- les croisements faiblement compatibles cf. Figure S2 :

3.6% des croisements parentaux (n = 5) concernant une souche de l'espèce *T.swD* [RC1= 20%, RC2=0.01% - n=5] qui pourraient s'expliquer par des conditions de développement défavorables (l'hôte de substitution), de la dépression de consanguinité, de l'incompatibilité liée à des symbiotes, une présence de pathogène, voire d'autres aspects non identifiés ;

3.4% (n = 10 ; 1♀i x 1♂j) des croisements intraspécifiques concernant une souche du complexe « brassicae-euproctidis », discuté dans le § 3-barcoding COI ;

9.6% (n = 74) des croisements interspécifiques concernent des souches du complexe « brassicae-euproctidis » (n= 26), des souches *T.misA* (1 souche) X *T.brassicae* (2 souches) (n = 22), et des souches *T.evanescens*(2 souches) X *T.brassicae* (3 souches) (n = 26) dont les résultats sont présentés dans le § Interspécifique non conforme ci-dessous.

II- Cas des croisements interspécifiques non conformes :

On peut distinguer deux types de croisements interspécifiques non conformes :

- les croisements impliquant la souche identifiée *T.misA* ;

- les croisements interspécifiques entre des souches de « *T. brassicae/euproctidis* » et « *T.evanescens/daumalae* » dans les deux sens de croisement.

Concernant le premier type de croisement, une souche *T.misA* (Al22) présentait une compatibilité moyenne avec une souche identifiée *T.swD* (JAP0018) (♀Al22 x ♂JAP0018 ; n = 7 ; RC1=0,43, RC2=0,25) dans un seul sens de croisement. Le sens inverse présentait une incompatibilité de reproduction conforme à ce qui était attendu. Cette modalité de croisement n'a pu être testée qu'à partir de ces deux souches seules représentantes de ces deux espèces. Les fils et filles restés en mélange ont produit des petites-filles la génération suivante (n=2, RC1=1 ; RC2=0,73). Les filles vierges n'ont pas été testées afin d'identifier une possible induction de thélytoquie.

Concernant les croisements entre la souche *T.misA* (Al22) et deux souches *T.brassicae* (PR010 et Tbra), ils présentaient des compatibilités faibles (Al22 x Tbra, n=12 ; Al22 x PR010, n=20) quel que soit le sens de croisement. Deux autres souches du complexe « *T.brassicae/euproctidis* » ont été croisées avec la souche *T.mis A* et ces croisements n'ont pas donné de filles dans les deux sens de croisement (♀al22 x ♂ brassicae/euproctidis, n=15 ; ♂al22x♀ brassicae/euproctidis , n=17).

Concernant les croisements entre « *T. brassicae/euproctidis* » et « *T.evanescens/daumalae* », (cf. Figure S3), des niveaux variables de (in)compatibilités ont été observés, tous croisements confondus (n = 83), avec un RC1 moyen de 0.20 [0% ; 100%] et un RC2 moyen de 0,87 [0% ;100].

Pour les croisements ont été étudiés à la génération suivante (voir S3 flèches bleues), aucune des filles vierges (n=16), ni aucune des filles mélangées aux fils (n=9) n'ont réussi à avoir de descendance (cf. Figure S3), tous croisements confondus. Dans notre cas, il semblerait que les filles obtenues soient stériles mais le nombre de souches testées et le nombre de répétition étaient faibles. Il est nécessaire de réaliser des expériences complémentaires afin de mettre réellement en évidence : soit une induction de thélytoquie déjà décrite par B.Pintureau (Pintureau and Babault, 1981), soit l'identification d'hybride non viable, soit l'obtention de femelles stériles.

Aucune des filles vierges (n=16), ni aucune des filles mélangées aux fils (n=9) n'ont réussi à avoir de descendance (cf. Figure S3), tous croisements confondus. Dans notre cas, il semblerait que les filles obtenues soient stériles mais le nombre de souches testées et le nombre de répétitions étaient faibles. Il est nécessaire de réaliser des expériences complémentaires afin de mettre en évidence soit une induction de thélytoquie déjà décrite par B.Pintureau (Pintureau and Babault 1981), soit l'identification d'hybride non viable, soit l'obtention de femelles stériles.

