



HAL
open science

Etude intégrative du statut des deux variants adaptatifs à la plante-hôte de *Spodoptera frugiperda* (Lepidoptera:Noctuidae)

Pascaline Dumas

► **To cite this version:**

Pascaline Dumas. Etude intégrative du statut des deux variants adaptatifs à la plante-hôte de *Spodoptera frugiperda* (Lepidoptera:Noctuidae). Sciences agricoles. Université Montpellier 2 (Sciences et Techniques), 2014. Français. NNT: . tel-02810985

HAL Id: tel-02810985

<https://hal.inrae.fr/tel-02810985>

Submitted on 6 Jun 2020

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

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

**UNIVERSITE MONTPELLIER II
SCIENCES ET TECHNIQUES DU LANGUEDOC**

THESE

Pour obtenir le diplôme de doctorat

**Formation doctorale : Biologie de l'évolution et écologie
Ecole Doctorale : Systèmes Intégrés en Biologie, Agronomie, Géosciences, Hydrosiences,
Environnement, SIBAGHE**

**Étude intégrative du statut des deux variants adaptatifs à la plante hôte
de *Spodoptera frugiperda* (Lepidoptera : Noctuidae)**

Présentée et soutenue publiquement par

Pascaline DUMAS

le 29 octobre 2013

Devant le jury composé de :

Emmanuelle d'ALENÇON, Chargée de recherche, INRA Montpellier (Directrice de Thèse)

Denis BOURGUET, Directeur de recherche, INRA Montpellier (Examineur)

Michel CUSSON, Chercheur scientifique, RNCAN Québec (Rapporteur)

Gaël KERGOAT, Chargé de recherche, INRA Montpellier (Invité, co-Directeur de Thèse)

Emmanuelle JOUSSELIN, Chargée de recherche, INRA Montpellier (Représentante Ecole Doctorale)

Jean-François SILVAIN, Directeur de recherche, CNRS Gif-sur-Yvette (Examineur)

Jean-Christophe SIMON, Directeur de recherche, INRA Rennes (Rapporteur)

**UNIVERSITE MONTPELLIER II
SCIENCES ET TECHNIQUES DU LANGUEDOC**

THESE

Pour obtenir le diplôme de doctorat

**Formation doctorale : Biologie de l'évolution et écologie
Ecole Doctorale : Systèmes Intégrés en Biologie, Agronomie, Géosciences, Hydrosiences,
Environnement, SIBAGHE**

**Étude intégrative du statut des deux variants adaptatifs à la plante hôte
de *Spodoptera frugiperda* (Lepidoptera : Noctuidae)**

Présentée et soutenue publiquement par

Pascaline DUMAS

le 29 octobre 2013

Devant le jury composé de :

Emmanuelle d'ALENÇON, Chargée de recherche, INRA Montpellier (Directrice de Thèse)

Denis BOURGUET, Directeur de recherche, INRA Montpellier (Examineur)

Michel CUSSON, Chercheur scientifique, RNCAN Québec (Rapporteur)

Gaël KERGOAT, Chargé de recherche, INRA Montpellier (Invité, co-Directeur de Thèse)

Emmanuelle JOUSSELIN, Chargée de recherche, INRA Montpellier (Représentante Ecole Doctorale)

Jean-François SILVAIN, Directeur de recherche, CNRS Gif-sur-Yvette (Examineur)

Jean-Christophe SIMON, Directeur de recherche, INRA Rennes (Rapporteur)

Financements :

Cette thèse a été financée pour moitié par l'INRA département SPE et par l'Université Montpellier II

Laboratoires de recherche :

- UMR 1333 Diversité, Génome et Interactions Microorganismes-Insectes (DGIMI)

Université Montpellier II
Bâtiment 24, 4^{ème} étage
Place Eugène Bataillon
34 095 Montpellier Cedex 5

- UMR 1062 Centre de Biologie pour la Gestion des Population (CBGP)

Campus International de Baillarguet
CS 30016
34 988 Montferrier-sur-Lez

Remerciements

Je tiens à remercier chacun des membres de ce jury, Denis Bourguet, Michel Cusson, Emmanuelle Jouselin, Jean-François Silvain et Jean-Christophe Simon pour avoir accepté de participer à la soutenance de cette thèse.

Je tiens également à remercier M. Philippe Fournier et son successeur M. Patrick Tailliez pour leur accueil au laboratoire DGIMI. Je remercie également leurs homologues du CBGP, M. Denis Bourguet puis Mme Flavie Vanlerberghe-Masutti pour m'avoir permis d'effectuer une partie de ma thèse dans leur laboratoire.

Emma, Gaël, merci à tous les deux pour votre encadrement pendant ces trois ans. Cela a été un vrai honneur d'être sous votre tutelle. Jusqu'au bout vous avez été présents et nos discussions m'ont beaucoup apportées, spécialement lors de ces derniers mois. Emma, je souhaite vous remercier du fond du cœur pour le soutien que vous m'avez apporté tout au long de cette thèse. Vous m'avez toujours fait confiance et je me suis toujours sentie à l'aise sous votre direction. Votre disponibilité, votre engagement et votre ouverture d'esprit sont un vrai modèle pour la future scientifique que je souhaiterais devenir. Gaël, toi aussi je te remercie sincèrement. Avec toi, c'était la découverte permanente : de nouveaux concepts, de nouvelles techniques, de nouveaux continents... Merci pour tes conseils toujours pertinents, pour ta rigueur et ton soutien. Merci de m'avoir poussée vers le haut.

Je souhaite également remercier toute l'équipe EHA pour leurs conseils et leur aide pendant cette thèse. Un merci tout spécial à Sylvie pour son encadrement à la paillasse et pour ses conseils techniques mais également pour tout le reste ! Merci à Philippe, François et Nicolas ainsi qu'à ma collègue de bureau Imène, la seule et unique ! Et bien sûr, merci à Marion.

Un très grand merci à tous les membres de l'insectarium et particulièrement à Clotilde et Gaëtan. Merci pour m'avoir formée aux élevages et pour avoir été toujours disponibles quand j'avais des soucis de boîtes ou de Poitout ! Sans votre travail, je n'aurais rien pu faire.

Même s'il n'y a pas d'équipe au CBGP, je tiens à remercier Anne-Laure pour son soutien technique et pas seulement ! Merci également pour m'avoir supportée en mission

poulette. Un merci très chaleureux à Anne et Bruno pour ne pas m'avoir laissée dans la forêt le jour où on s'est perdu...lors de la mission en Guyane.

Ces trois années de thèses se sont partagées entre deux laboratoires différents ; un pied au laboratoire DGIMI et l'autre au CBGP. Grâce à cela j'ai eu la chance de rencontrer deux fois plus de gens intéressants, qui à un moment ou l'autre de ma thèse m'ont aidée, conseillée, soutenue et beaucoup apporté. Je tiens donc à remercier toutes les personnes du laboratoire DGIMI et du CBGP.

Un grand merci aux « glougloux » de DGIMI : Sylvie, Amélie, Cécilia, Aurélie, Imène, Laure, Doriane, Anne-Sophie, Fanny, Véro, Marie-Elsie, Apolline, Marion. Nous avons beaucoup ri ensemble et vous m'avez apporté un soutien permanent tout au long de cette thèse. Je suis vraiment heureuse de vous avoir rencontrées, merci les filles ! Bien-sûr je ne vous oublie pas les garçons.

Un autre immense merci aux poulettes du CBGP : Anne-Laure, Joséphine, Aléna, Laure, Caro, Marie et Morgane (poulette de SupAgo !). Je vous remercie pour tout : votre accueil, le soutien moral et scientifique, les encouragements de ces derniers mois, les soirées filles, les conseils... Vous avez rendu ma vie à Montpellier encore plus agréable ! Merci à tous les autres du CBGP avec qui j'ai partagé des moments formidables autour d'une bière ou lors d'une partie de pétanque : Fabien, Jeff, Max et Virginie, Manu, Ashraf, Guillaume, Laurent, Julie, Réjane, Manue, Armelle, Alex ...

Il y a aussi les copains du Limousin et de la Dordogne, Poppy, Léa, Sakout, Clément, Mathieu, Villou merci pour votre amitié qui dure depuis si longtemps et merci à *Skype* de nous avoir donné la possibilité de nous parler tous les jours (ou presque) pendant ces trois ans qui ne nous ont pas permis de nous voir autant que nous l'aurions souhaité.

Et puis merci à maman, « ma papa » et Manie, merci pour votre soutien inconditionnel et pour vos encouragements. Vous m'avez manqué pendant ces 9 derniers mois, j'ai hâte de vous revoir enfin. Pour finir, merci à Paco, pendant cette thèse tu as été ma boussole, ma Grande Ourse. Avec toi, la vie est plus belle mais bon, « suffit la parlotte ! » Tout ce que je n'écris pas ici, tu le sais déjà.

Liste des Figures

Figure 1 : Représentation schématique de la notion d'espèce.

Figure 2 : Différentes terminologies des états de caractères en phylogénie.

Figure 3 : Exemples de cas d'espèces cryptiques, de convergence évolutive et de dimorphisme sexuel.

Figure 4 : Figure du *Coalescent* standard.

Figure 5 : Histogrammes du nombre de publication par année pour une recherche sur *Web of science* comprenant les mots *integrative taxonomy*.

Figure 6 : Représentation schématique des étapes théoriques de la spéciation allopatrique vicariante et de la spéciation péripatrique.

Figure 7 : Représentation schématique des étapes théoriques de la spéciation sympatrique et de la spéciation parapatrique.

Figure 8 : Classification des mécanismes d'isolement reproducteur.

Figure 9 : Représentation schématique d'incompatibilité cytoplasmique.

Figure 10 : Représentation schématique de l'incompatibilité de type Dobzhansky–Muller.

Figure 11 : Evolution des incompatibilités génétiques qui peuvent induire la non-viabilité et la stérilité des hybrides.

Figure 12 : Mécanismes génétiques qui lient la sélection divergente à l'isolement reproducteur.

Figure 13 : Représentation schématique de la vision en île ou en continent de la divergence génomique.

Figure 14 : Aires de répartition des six plus grandes espèces de ravageurs du genre *Spodoptera*.

Figure 15 : Cycle de vie de *Spodoptera frugiperda*.

Figure 16 : Diagramme de la composition des différents composés présents dans le cocktail de phéromones sexuelles du variant maïs et riz de *S. frugiperda*.

Figure 17 : Photographies de dommages engendrés par les larves de *S. frugiperda*.

Figure 18 : Organigramme présentant chronologiquement les différentes étapes de la thèse.

Figure 19 : Photographie du protocole expérimental utilisé pour réaliser des croisements intra et inter variants.

Figure 20 : Comparaison entre variant maïs et variant riz des séquences d'un fragment du gène COI.

Figure 21 : Représentation schématique du processus de crossing-over lors de la formation des gamètes.

Figure 22 : Carte du continent américain présentant l'échantillonnage de *Spodoptera frugiperda*.

Figure 23 : Photographies prises lors de l'échantillonnage sur le terrain de FAW adultes.

Figure 24 : Analyse de *barcode* pour une région du gène COI.

Figure 25 : Schéma de la méthode utilisée pour assembler les *contigs* et les *scaffolds*.

Figure 26 : Pipeline d'analyses bio-informatiques réalisées pour annoter les régions flanquantes des microsatellites étudiés et comparer les séquences du variant riz et du variant maïs de *S. frugiperda*.

Figure 27 : Représentation schématique de la spéciation par renforcement.

Liste des tableaux

Tableau 1 : Cas d'études détaillées d'isolement reproducteur.

Tableau 2 : Gènes d'incompatibilité hybride.

Tableau 3 : Statistiques de l'assemblage du génome de *S. frugiperda* version 3.

Sommaire

PARTIE 1 : Introduction générale	17
1. La spéciation	17
1.1. Le concept d'espèce : difficulté de définition	17
1.1.1. Evolution de la définition de l'espèce	17
1.1.2. Critères de définition d'une espèce et méthodes de délimitation associées	19
1.2. Les différents modèles de spéciation	27
1.2.1. Spéciation avec rupture du flux de gènes	28
1.2.2. Spéciation avec flux de gènes.....	30
1.3. Spéciation écologique : cas des insectes phytophages	34
1.3.1. Définition.....	35
1.3.2. Le concept de race d'hôte	35
1.4. Etude des mécanismes d'isolement reproducteur	36
1.4.1. Les barrières qui conduisent à l'isolement reproducteur	37
1.4.2. Signatures génétiques et génomiques de la spéciation	43
2. Le modèle biologique : <i>Spodoptera frugiperda</i>	52
2.1. Généralités	52
2.2. Biologie générale et cycle de vie	54
2.3. Différences existant entre les deux variants	56
2.3.1. Ecologiques.....	56
2.3.2. Comportementales.....	57
2.3.3. Génétiques	57
2.4. Un ravageur des cultures.....	58
3. Principaux objectifs de la thèse.....	60
PARTIE 2 : Matériel et Méthode Général	61
1. Les élevages.....	64
1.1. <i>Spodoptera frugiperda</i> , insecte de quarantaine NS2	64
1.2. Les conditions d'élevage	64
2. La caractérisation moléculaire des variants de <i>S. frugiperda</i>	66
2.1. Sur population d'élevage.....	66
2.2. Sur populations naturelles	69
3. Le génome de <i>Spodoptera frugiperda</i>	72
3.1. Assemblage du génome	72
3.2. Annotation du génome et alignement	73

3.3. Comparaison des <i>scaffolds</i> <i>S. frugiperda</i> variant riz et maïs	75
PARTIE 3 : Principaux résultats	77
1. Premier niveau d'analyse : études sur populations naturelles	79
1.1. Phylogénie et évolution du genre <i>Spodoptera</i>	81
1.2. Analyse de délimitation d'espèce.....	99
1.3. Différenciation génétique et structuration des populations des deux variants de <i>S. frugiperda</i>	133
2. Deuxième niveau d'analyse : analyse génétique et génomique sur population de laboratoire.	153
2.1. Caractérisation des marqueurs microsatellites.....	155
2.2. Isolement reproducteur entre les deux variants de <i>S. frugiperda</i>	169
PARTIE 4 : Conclusion générale et Perspectives	191
Annexes	199
Références.....	203

Partie 1 : Introduction Générale

Avant-propos

Parmi l'extraordinaire biodiversité présente sur Terre, les insectes représentent plus de 70% du règne animal et sont répartis en plus de 880 000 espèces. Aujourd'hui plus de 10-20% de la production agricole mondiale est détruite quotidiennement par des insectes herbivores. Des changements globaux peuvent perturber le contrôle naturel des bio agresseurs en créant un déséquilibre entre ces derniers et leurs ennemis naturels, augmentant alors le pourcentage de perte et rendant urgent le contrôle de ces populations d'insectes. Dans ce contexte, l'enjeu est important de bien comprendre les mécanismes d'évolution, d'adaptation et de spéciation de ces insectes. Cette thèse se concentre autour d'un modèle biologique *Spodoptera frugiperda* et de ses variants adaptatifs à la plante hôte. Les questions sur la biologie de ce lépidoptère ravageur des cultures sont nombreuses et ont déjà fait l'objet d'un grand nombre d'études. Pourtant certains points restent controversés. En effet plusieurs études sur la biologie de ces deux variants suggèrent que le niveau de différenciation entre les deux variants est élevé et que *S. frugiperda* pourrait être un bon modèle de spéciation en cours. Cette thèse s'inscrit dans le cadre d'une collaboration entre deux laboratoires différents, le laboratoire DGIMI (Diversité, Génomes & Interactions Microorganismes-Insectes) et le CBGP (Centre de Biologie pour la Gestion des Populations) et vise à essayer de comprendre quel est le statut de ces deux variants de *S. frugiperda*. La collaboration de ces deux laboratoires a permis d'appréhender la question de manière intégrative avec plusieurs niveaux d'analyses, allant d'une échelle d'analyse large, avec des études sur populations naturelles, vers une échelle d'analyse plus fine, à travers la comparaison des génomes d'individus de population d'élevages. Cette introduction présentera donc le contexte de la spéciation avec ses complexités et les nuances qui existent autour des différents concepts qu'elle implique ainsi que les mécanismes génétiques et génomiques sous-jacents. Puis les connaissances et les caractéristiques biologiques connues jusqu'alors sur les deux variants de *Spodoptera frugiperda* seront passées en revue.

PARTIE 1 : Introduction générale

1. La spéciation

Une des grandes questions de la biologie évolutive est d'essayer de comprendre l'origine et les mécanismes conduisant à l'incroyable diversité phénotypique et génotypique du monde vivant. Dans ce chapitre, nous commencerons par définir ce qu'est une espèce vivante et quels sont les critères pris en compte pour distinguer deux espèces entre elles en discutant la validité de chacun d'eux. Nous passerons ensuite en revue les différents modes de spéciation, c'est-à-dire les processus connus pour conduire à la formation de nouvelles espèces. Puis nous aborderons les connaissances concernant les bases moléculaires du phénomène de spéciation. Enfin, nous terminerons par une synthèse bibliographique sur *Spodoptera frugiperda*, notre modèle biologique pour lequel nous interrogerons le statut d'espèce dans la suite de cette thèse.

1.1. Le concept d'espèce : difficulté de définition

La notion d'espèce, unité du décompte de la biodiversité, est inhérente à l'étude de la spéciation, et a toujours été la source de nombreuses discussions autour de sa définition. Le grand nombre d'études sur ce domaine illustre bien la complexité du problème (Dobzhansky 1937; Mayr 1942; Harlan et De Wet 1963; Avise et Ball 1990; Mayden 1997; Wilson 1999; Wheeler et Meier 2000; Noor 2002; de Queiroz 2007). Au cours du XX^{ème} siècle, plus d'une vingtaine de définitions de l'espèce ont été proposées (Mayden 1997). Une partie des problèmes de définition de la notion d'espèce semble directement liée à la difficulté de définir les critères qui caractérisent et permettent de délimiter pratiquement ces entités.

1.1.1. Evolution de la définition de l'espèce

La difficulté de définition de l'espèce n'est pas nouvelle et réside peut-être dans l'origine même du mot *espèce*. En effet étymologiquement *species* (qui a donné espèce en français) est la traduction latine du terme grec *eidos* proposée par Cicéron, qui est le plus souvent traduit en français par « idée » ou « forme » selon les philosophes. Durant l'antiquité et le moyen-âge, le débat philosophique autour de la notion d'espèce était de

savoir si ce mot renvoyait à quelque chose de réel ou bien s'il était une simple convention nominale.

Cet héritage philosophique a lourdement pesé sur les discussions quand les biologistes ont commencé à donner au terme espèce un sens particulier, réservé aux êtres vivants. Ce n'est réellement qu'à la fin du XVI^{ème} siècle, avec Buffon et Ray que la notion d'espèce philosophique va se distinguer de la notion naturaliste, qui dans ce cas renvoie à une lignée d'individus connectés par la génération dans le temps. Les espèces sont alors définies selon des critères qui sont la ressemblance, la descendance et l'interfécondité. Ainsi Cuvier (1769-1862) définit la notion d'espèce comme « *la collection de tous les corps organisés nés les uns des autres ou de parents communs et de ceux qui leur ressemblent autant qu'ils se ressemblent entre eux* ».

La théorie de l'évolution proposée par Darwin représente une vraie révolution dans la définition du concept d'espèce (Darwin 1859). En effet il ne gardera de la définition « pré-évolutionniste » de l'espèce que le critère de la descendance, et amènera la notion de sélection naturelle pour expliquer la « transformation des espèces ». Avec la théorie de l'évolution, les organismes vivants passés ou présents sont capables de se reproduire et donnent naissance à de nouveaux organismes capables eux aussi de se reproduire. Ernst Mayr (1942) définit alors un concept de l'espèce biologique (CEB) correspondant à des groupes de populations naturelles interfécondes qui sont isolées sur le plan reproductif des autres groupes du même type. Les organismes font alors partie d'un réseau généalogique souvent représenté par une arborescence (Figure 1). Dans ce contexte, Samadi et Barberousse (2006) définissent l'espèce comme étant une branche du réseau généalogique, c'est à dire un fragment du réseau compris soit entre deux événements de spéciation, soit entre un événement de spéciation et un événement d'extinction (Figure 1). L'espèce est alors un groupe d'individus ayant une histoire évolutive autonome et capables de se reproduire.