8.2 Suppléments S2 : Présence de filles par catégorie de croisement

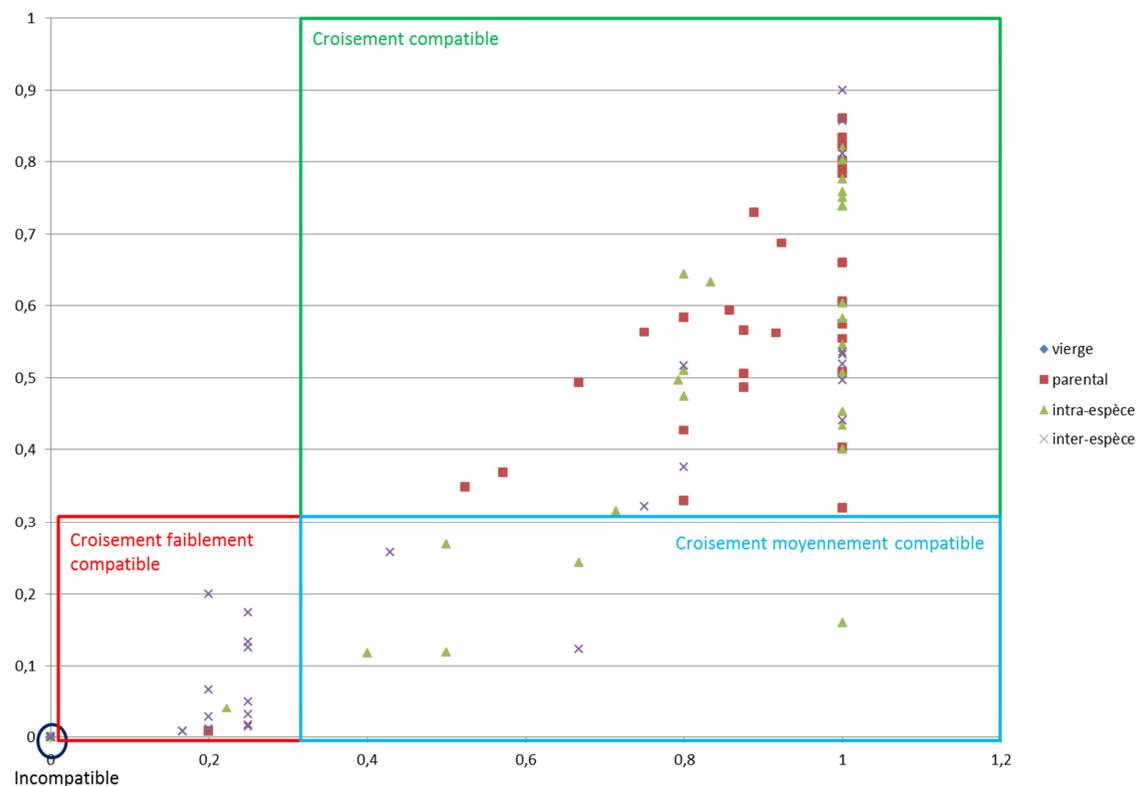


Figure S2 : Présence de filles par catégorie de croisement.

Les catégories de croisement (femelles vierges, parentaux, intraspécifiques et interspécifiques) sont représentées en fonction de leurs RC1 (proportion de croisements qui ont fourni au moins une fille) en abscisse par rapport à leurs RC2 (proportion moyenne de filles dans les croisements) en ordonnée. Les compatibilités de reproduction sont identifiées par des couleurs, vert pour les croisements compatibles, bleu pour les croisements moyennement compatibles, rouge pour les croisements faiblement compatibles et noir pour les croisements incompatibles.