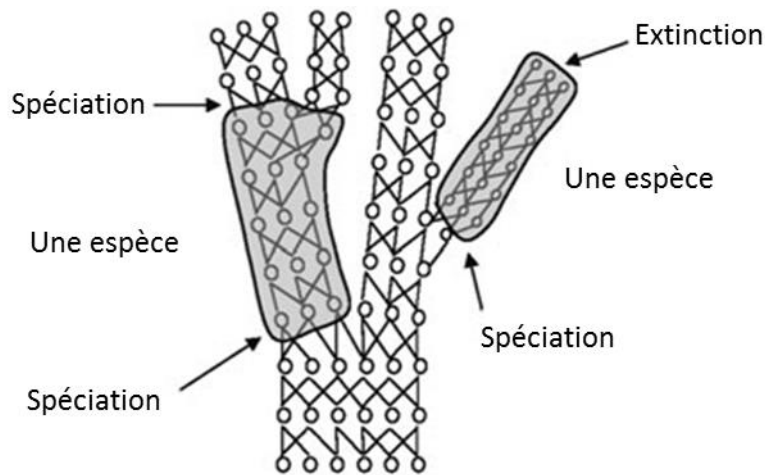


Figure 1 Représentation schématique de la notion d'espèce (Samadi et Barberousse, 2006). Les cercles représentent les individus connectés les uns aux autres par des traits. Les cercles grisés correspondent aux espèces.

1.1.2. Critères de définition d'une espèce et méthodes de délimitation associées

Dans le contexte de cette définition de l'espèce, basée sur la théorie de l'évolution, plusieurs critères ont été décrits afin de délimiter les espèces (Sites et Marshall 2004). Nous allons passer en revue et discuter la pertinence de ces critères dans ce paragraphe.

1.1.2.1. Critère morphologique

Le critère morphologique a été jusque dans les années 1940 l'une des méthodes d'identification et de définition des espèces les plus utilisées (Coyne 1994). Ce critère repose sur l'utilisation de caractères morphologiques communs ou non à plusieurs individus. En théorie les espèces, sous l'effet de la dérive et de la sélection, vont accumuler des caractères particuliers conduisant à des différences et à la formation de nouveaux caractères propres à chaque espèce. Les caractères morphologiques définis sont décidés *a priori*, généralement, sur la base de critères phénétiques, notamment basés sur la ressemblance et faisant appel à des notions de cladistique (Encadré 1). Cependant il apparaît un certain nombre de limites liées en partie au fait que ce critère phénétique ne prend pas en compte la notion primitive ou dérivée de l'état du caractère.

Encadré 1 : définitions de caractère et état de caractère

Un caractère est défini comme étant un attribut observable d'un organisme et l'état de caractère correspond à une modalité qui est prise par ce caractère. L'exemple classiquement utilisé est celui de l'œil : si « couleur de l'œil » est un caractère alors « bleu » est l'état du caractère « couleur de l'œil » (Thomas *et al.* 2010). L'état primitif d'un caractère ou **plésiomorphie** correspond à un état de caractère ancestral c'est-à-dire apparu anciennement alors que l'état dérivé d'un caractère ou **apomorphie** désigne un état de caractère dérivé de l'ancêtre commun le plus proche.

Une **synapomorphie** est un caractère dérivé partagé par deux ou plusieurs taxons également partagés par leur ancêtre commun le plus proche.

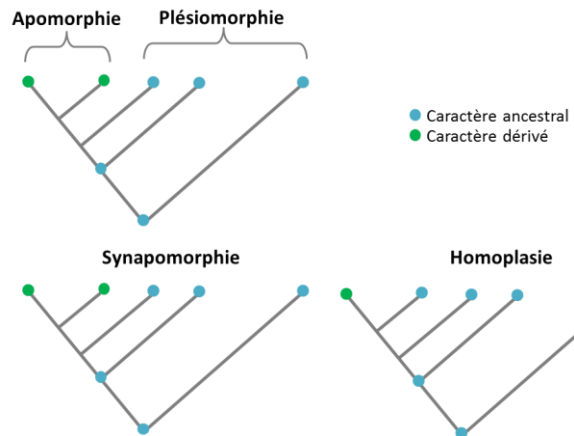


Figure 2 : différentes terminologies des états de caractères en phylogénie

Un caractère **homoplasique** quant à lui est un caractère plésiomorphe présent chez deux ou plusieurs taxons appartenant à des lignées différentes.

Une des limites de ce critère est qu'il ne permet pas d'identifier les cas de convergence évolutive. En effet un caractère morphologique partagé ne signifie pas obligatoirement une histoire évolutive commune. La convergence évolutive est un mécanisme évolutif qui peut permettre d'expliquer la similitude non héritée (ou homoplasie (Encadré 1)). Les individus appartenant à des espèces différentes partagent des caractères morphologiques semblables mais sont reproductivement isolés (Giraldo *et al.* 2008). Le critère morphologique ne permet pas non plus d'identifier les espèces cryptiques, c'est-à-dire des espèces morphologiquement identiques mais génétiquement différenciables (Bickford *et al.* 2007). Enfin des erreurs d'identification d'espèces, basées uniquement sur le critère morphologique, peuvent être dues à la présence de dimorphisme sexuel (Figure 3).

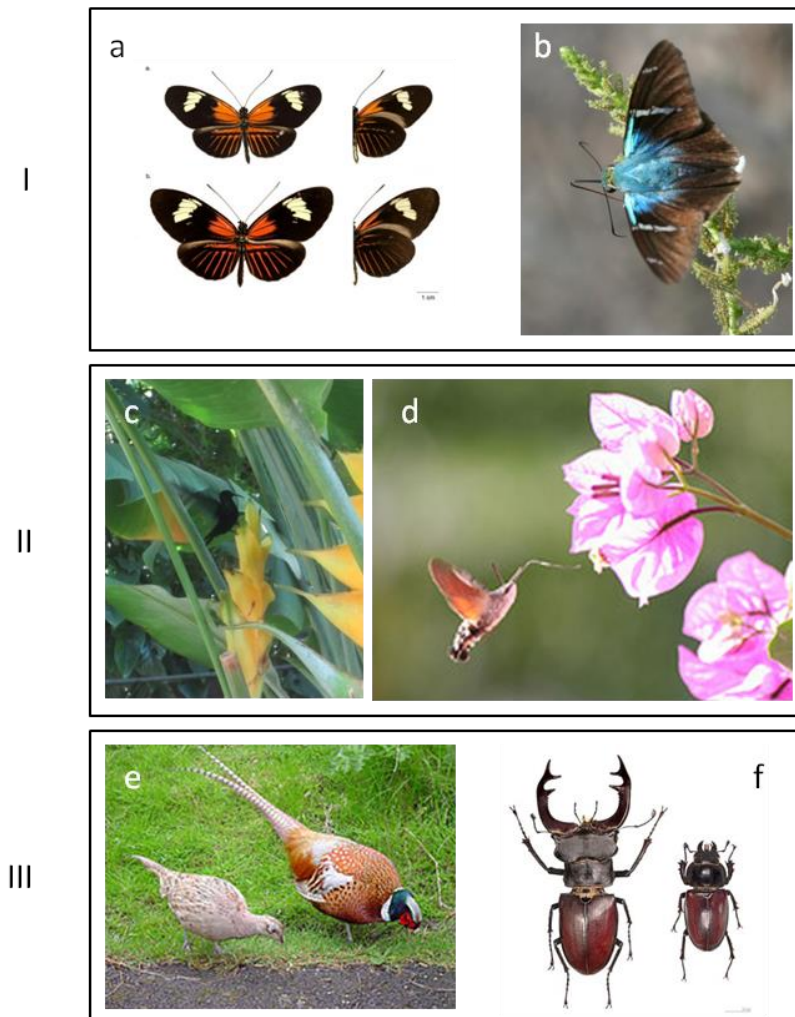


Figure 3 : Exemples de cas : I- d'espèces cryptiques avec (a) le papillon *Heliconius m. malleti* and *Heliconius timareta* (Figure Giraldo *et al.* 2008) et (b) le papillon *Astrartes fulgerator*. II-de convergence évolutive avec le colibri *Eulampis jugularis* (c) et le papillon *Macroglossum stellatarum* (d). Et III- de dimorphisme sexuel chez les faisans *Phasianus colchicus* (e) et chez le coléoptère *Lucanus cervus*(f).

1.1.2.2 Critère phylogénétique

C'est en 1966 que le concept d'espèce phylogénétique est proposé pour la première fois par Hennig. Le critère phylogénétique correspond à la fixation au cours de l'évolution de caractères dérivés dans des lignées autonomes. Ce concept repose sur le principe que seuls les caractères dérivés (apomorphiques (Encadré 1)) permettent de retracer l'histoire évolutive des lignées. Ce critère considère ainsi que seuls les groupes monophylétiques, c'est-à-dire les groupes constitués de l'ancêtre commun le plus récent et de ses descendants, permettent de définir les espèces (Rosen 1979; Mishler 1985).

Le critère phylogénétique est à la base de nombreuses méthodes utilisées aujourd'hui pour identifier et nommer les taxons (la taxonomie) mais aussi pour délimiter les espèces (Sites et Marshall 2003; O'Meara *et al.* 2006; Pons *et al.* 2006; Ence et Carstens 2011; Hamback *et al.* 2013). Ces méthodes utilisent des arbres phylogénétiques et sont donc appelées en anglais *tree-based methods*. Parmi celles-ci on distingue les méthodes de distance, les méthodes basées sur la parcimonie et les méthodes probabilistes.

Les méthodes de distance

Les méthodes de distance (ou phénétiques) calculent un indice de similitude entre deux taxons, correspondant à des séquences alignées. La distance correspond au nombre de nucléotides différents entre les deux espèces, divisé par le nombre de sites examinés, ce qui permet de définir un pourcentage de différences entre les séquences de deux espèces. Les distances peuvent être corrigées afin d'incorporer une information évolutive, par exemple afin de tenir compte du fait que certains types de mutations ont lieu plus souvent que d'autres. Il existe plusieurs techniques de construction d'un arbre phylogénétique à partir d'une telle matrice comme par exemple la méthode de l'UPGMA (*Unweight Pair Group Method with Arithmetic mean*), qui identifie les deux séquences les plus proches et ce groupe est ensuite traité comme un tout, puis la séquence la plus proche est recherchée et ainsi de suite jusqu'à ce qu'il n'y ait plus que deux groupes. Une deuxième méthode correspond au NJ (*Neighbor-Joining*) (Saitou et Nei 1987) qui utilise la méthode de l'UPGMA en la corrigeant afin d'autoriser un taux de mutation différent sur les branches.

La parcimonie

Les méthodes basées sur la parcimonie visent à établir les relations de parenté entre individus en identifiant la co-occurrence de caractères dérivés parmi les états de caractères homologues, c'est-à-dire les synapomorphies (Encadré 1). Le maximum de parcimonie (ou méthode cladistique) comme son nom l'indique, est une méthode qui a pour but de reconstruire l'arbre phylogénétique le plus parcimonieux en minimisant le nombre d'homoplasies, c'est-à-dire le nombre de ressemblances sans ascendance commune. Toutefois cette méthode présente des inconvénients car elle est sensible, par exemple au phénomène d'attraction des longues branches (Simmons 2004), qui conduit au regroupement de taxons non apparentés. Cette approche a été utilisée dans plusieurs études de délimitation d'espèces comme par exemple chez les salamandres (Templeton *et*

al. 1995) et chez les insectes (Wilder et Hollocher 2003; Cardoso et Vogler 2005; Chen *et al.* 2010).

Les méthodes probabilistes

Les méthodes probabilistes sont basées sur le concept de vraisemblance, qui est la probabilité conditionnelle d'observer les données pour un ou des modèles d'évolution particuliers (Delsuc et Douzery 2004). Deux types de méthodes probabilistes ont été développées à partir de ce concept de vraisemblance : le maximum de vraisemblance (Felsenstein 1973, 1981), qui va chercher l'arbre phylogénétique qui maximise cette vraisemblance et l'inférence Bayésienne (Li 1996; Mau 1996; Rannala et Yang 1996), qui recherche un grand nombre de topologies ayant une vraisemblance élevée et considère alors que la meilleure topologie correspondra au consensus de toutes les topologies analysées. Ces deux approches utilisent la recherche d'un « résultat » approché (recherches heuristiques) étant donné l'importance du nombre de topologies à examiner. Le résultat obtenu par ces méthodes n'est donc pas une solution exacte. L'utilisation de ces approches s'est accrue de manière spectaculaire avec l'accessibilité et la facilité d'acquisition des données moléculaires. Dans les méthodes probabilistes, l'évolution est modélisée comme un processus aléatoire (*stochastique*) qui est décrit comme la probabilité de substitution au cours du temps et le long des séquences génétiques (Thomas *et al.* 2010). Ces approches impliquent, dans un premier temps, de sélectionner un modèle d'évolution qui a été généralement choisi à partir d'un arbre phylogénétique reconstruit grâce aux méthodes de distance (NJ la plupart du temps). A partir de cette topologie, les valeurs de paramètres de différents modèles d'évolution (vitesse d'évolution ou taux de substitutions, variations d'évolution au cours du temps ou entre les marqueurs utilisés), de plus en plus complexes, vont être associées et la vraisemblance de chacun sera estimée. Le modèle d'évolution sélectionné correspond alors au modèle qui optimise le mieux le nombre de paramètre par rapport à la vraisemblance.

Plusieurs méthodes de délimitation d'espèces ont été développées à partir de ces approches phylogénétiques probabilistes (O'Meara *et al.* 2006; Ence et Carstens 2011; Fujita *et al.* 2012). Ces approches se placent dans le contexte de la coalescence (Kingman 1982) (Encadré 2), qui a pour but d'identifier les lignées indépendantes, sans flux de gènes et sur lesquelles agissent la dérive et la sélection. Parmi ces méthodes, le modèle mixte dit GMYC

(General Mixed Yule-Coalescent) (Pons *et al.* 2006) est de plus en plus utilisé dans les études récentes de délimitation d'espèce (McBride *et al.* 2009; Fontaneto *et al.* 2011; Kergoat *et al.* 2011; Vuataz *et al.* 2011; Astrin *et al.* 2012; Bryson *et al.* 2013; Fernández-Flores *et al.* 2013). Le principe de cette méthode est de définir, grâce à la notion de maximum de vraisemblance, un seuil virtuel délimitant les relations intraspécifiques et des relations interspécifiques au sein d'un arbre phylogénétique rendu arbitrairement ultramétrique (longueur des branches proportionnelles au temps). Plus précisément, cette méthode associe simultanément deux modèles différents : le premier est un modèle de coalescence et le deuxième est un modèle de spéciation de Yule (Yule 1925) (Encadré 2).

Encadré 2 : Modèle de Yule et modèle de Coalescence

Modèle de spéciation :

Dès 1924, Yule décrit un processus en lien avec la théorie de l'évolution, où il considère que le taux d'apparition des nouvelles espèces est linéaire par rapport au nombre d'espèce déjà présentes. Un processus de branchement à taux constant (aussi appelé processus de naissance pure à taux constant) où les événements de branchements correspondent à des événements de spéciation. Chaque lignée a une durée de vie qui suit une loi exponentielle de paramètre λ et se sépare à la fin de sa vie en deux lignées : une correspondant à l'espèce ancestrale et l'autre à la nouvelle espèce.

Modèle de Coalescence :

La Coalescence est un modèle rétrospectif, c'est-à-dire qu'à partir des données génétiques d'une population, il retrace l'évolution de tous les allèles d'un gène donné pour cette population, en remontant le temps jusqu'à trouver une seule copie ancestrale de ce gène, qui représente alors l'ancêtre commun le plus récent. La coalescence peut être représentée sous la forme d'une arborescence appelée *coalescent* (Figure 4).

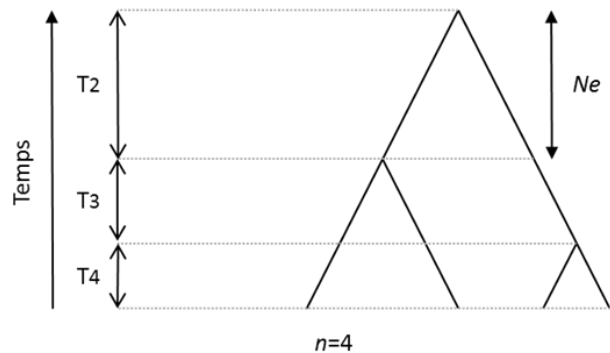


Figure 4: Figure du *Coalescent* standard pour un nombre de gènes : $n=4$ dans une population de taille constante

La probabilité que deux lignées coalescent (c'est-à-dire la réunion de deux lignées en une lignée unique) à la génération précédente est dépendante de la taille de l'échantillon et non de la taille de la population. Ainsi dans sa forme la plus simple, la théorie de la coalescence prédit que dans une population haploïde, de taille constante et égale à N_e , et pour un nombre de gènes échantillonnés : $n=2$, alors la probabilité que les deux gènes de l'échantillon aient un ancêtre commun à la génération précédente est :

$$prob_{coal}(2, N_e) = \frac{1}{N_e}$$

Le premier modèle (Coalescent) teste tout d'abord l'hypothèse selon laquelle les différents embranchements de l'arbre font partie d'une même population. Alors que le deuxième (de Yule) va tester la seconde hypothèse qui suggère que les différentes branches de l'arbre correspondent à des lignées évolutives indépendantes. Le modèle GMYC détermine alors un seuil de maximum de vraisemblance au-dessous duquel les nœuds de la phylogénie correspondent à des événements interspécifiques et au-dessus duquel les nœuds sont considérés comme des événements intra spécifiques. A partir de ce modèle mixte GMYC, Monaghan et ses collaborateurs (2009) ont développé un modèle capable de détecter plusieurs seuils de différenciation d'évènement intra et interspécifique, au sein d'un même arbre phylogénétique, ce qui permet d'analyser des lignées hétérogènes (différences marquées en termes d'ancienneté et/ou de taux d'évolution).

1.1.2.3. Critère biologique

Le concept biologique de l'espèce est décrit pour la première fois par Mayr (Mayr 1942). Il y définit les espèces comme étant « *des groupes de populations naturelles, effectivement ou potentiellement interfécondes, qui sont génétiquement isolées d'autres groupes similaires* ». Le critère biologique ou encore le critère d'interfécondité s'applique à des populations. L'appartenance à une espèce n'est donc plus diagnostiquée sur la base de caractères essentiels individuels. Le critère biologique est un critère puissant de délimitation qui repose entièrement sur la notion d'isolement reproductif, c'est-à-dire l'ensemble des barrières biotiques et abiotiques (*cf* paragraphe 1.4.1.) qui vont empêcher les individus de deux lignées distinctes d'engendrer une descendance viable et fertile. Cependant si dans certains cas l'isolement reproductif apparaît de manière évidente lorsque les différences anatomiques sont grandes, dans d'autre cas les barrières sont beaucoup moins nettes. Par exemple dans le cas du cheval et de l'âne, il n'y a pas d'isolement reproductif à première vue car ils sont capables de se reproduire. Toutefois leurs hybrides (le bardot et le mulet), eux, sont stériles. Cet exemple montre qu'il est donc important de préciser, dans la définition du critère biologique de l'espèce, la notion de viabilité et de fécondité des hybrides.