8.3 Suppléments S3 : Résultats des croisements interspécifiques entre complexes

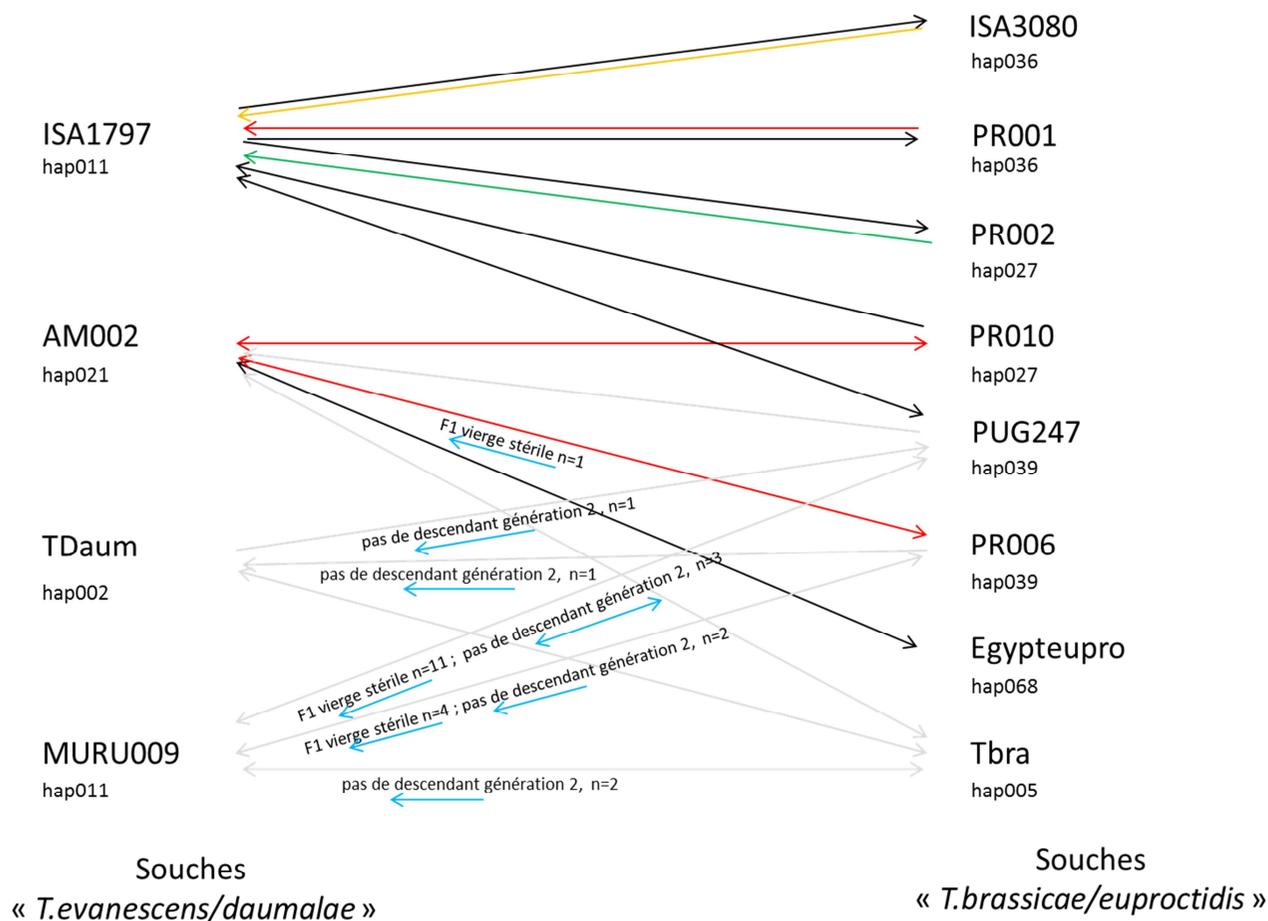


Figure S3 : Résultats des croisements interspécifiques entre le complexe « *T. brassicae/euproctidis* » et le complexe « *T. daumalae-evanescens* » sur deux générations.

Test de la G0 : Les compatibilités de reproduction de la G0 sont identifiées en fonction de la couleur de la flèche : vert pour les croisements compatibles, bleu pour les croisements moyennement compatibles, rouge pour les croisements faiblement compatibles et noir pour les croisements incompatibles (voir Figure S2).

Les flèches grises indiquent les croisements G0 ayant eu au moins une descendance fille F1 mais non quantifiée, utilisée pour les tests de croisements de la génération suivante.

Test des filles F1 : Les flèches bleues indiquent les croisements réalisés en F1 avec le résultat de la fécondité des F1 au-dessus et les n répétitions.

Dans tous les cas, la flèche pointe sur la femelle du couple.

8.4 Autres suppléments

Compte tenu de leur nombre et de leurs tailles, les autres figures sont disponibles sur FileSender via le lien :

<https://filesender.renater.fr/?s=download&token=3736ae92-d701-b490-4893-92e92fd8ee6c> jusqu'au 11/10/2018.

Seules les légendes sont indiquées ici.

Figure S4 : Arbre ML sur l'alignement du gène *F2-Ef1alpha*

Arbre ML basé sur les données de *Trichogramma* obtenu à partir de l'alignement ClustalW de l'ensemble de données *F2-Ef1alpha*. Les supports BP (100 répliqués) des clades sont indiqués aux nœuds en pourcentage.

Figure S5 : Arbre RaxML sur l'alignement du gène *Wingless*

Arbre ML basé sur les données de *Trichogramma* obtenu à partir de l'alignement ClustalW de l'ensemble de données *Wingless*. Les supports BP (100 répliqués) des clades sont indiqués aux nœuds en pourcentage.

Figure S6 : Arbre RaxML sur l'alignement du gène *COI*

Arbre ML basé sur les données de *Trichogramma* obtenu à partir de l'alignement ClustalW de l'ensemble de données *COI*. Les supports BP (100 répliqués) des clades sont indiqués aux nœuds en pourcentage.