Bien que ce critère biologique de l'espèce soit un critère puissant et qui fait le plus consensus de nos jours, il présente un certain nombre de limites. En effet, ce critère ne

prend pas en compte les cas d'hybridation ponctuelle, ce qui est, entre autre, le cas chez les plantes. De plus, même si ce critère est théoriquement très discriminant, il apparaît plus difficile d'un point de vue pratique, quand il s'agit de mesurer l'isolement reproductif entre organismes au sein de populations naturelles. Dans ce cas l'étude de l'isolement reproductif peut être ramenée à la présence ou non de flux de gènes (Sites et Marshall 2004; Petit et Excoffier 2009). Enfin, le critère biologique ne prend pas en compte le cas des organismes asexués (insectes parthénogénétiques ou bactéries par exemple) ou des organismes ne participant pas directement à la reproduction comme par exemple les formes stériles spécialisées (cas des abeilles ouvrières ou des soldats chez les termites) (Barraclough *et al.* 2003; Birky *et al.* 2005). Pour ces derniers le critère de l'isolement reproductif conduirait à considérer chaque individu comme une espèce, ce qui n'est évidemment pas le cas. Ainsi de nouveaux critères particuliers sont à chercher pour mieux cerner l'espèce, comme par exemple le critère écologique (Valen 1976) ou généalogique (Taylor *et al.* 2000).

1.1.2.4. Approches intégratives

Bien que le critère biologique soit le critère le plus largement admis au moins pour la définition des organismes eucaryotes à reproduction sexuée, il n'est en pratique pas le plus utilisé pour délimiter les espèces. En effet, la plupart des espèces ont-été et sont encore décrites sur la base du critère morphologique car il est difficile dans la plupart des cas de vérifier les critères biologiques dans la nature. C'est pourquoi, de plus en plus d'études multiplient les types de données expérimentales, pour valider ou invalider une hypothèse de départ. Ces études permettent de contourner en partie le problème de la pertinence du choix du critère utilisé pour définir l'espèce, en combinant plusieurs critères de définitions précédemment décrits. Cette nouvelle approche méthodologique est souvent désignée par le terme de taxonomie intégrative (Dayrat 2005; Schlick-Steiner *et al.* 2010) et est de plus en plus utilisée (Figure 5) dans les études d'identification et de délimitation d'espèces (Alstrom *et al.* 2008; Lumley et Sperling 2010; Gebiola *et al.* 2012; Rougerie *et al.* 2012; Lumley et Cusson 2013). Les approches intégratives de délimitation des espèces offrent un moyen de répondre à la difficulté du critère biologique de l'espèce même si dans beaucoup de cas ces approches nécessitent de combiner des domaines différents de la biologie.

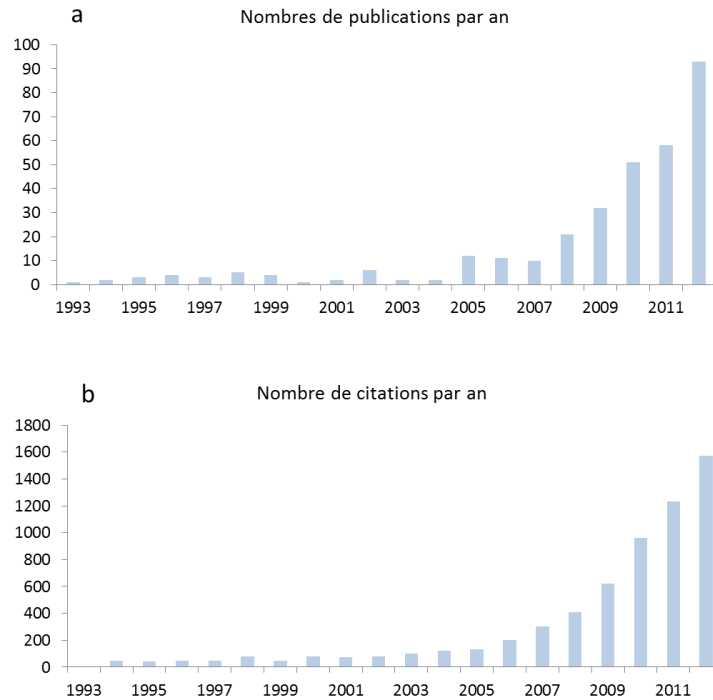


Figure 5 : Histogrammes obtenus grâce à l'outil *citation report* de la base de données *Web of science* pour une recherche comprenant les mots *integrative taxonomy*. L'histogramme a) représente le nombre de publications par an correspondant à cette recherche et le b) représente le nombre de citations par an de ces publications.

1.2. Les différents modèles de spéciation

Dans ce paragraphe nous allons nous concentrer sur la spéciation chez les organismes à reproduction sexuée, qui s'étudie en grande partie au travers de l'isolement reproducteur existant entre les populations. Un des débats autour de la question de la spéciation porte sur la pertinence de classification biogéographique des différents mécanismes qui la composent. Initialement les principaux modes de spéciation (Mayr 1942) ont été définis en fonction de l'existence de barrières géographiques entre les populations pendant leur divergence supposée (allopatrie), qui conduirait à la spéciation. Dans ce contexte le flux de gènes entre populations non isolées géographiquement empêche l'évolution de l'isolement reproducteur et la spéciation allopatrique devient alors le modèle principal de spéciation. Les autres modes de spéciation (sympatrie, parapatrie) sont généralement considérés comme rares voire inexistants.

Cependant, une autre approche de la spéciation est envisagée depuis ces dix dernières années (Kirkpatrick et Ravigné 2002; Mallet *et al.* 2007; Butlin 2008), considérant plus de facteurs (comme les différences de capacités de dispersion et donc la possibilité qu'il

existe un échange génétique entre les populations). De plus, une nouvelle dimension est apportée à l'étude de la spéciation, avec la spéciation écologique (Funk 1998; Schluter 2001; Nosil *et al.* 2002), « où la sélection divergente survient due à des différences entre environnements, et résulte en l'évolution de l'isolement reproducteur. » (Nosil 2008), dans ce cas la sélection divergente est le processus clé de la spéciation. Avec ces nouvelles approches des modèles classiques de spéciation, l'accent est mis sur l'étude de l'origine et des mécanismes qui conduisent à l'évolution de l'isolement reproducteur. Enfin, l'avancée technologique de ces dernières années avec les nouvelles méthodes de séquençage, plus rapides et moins onéreuses ont fait émerger un nouvel axe de recherche dans la spéciation : l'étude des bases génétiques et génomiques de la spéciation.

1.2.1. Spéciation avec rupture du flux de gènes

L'allopatrie est classiquement divisée en deux catégories : la vicariance et la péripatrie (Figure 6). Dans les deux cas, la rupture de flux de gènes fait suite à l'isolement d'une population en deux par des barrières géographiques ou extrinsèques. La seule différence entre la vicariance et la péripatrie réside dans la taille de la population concernée par l'isolement.

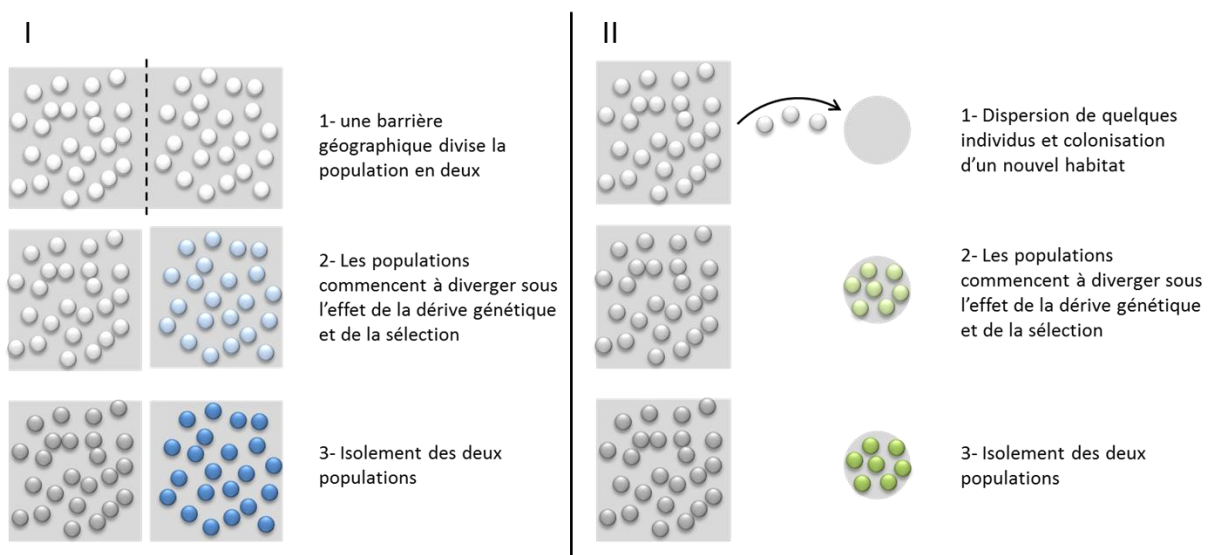


Figure 6 : Représentation schématique des étapes théoriques de la spéciation allopatrique vicariante (I) et de la spéciation péripatrique (II).

1.2.1.1. Spéciation allopatrique par vicariance

La définition de la vicariance est en théorie assez intuitive. Elle correspond à la séparation géographique de populations précédemment interfécondes par des barrières extrinsèques (glaciation, formation de chaîne de montagne, séparation des continents...). Si seule, la dérive génétique agit pour conduire à la fixation d'allèles différents entre les deux populations alors la divergence sera lente (Nei *et al.* 1983). Si les habitats sont différents de part et d'autres des barrières, la divergence entre les populations peut s'accélérer du fait de pressions de sélection différentes exercées par les différents habitats (Schluter 2001). La divergence sera encore plus rapide si la sélection, qui s'exerce sur les populations, est sexuelle.

On retrouve plusieurs cas de spéciation allopatrique par vicariance dans la nature que ce soit chez des organismes marins (Colborn *et al.* 2001), des angiospermes (Xiang *et al.*, 1998, 2000) ou des insectes (Sanmartin *et al.* 2001). Un des exemples emblématiques de spéciation allopatrique concerne la faune marine de part et d'autre de l'Isthme du Panama. Plusieurs couples d'espèces sœurs ont été identifiés (Lessios *et al.* 1998; Lessios *et al.* 2001). Cela a été particulièrement bien étudié chez les crevettes marines du genre *Alpheus* (Knowlton *et al.* 1993; Knowlton et Weigt 1998), où sept couples d'espèces jumelles ont été identifiés au sein du genre avec, pour chacun, la présence d'une espèce sœur provenant de la mer des Caraïbes et l'autre de l'océan Pacifique.

1.2.1.2. Spéciation péripatrique

Le principe de la spéciation péripatrique est en théorie le même que celui de la vicariance, seule la taille de la population concernée change, il s'agit en effet d'une très petite population. La péripatrie peut alors mener d'une part à l'invasion d'un nouvel habitat et donc induire une forte sélection. D'autre part pour des petites populations, la dérive génétique devient forte et peut alors jouer un rôle dans la spéciation.

La spéciation péripatrique peut se produire selon deux mécanismes différents. Le premier correspond à l'isolement d'une petite partie de la population par une barrière géographique. Dans ce cas les mécanismes de la spéciation sont les mêmes que ceux de la

vicariance. Le deuxième correspond à la colonisation, par un petit groupe d'individus de la population ancestrale, d'un nouvel habitat éloigné. Dans ce cas, la colonisation est souvent suivie de l'adaptation au nouvel habitat, qui peut être très différent de celui d'origine et donc induire une forte pression de sélection. La spéciation est alors une conséquence de cette sélection. Toutefois, même si les pressions de sélections sont identiques dans les deux habitats, le petit effectif de la population colonisatrice ne va refléter qu'une partie du paysage génétique de la population ancestrale et donc sous l'effet de la dérive génétique, des allèles différents peuvent être fixés dans chacune des deux populations et induire des incompatibilités entre les individus des deux populations.

Un des exemples le plus étudiés de spéciation péripatrique concerne les drosophiles de l'archipel d'Hawaï (Carson et Kaneshiro 1976), où la plupart des espèces retrouvées sont endémiques d'une des îles de l'archipel et descendent toutes du même ancêtre commun (DeSalle 1995). De même, une étude sur le mode de spéciation d'espèces de criquets de l'archipel d'Hawaï (Shaw 2002), établit, grâce à des analyses de phylogénie moléculaire, un mode de spéciation péripatrique pour au moins la moitié des espèces de criquet étudiées, ce qui correspond également à la proportion de cas de spéciation péripatrique pour les drosophiles (Carson et Kaneshiro 1976).

1.2.2. Spéciation avec flux de gènes

La spéciation avec flux de gènes désigne les modèles où il y a spéciation en l'absence de barrière géographique. Bien que cette conception de la spéciation ait longtemps été source de débat, il est maintenant accepté que l'évolution de l'isolement reproducteur sans séparation spatiale ou consécutif à une zone de contact secondaire puisse être un mode de spéciation (Bolnick et Fitzpatrick 2007; Butlin *et al.* 2008; Fitzpatrick *et al.* 2009; Smadja et Butlin 2011). La spéciation avec du flux de gènes regroupe classiquement deux modèles de spéciation, tout d'abord définis dans un contexte biogéographique : la spéciation parapatrique et la spéciation sympatrique (Figure 7). La spéciation parapatrique étant considérée comme à « mi-chemin » entre la spéciation allopatrique et sympatrique.

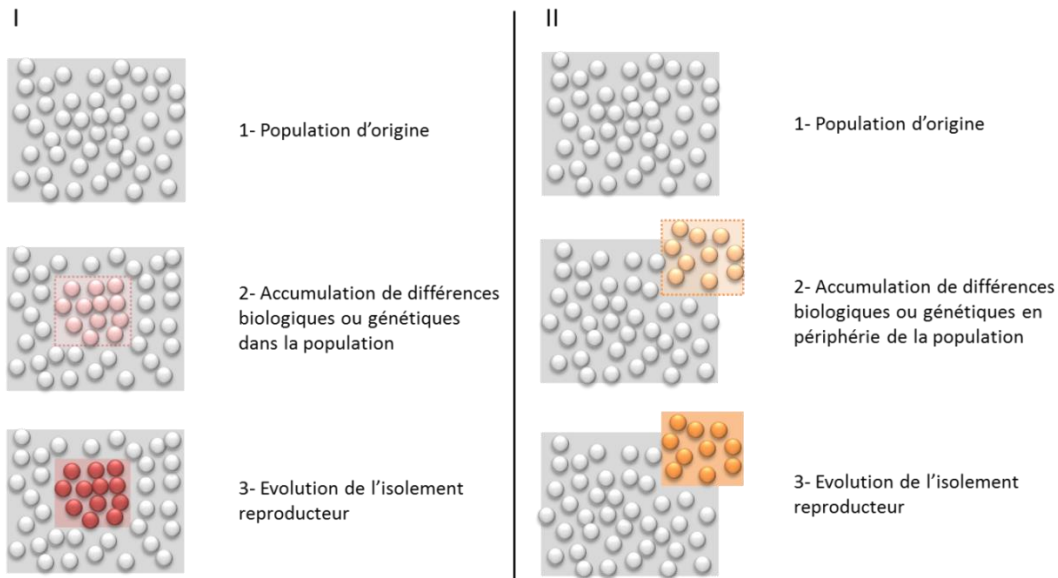


Figure 7 : Représentation schématique des étapes théoriques de la spéciation sympatrique (I) et de la spéciation parapatricque (II).

1.2.2.1. Spéciation parapatricque

La spéciation parapatricque correspond à une divergence entre populations proches qui échangent des migrants mais avec des flux de gènes limités. Dans ce cas l'isolement reproducteur entre les deux populations évolue sous l'effet de la sélection ou de la dérive génétique qui alors fait contrepoids au flux de gènes, et engendre de la spéciation. Deux modèles de spéciation parapatricque sont alors envisagés. La parapatricque selon un gradient ou un cline géographique considère une espèce distribuée en continu à travers un environnement variable, où les sous-populations s'adaptent à leur environnement malgré l'existence de flux de gènes entre populations adjacentes, et deviendraient des populations écologiquement différenciées. Le deuxième modèle de spéciation parapatricque envisage l'existence de populations discrètes avec un taux de migration faible et donc de flux de gènes réduit.

La spéciation parapatricque a fait l'objet d'études expérimentales et de modélisation (Gavrilets *et al.* 2000; Doebeli et Dieckmann 2003), mais l'existence de cas de spéciation parapatricque dans la nature est quasiment impossible à identifier. En effet, la présence de deux espèces sœurs dans des zones géographiques adjacentes n'est pas exclusif à la

parapatric, d'autres explications sont possibles comme par exemple un contact secondaire d'espèces ayant préalablement divergé en allopatric.

1.2.2.2. Spéciation sympatrique

La spéciation sympatrique est sûrement le modèle de spéciation qui a soulevé le plus de questionnements et de discussions. Dans le contexte biogéographique la sympatrie est définie comme l'occupation d'une même aire géographique par deux populations différentes (Mayr 1942). Une approche plus spatiale de la définition de la sympatrie a été apportée récemment par Mallet *et al.* (2009), où il incorpore la notion de génétique des populations et la capacité des individus à se disperser. Mallet propose ainsi une définition spatiale de la sympatrie : « où les individus sont physiquement capables de se rencontrer à une fréquence relativement haute ». Il définit également que « des populations peuvent être sympatriques même si elles sont écologiquement isolées, tant qu'une proportion élevée de chaque population est capable de se rencontrer ». Parallèlement à ce débat, une autre approche de la spéciation est appréhendée et préfère se concentrer sur l'évaluation des processus biologiques qui affectent la divergence plutôt que d'essayer de définir la sympatrie (Fitzpatrick *et al.* 2008, 2009; The Marie Curie speciation Network 2012).

En théorie, le modèle de spéciation sympatrique correspond à la divergence de deux populations sous l'effet de la sélection, sans aucune barrière extrinsèque. En sympatrie l'accouplement entre individus est supposé être aléatoire et ainsi le taux d'échange de gènes maximal ($m=0.5$) (Gavrilets 2003). Il semble donc difficile qu'il y ait dans ce contexte présence et évolution d'un isolement reproducteur entre deux populations. Pour lutter contre l'effet d'homogénéisation du flux de gènes, la sélection doit être divergente (Maynard Smith 1966; Bush 1969; Ravigné *et al.* 2004). La sélection divergente correspond à la sélection de deux optimums distincts. Classiquement, la sélection divergente a lieu en allopatric où les individus de populations différentes sont soumis à des pressions de sélections différentes liées à la spécificité de leur habitat. Chaque population isolée s'adapte alors à son nouveau milieu. En sympatrie, la sélection divergente a lieu quand le milieu est spatialement hétérogène avec des niches écologiques différentes mais non isolées, ce qui laisse aux individus la possibilité d'accéder à n'importe quelle niche.

La spéciation sympatrique étant difficile à démontrer, le nombre d'exemples non controversés dans la littérature est rare. Ainsi plusieurs cas ont été décrits avant d'être revus par la suite. Le cas de la mouche du pommier (*Rhagoletis pomonella*) en est l'exemple le plus emblématique. Il existe pour ce parasite de fruit deux races d'hôtes retrouvées sur le continent américain en sympatrie : une parasite l'aubépine (hôte initial) et l'autre parasite le pommier (hôte introduit) (Bush 1969; Feder *et al.* 1998). Les deux races d'hôtes ont les mêmes cycles de vie et sont génétiquement différenciés, ce qui d'après la définition de Berlocher et Feder (2002) fait de *R. pomonella* un excellent exemple de spéciation sympatrique en cours. Cependant des données moléculaires tendent à montrer qu'en fait les deux races d'hôtes auraient divergé en allopatrie et que la sympatrie serait une conséquence d'un contact secondaire (Feder *et al.* 2003; Xie *et al.* 2007).

Le cas des poissons cichlides du lac Apoyo au Nicaragua est un exemple presque non controversé de spéciation sympatrique dans la nature. Dans ce lac, les deux espèces de poissons rencontrées (*Amphilophus citrinellus* et *Amphilophus zalius*) proviennent d'une même population ancestrale, la deuxième espèce n'a été retrouvée que dans le lac, les deux espèces sont reproductivement isolées et elles possèdent des niches écologiques différentes (Barluenga 2006). Toutes ces caractéristiques font de ces poissons cichlides un parfait exemple de spéciation sympatrique.