Figure S7 : Arbre ML sur l'alignement du gène *CytB*

Arbre ML basé sur les données de *Trichogramma* obtenu à partir de l'alignement ClustalW de l'ensemble de données *CytB*. Les supports BP (100 répliqués) des clades sont indiqués aux nœuds en pourcentage.

Figure S8 : Arbre RaxML sur l'alignement du gène *RpS4*

Arbre ML basé sur les données de *Trichogramma* obtenu à partir de l'alignement Muscle de l'ensemble de données *RpS4*. Les supports BP (100 répliqués) des clades sont indiqués aux nœuds en pourcentage.

Figure S9 : Arbre RaxML sur la concaténation de 2 gènes mitochondriaux (*COI* et *CytB*) avec 2 partitions

Arbre ML basé sur les données de *Trichogramma* obtenu à partir de l'alignement ClustalW de l'ensemble de données *COI*, *CytB* combinées avec 2 partitionnements, un par gène. Les arbres ML correspondant par gène sont donnés sur les Figures S6 et S7. Les supports BP (100 répliqués) des clades sont indiqués aux nœuds en pourcentage. Les groupes renvoient aux clades discutés dans le §4-Multilocus voir aussi §3-Barcoding Figure 4.

Figure S10 : Arbre RaxML sur concaténation de 2 gènes mitochondriaux (*COI* et *CytB*) avec 6 partitions (PF2)

Arbre ML basé sur les données de *Trichogramma* obtenu à partir de l'alignement ClustalW de l'ensemble de données *COI* et *CytB* combinées sous le schéma de 6 partitions (*COI* position1 P1, *COI* P2, *COI*-P3, *CytB* P1, *CytB* P2, *CytB* P3). Les arbres ML correspondant par gène

sont donnés sur les Figures S6, S7. Les supports BP (100 répliqués) des clades sont indiqués aux nœuds en pourcentage.

Figure S11 : Arbre RaxML sur concaténation de 4 gènes nucléaires (*EF1alpha-F2*, *RpL27a*, *RpS4*, *Wingless*) avec 4 partitions

Arbre ML basé sur les données de *Trichogramma* obtenu à partir de l'alignement ClustalW de l'ensemble de données *EF1alpha-F2*, *RpL27a*, *RpS4*, *Wingless* combinées sous le schéma de 4 partitions, une pour chaque gène. Les arbres ML correspondant par gène sont donnés sur les Figures S4, S13, S8, S5. Les supports BP (100 répliqués) des clades sont indiqués aux nœuds en pourcentage. Les groupes renvoient aux clades discutés dans le §4-Multilocus voir aussi §3-Barcoding Figure 4.

Figure 12 : Arbre RaxML sur concaténation de 4 gènes nucléaires (*EF1alpha-F2*, *RpL27a*, *RpS4*, *Wingless*) avec 7 partitions (PF2)

Arbre ML basé sur les données de *Trichogramma* obtenu à partir de l'alignement ClustalW de l'ensemble de données *F2-EF1alpha*, *RpL27a*, *RpS4*, *Wingless* combinées sous le schéma de 7 partitions (*EF1alpha-F2* position1 P1, *EF1alpha-P2*, *EF1alpha-P3* + *Wingless P1*, *RpL27a*, *RpS4*, *Wingless P2*, *Wingless P3*). Les arbres ML correspondant par gène sont donnés sur les Figures S4, S13, S8, S5. Les supports BP (100 répliqués) des clades sont indiqués aux nœuds en pourcentage.

Figure S13 : Arbre RaxML sur l'alignement du gène *RpL27a*

Arbre ML basé sur les données de *Trichogramma* obtenu à partir de l'alignement Muscle de l'ensemble de données *RpL27a*. Les supports BP (100 répliqués) des clades sont indiqués aux nœuds en pourcentage.

Figure S14 : Arbre ML sur concaténation de 6 gènes *COI*, *CytB*, *EF1alpha-F2*, *RpL27a*, *RpS4*, *Wingless* avec 12 partitions (PF2)

Arbres ML basé sur les données de *Trichogramma* obtenu à partir de l'alignement ClustalW de l'ensemble de données *COI*, *CytB*, *F2-EF1alpha*, *RpL27a*, *RpS4*, *Wingless* combinées sous le schéma de 12 partitions. Les arbres ML correspondant par gène sont donnés sur les Figures S4, S5, S6, S7, S8, S13. Les supports BP (100 répliqués) des clades sont indiqués aux nœuds en pourcentage.