1.2.2.3. Spéciation par hybridation

Comme cela a été décrit dans le paragraphe précédent, dans beaucoup d'études la formation d'hybride est considérée comme un frein au processus de la spéciation même si l'existence d'une spéciation avec flux de gènes et donc de l'hybridation est depuis peu prise en compte (Fitzpatrick *et al.* 2009; Nei et Nozawa 2011; Smadja et Butlin 2011). Cependant dans certains cas l'hybridation peut être considérée comme ayant un rôle dans la divergence des espèces (Seehausen 2004; Abbott *et al.* 2013). Chez les plantes, par exemple, la spéciation par hybridation est depuis longtemps considérée comme un moyen potentiel de diversification (Stebbins 1959; Grant 1981), grâce à la polyploïdie (un organisme est polyploïde lorsqu'il possède plus de deux jeux de chromosomes). Deux types de polyploïdies sont décrits : l'autopolyploïdie (les jeux de chromosomes proviennent de la même espèce),

et l'allopolyploïdie (les jeux de chromosomes proviennent d'espèces différentes) qui est une conséquence directe de l'hybridation interspécifique. Si chez les angiospermes 15% des évènements de spéciation impliquent la polyploïdie (Wood *et al.* 2009), il semble que cela ne concerne que quelques animaux (Mable *et al.* 2011). Chez les animaux un résultat potentiel de l'hybridation est la spéciation hybride homoploïde, qui contrairement à précédemment n'implique pas de changement de ploïdie chez les hybrides (Mallet *et al.* 2007; Abbott *et al.* 2010; Abbott *et al.* 2013). Dans ce cas, la combinaison nouvelle des allèles parentaux chez l'hybride conduit à l'établissement et au maintien d'une nouvelle population par la mise en place d'un isolement reproducteur avec les deux parents (Seehausen 2004; Abbott *et al.* 2013). *Heliconius heurippa* (Salazar *et al.* 2010) est un exemple de spéciation par hybridation et implique l'introgession, dans un contexte génétique différent, d'un petit nombre de loci et ayant un rôle direct dans la rupture du flux de gènes. Dans ce cas, l'apparition d'un nouveau patron alaire peut se faire par recombinaison génétique lors de croisements interspécifiques dans des régions sympatriques en générant une population hybride stable ne se croisant plus avec les espèces parentales : le patron alaire permettant à la fois une protection contre les prédateurs et la reconnaissance du partenaire sexuel sous forte sélection

1.3. Spéciation écologique : cas des insectes phytophages

L'idée selon laquelle des espèces peuvent naître de l'adaptation des organismes à des environnements écologiques différents n'est pas nouvelle (Darwin 1859) et l'étude des insectes phytophages a largement contribué à l'évolution de cette idée. En effet, Walsh (Walsh 1867) défend l'idée que la spéciation chez les insectes phytophages se fait par un changement vers un nouvel hôte. Pendant longtemps les insectes phytophages ont été utilisés comme modèles biologiques pour l'étude de la spéciation sympatrique (Berlocher et Feder 2002; Drès et Mallet 2002). Et c'est à partir de ces études que l'hypothèse d'une spéciation liée aux interactions biotiques des organismes avec leur environnement est née, identifiée sous le terme de spéciation écologique (Funk 1998; Schluter 2001; Rundle et Nosil 2005).

1.3.1. Définition

La spéciation écologique correspond à l'évolution de l'isolement reproducteur entre deux populations, lorsque cela est uniquement dû à la sélection divergente entre environnements biotiques différents, liée à des interactions écologiques. La sélection est donc écologique quand elle résulte des interactions entre les individus et leur environnement biotique. De plus, la sélection écologique peut avoir lieu dans n'importe quel contexte géographique, avec ou sans barrière au flux de gènes et peut impliquer n'importe quelle barrières reproductrices (Rundle et Nosil 2005; Nosil 2012b). Les insectes phytophages Figurent en bonne place dans la littérature dédiée à l'étude de la spéciation écologique (Via 2001; Rundle et Nosil 2005; Peccoud *et al.* 2009; Matsubayashi *et al.* 2010). En effet, un nombre important d'études réalisées sur l'évolution des modèles plante-insectes ont été consacrées à l'hypothèse que la spéciation était écologique chez les insectes phytophages et qu'elle se faisait en sympatrie (Drès et Mallet 2002; Via 2002), définissant alors un mode de spéciation sympatrique écologique.

Un autre des points clés de la spéciation sympatrique écologique est la notion de continuum. En effet, la spéciation est un processus continu et graduel allant du polymorphisme entre individus d'une même population à la formation de nouvelles espèces totalement différenciées et isolées d'un point de vue reproducteur. L'importance de l'isolement reproducteur observé entre deux populations peut être utilisée comme un moyen de mesurer jusqu'où le processus de spéciation est arrivé. Ainsi, on retrouve dans la littérature une série de termes pour caractériser les organismes qui sont en cours de spéciation en fonction de leur « place » le long du continuum de spéciation comme par exemple chez les insectes : les races d'hôtes, les écotypes, les biotypes et les espèces cryptiques (Funk *et al.* 2002).

1.3.2. Le concept de race d'hôte

Les races d'hôtes définissent un état de différenciation le long du continuum de spéciation et sont caractérisées par la présence d'un isolement reproducteur important mais incomplet. Plusieurs critères ont été établis afin de définir et identifier les races d'hôtes (Berlocher et Feder 2002; Drès et Mallet 2002) : 1) elles utilisent des plantes hôtes différentes, 2) elles coexistent le plus souvent en sympatrie, 3) elles sont génétiquement

différentiées, 4) il existe une corrélation entre le choix de l'hôte et le choix d'accouplement et 5) il y a présence de flux de gènes entre elles. Ces caractéristiques constituent un argument en faveur de la possibilité d'un processus de spéciation en présence de flux de gènes ou autrement dit de spéciation sympatrique. Cependant, le statut de race d'hôte reste flou ; comme le souligne Funk (Funk 2012), rien n'est connu du devenir de ces organismes. En effet, est-ce que les races d'hôtes vont continuer à diverger le long du continuum de spéciation, ou bien vont-elles rester indéfiniment à cet équilibre ou encore vont-elles redevenir une seule espèce ? De plus, comme nous l'avons vu précédemment, avec la mouche du pommier, les races d'hôtes peuvent être le reflet d'un contact secondaire et du maintien de la différenciation entre populations ayant initialement divergé en allopatrie.

Le puceron du pois *Acyrtosiphon pisum* a longtemps correspondu à un bon exemple de race d'hôtes, l'une préférant le trèfle (*Trifolium pratense*) et l'autre la luzerne (*Medicago sativa*). Ces deux races d'hôtes correspondent parfaitement à la définition proposée par Drès et Mallet car de nombreuses études menées sur ce modèle ont permis de mettre en évidence des différences génétiques entre les deux (Via 1999), la présence de migrants et une meilleure performance des migrants sur leur hôte natif (Via 1999; Via *et al.* 2000). Cependant, des études récentes ont permis d'identifier onze races associées à différents hôtes pour *A. pisum* (Peccoud *et al.* 2009), ce qui a amené les scientifiques à revoir leur qualification en tant que races d'hôte, et leur préférant le terme de biotypes (Peccoud et Simon 2010). Les biotypes sont définis comme des populations entre lesquelles il y a un niveau de différenciation génotypique ou phénotypique indéterminé mais dont le niveau de divergence est inférieur ou égal à celui qui existe entre espèces (Drès et Mallet 2002). Toutefois le puceron du pois avec ses différents biotypes semble correspondre à un bon modèle de spéciation écologique, avec des événements répétés de spéciation écologique vers différents hôtes (Peccoud et Simon 2010).

1.4. Etude des mécanismes d'isolement reproducteur

Au cours du processus de spéciation écologique, la sélection divergente engendre l'évolution de barrières au flux de gènes vers un isolement reproducteur plus ou moins total. L'isolement reproducteur peut avoir lieu avant (pré-zygotique) ou après (post-zygotique) l'accouplement et implique différents types de barrières (Figure 8). Ce paragraphe sera

consacré à l'étude des barrières reproductrices impliquées, le plus souvent, dans la spéciation chez les insectes phytophages.

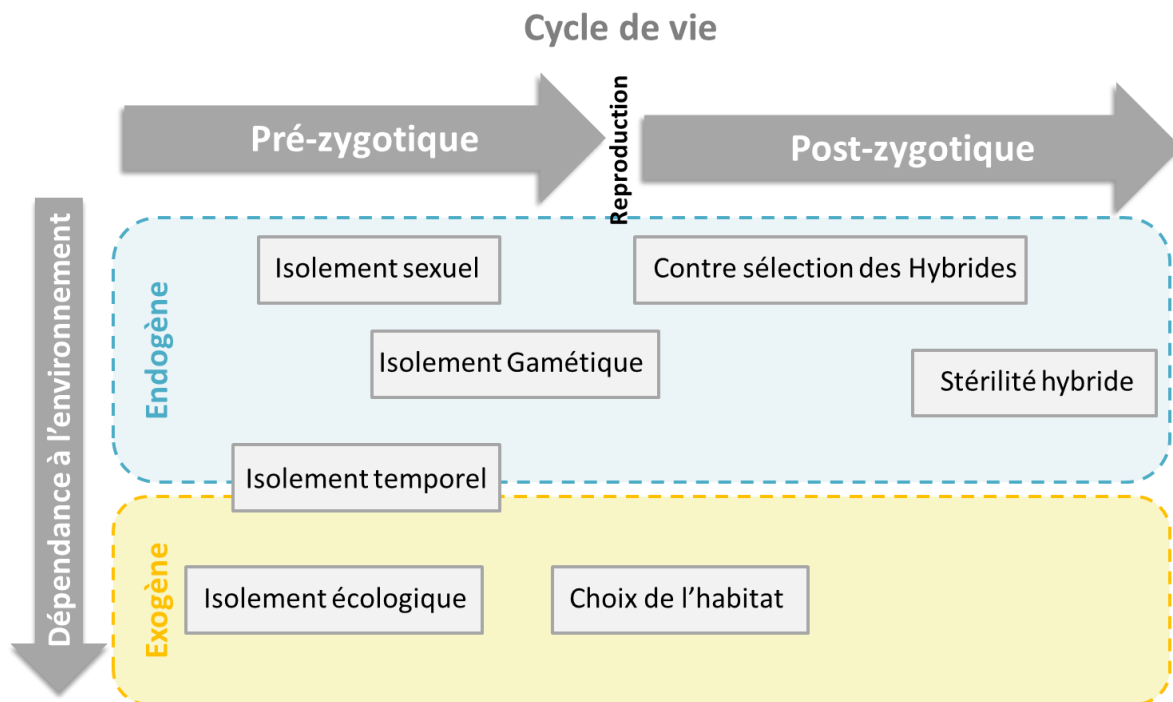


Figure 8 : Classification des mécanismes d'isolement reproducteur en fonction de leur dépendance à l'environnement et du stade du cycle de vie au moment duquel ils interviennent (adapté d'après Thomas *et al.*, 2010 p84)

1.4.1. Les barrières qui conduisent à l'isolement reproducteur

1.4.1.1. Isolement pré-zygotique

Isolement écologique

L'isolement pré-zygotique désigne les barrières extrinsèques qui tendent à réduire la probabilité de rencontre entre deux espèces. Cette catégorie d'isolement reproducteur peut prendre plusieurs formes en fonction de la nature de la barrière mise en jeu. On distingue alors l'isolement lié à l'habitat et l'isolement temporel, mais également l'isolement par la contre-sélection des migrants.