Figure S15 : Phylogénie MLST de *Wolbachia* présent chez *Trichogramma* à partir de 5 gènes (*FtsZ*, *CoxA*, *GatB*, *hcpA*, *fbpA*)

Phylogénie Multi Locus Sequence Typing utilisant du ML maximum likelihood, avec un modèle de substitution GTR General Time Reversible chez des *Wolbachia* de *Trichogramma*. Les supports BP (100 répliqués) des clades sont indiqués aux nœuds en pourcentage. L'arbre a été raciné sur des séquences concaténées de NCBI de *Wolbachia* présentes chez l'out-group *Brugia malayi*. Les séquences de *Wolbachia* chez *T. deion* et *T. brassicae* sont également des séquences de NCBI. Le numéro d'haplotype fait référence aux haplotypes déterminés dans le § 3 Barcoding (voir aussi Figure 4).

ÉCOLE PRATIQUE DES HAUTES ÉTUDES
SCIENCES DE LA VIE ET DE LA TERRE

Caractérisation moléculaire et isolements reproducteurs chez des auxiliaires de lutte biologique

WAROT Sylvie

22 novembre 2018

RÉSUMÉ

Le genre *Trichogramma* (Hymenoptera : Trichogrammatidae) comprend, à l'échelle mondiale, environ 200 espèces. Les caractéristiques biologiques de ces parasitoïdes oophages en font des taxons particulièrement étudiés pour des questions d'écologie évolutive ou des applications en lutte biologique. La taxonomie de ce genre reste cependant mal résolue et l'identification en routine de souches est difficile, du fait de limitations tant aux niveaux morphologiques que moléculaires.

Dans ce contexte, mon travail vise à aborder expérimentalement la question des « délimitations d'espèces » entre les espèces présentes dans la zone Ouest-Paléarctique en combinant : (i) une activité intense de « barcoding » moléculaire sur une portion de la Cytochrome oxydase 1 (>3000 individus séquencés) ; (ii) un travail de phylogénie multilocus incluant 2 gènes mitochondriaux (*Cytochrome oxydase 1* et *Cytochrome B*) et 4 gènes nucléaires, codants (*Wingless* et *Elongation Factor 1 alpha*) ou non (*Ribosomal protein L27a* et *Ribosomal protein S4*) ; (iii) la recherche d'endosymbiotes du genre *Wolbachia*, possiblement impliqués dans l'isolement reproducteur de certaines souches ; (iv) pour une trentaine de souches sexuées, la réalisation de tests de croisements en conditions contrôlées. Au final, les travaux de caractérisation moléculaire ont permis de mettre en évidence plus d'une vingtaine de clusters nettement différenciés qui incluent, pour la plupart, des espèces déjà décrites. Parmi ces clusters figurent notamment ce que nous avons considéré comme deux complexes d'espèces, « brassicae-euproctidis » et « cacoeciae-embryophagum », chez lesquels ont été observés des sous-structurations moléculaires et/ou , des patrons complexes d'(in)compatibilités reproductives voire un polymorphisme du mode de déterminisme du sexe. Nos résultats mettent également en évidence une proximité moléculaire et une compatibilité reproductrice entre des souches initialement identifiées comme *T. daumalae* et *T. evanescens*, deux espèces pourtant classées dans deux sections différentes sur la base de caractères moléculaires. Pour le reste, les résultats issus des croisements contrôlés confortent bien les délimitations observées sur des bases moléculaires. Contrairement à d'autres hyménoptères parasitoïdes, les infections à *Wolbachia* semblent relativement rares, le seul effet identifié étant l'induction de la thélytoquie. D'une façon générale, ce travail de caractérisation intégrative a vocation à (i) être complété par des travaux de caractérisation morphologique, (ii) être progressivement étendu à d'autres espèces, (iii) servir de base pour développer des recherches futures, à vocations académiques (analyses comparatives) ou plus finalisées (identification ou optimisation d'auxiliaires de lutte biologique).

MOTS-CLÉS : Barcoding, Délimitation d'espèces, Isolement reproducteur, Phylogénie, Taxonomie intégrative, *Trichogramma*, *Wolbachia*

Pinto 1991

Pinto, J.D., Stouthamer, R., Platner, G.R., and Oatman, E.R. (1991). Variation in Reproductive Compatibility in *Trichogramma* and Its Taxonomic Significance (Hymenoptera: Trichogrammatidae). *Ann. Entomol. Soc. Am.* 84, 37–46.

Pintureau, B., and Babault, M. (1981). Caractérisation enzymatique de *Trichogramma evanescens* et de *T. Maidis* [Hym.: Trichogrammatidae]; Étude des hybrides. *Entomophaga* 26, 11–22.