Il y a isolement par l'habitat quand les populations montrent des préférences pour des habitats différents et que cela diminue la probabilité de rencontrer des individus d'une population différente. L'isolement pré-zygotique lié à l'habitat est souvent rencontré lorsque

l'accouplement a lieu dans l'habitat ou encore quand il y a une préférence pour l'hôte natif, ce qui est le cas pour certains insectes phytophages (Funk *et al.* 2002). Beaucoup d'études sur les divergences entre hôtes différents sont conduites sur des populations de laboratoire (Funk *et al.* 2002; Matsubayashi *et al.* 2010). Cependant, il existe peu d'exemples avérés pour lesquels l'observation d'une divergence de préférence d'hôtes conduit à la formation de barrières reproductrices. Des études menées en cages sur la mouche *Eurosta solidaginis* (Craig *et al.* 1993; Itami *et al.* 1998) ou encore des études des marquages/recaptures sur une autre mouche *Rhagoletis pomonella* (Feder *et al.* 1994), font exception.

L'isolement temporel repose sur le même principe que l'isolement lié à l'habitat. Si des populations ont des temps de développement différents ou des périodes d'éclosion différentes ou encore présentent un décalage dans le temps de la période d'accouplement, la probabilité que les individus des deux populations se rencontrent et échangent des gènes sera très faible, entraînant ainsi la formation d'un isolement pré-zygotique entre les deux populations. Chez la noctuelle *Ostrinia nubilalis* les deux races d'hôtes (l'une sur maïs et l'autre sur armoise) présentent un décalage dans le temps lors de l'émergence des adultes (Thomas *et al.* 2003). Pour un autre lépidoptère, *Enchenopa binotata*, on observe un décalage de la période d'éclosion des œufs : les œufs d'*E. binotata* dont la plante hôte est l'orme éclosent plus tôt que ceux de *E. binotata* dont la plante hôte est le robinier faux acacia (Wood et Sheldon 1982).

La contre sélection des migrants est une barrière pré-zygotique spécifique à la spéciation écologique, qui se produit quand les migrants d'une population ont une survie inférieure à celle des non migrants car ils sont moins bien adaptés à l'habitat non natif. Cela tend à diminuer la probabilité de rencontre entre populations différentes et donc à réduire le flux de gènes entre populations. La « non viabilité des migrants » est une barrière pré-zygotique qui a été décrite assez récemment (Nosil 2004; Nosil *et al.* 2005; Giraud 2006) mais il apparaît que cette barrière est assez répandue, notamment chez les insectes phytophages (Nosil 2012b).

Isolement par reconnaissance sexuelle

Dans ce cas, quand il n'y a pas de barrière écologique, les individus de deux populations différentes peuvent se retrouver au même endroit et au même moment.

Cependant l'accouplement est la plupart du temps précédé par une phase de reconnaissance du partenaire soit par reconnaissance visuelle, tactile, auditive ou encore chimique. Et c'est lors de cette phase de reconnaissance que peuvent exister des barrières reproductrices. Chez les insectes la reconnaissance du partenaire se fait très souvent par la communication chimique, et notamment par l'émission de phéromones sexuelles. Plusieurs études ont montré l'importance du rôle des phéromones dans l'évolution et la spéciation (Löfstedt *et al.* 1991; Wicker-Thomas 2011), comme par exemple, pour les lépidoptères (Roelofs et Rooney 2003; Thomas *et al.* 2003) ou les coléoptères (Symonds et Elgar 2004) ou encore les diptères (Fang *et al.* 2002). L'isolement reproducteur peut également être lié à la reconnaissance visuelle du partenaire, comme par exemple, dans le cas des espèces sœurs de papillons mimétiques : *Heliconius melpomene* et *Heliconius cnydo*. Les deux espèces sœurs ont divergé en sympatrie par mimétisme vers des papillons appartenant à des taxons différents (McMillan *et al.* 1997). Le patron alaire des deux espèces étant très différent, la reconnaissance entre les individus des deux populations est réduite et ainsi l'hybridation dans la nature est rare (Jiggins *et al.* 2001).

Isolement pré-zygotique ayant lieu après l'accouplement

Une fois que la reconnaissance sexuelle entre partenaire est effectuée, des barrières reproductrices peuvent agir et empêcher la fécondation et donc la formation du zygote. Tout d'abord il peut y avoir un isolement mécanique entre les partenaires, qui correspond à une incompatibilité physique entre les organes reproducteurs empêchant la fécondation. Chez les insectes, la grande diversité des *genitalia* laisse penser que l'isolement mécanique, « système clé-serrure », peut être une barrière reproductrice assez courante (Sota et Kubota 1998; Hosken et Stockley 2004; Arbuthnott *et al.* 2010). De plus il peut y avoir des incompatibilités entre les gamètes de partenaires de populations différentes, au niveau par exemple de la correspondance moléculaire de l'enveloppe externe des gamètes (Swanson et Vacquier 2002).

La présence d'isolement pré-zygotique peut être également liée à la présence de bactéries chez l'hôte. Le cas des bactéries endosymbiotiques du genre *Wolbachia* est un exemple bien connu, notamment chez les arthropodes (bactéries présentes dans environ 20% des espèces). Ces bactéries sont connues pour s'accumuler dans les organes génitaux et induire une féminisation de la descendance via des processus d'élimination des mâles ou

« *male killing* », de féminisation des mâles, ou encore d'incompatibilité cytoplasmique (Engelstädter et Hurst 2009; Cordaux *et al.* 2011). La présence de ces bactéries peut engendrer des incompatibilités génétiques chez les hybrides, comme l'incompatibilité cytoplasmique (IC) (Figure 9), qui suggère que la présence de bactéries endosymbiotiques aurait un rôle dans la spéciation chez les eucaryotes (Brucker et Bordenstein 2012).

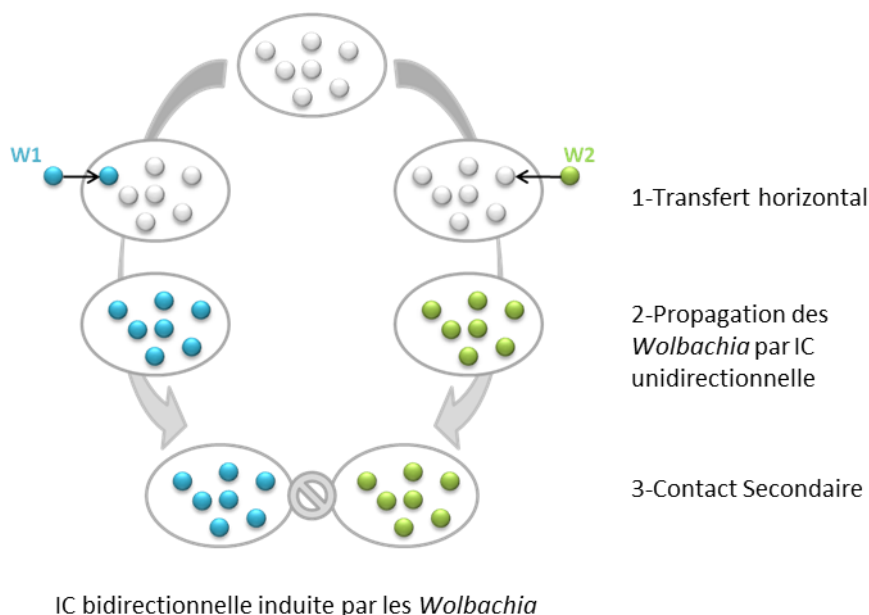


Figure 9 : Représentation schématique d'incompatibilité cytoplasmique (IC) induite par la présence de *Wolbachia* (d'après Brucker et Bordenstein, 2010). Une population ancestrale (cercle gris) contenant un ensemble d'individus non infectés (ronds blancs) se divise (en allopatrie) en deux populations qui sont ensuite contaminées par transfert horizontal, par des *Wolbachia* de souches différentes (ronds de couleur différente W1 et W2). Cette infection se propage dans la population par incompatibilité cytoplasmique unidirectionnelle. Ce phénomène donne un avantage sélectif aux femelles contaminées (qui transmettent les bactéries cytoplasmiques) en provoquant la mort des embryons issus d'un croisement entre un mâle infecté et une femelle non infectée. Les femelles infectées ne sont pas affectées par une diminution de leur valeur sélective car elles sont compatibles avec les mâles infectés ainsi qu'avec les non infectés. Par conséquent, l'IC unidirectionnelle peut propager rapidement les bactéries dans les populations d'accueil. Lors d'un contact secondaire de ces populations en sympatrie ou en parapatricie, l'IC bidirectionnelle provoque incompatibilité réciproque dans les deux sens de croisements. Ainsi des nouvelles espèces peuvent apparaître sans différences morphologiques ou génétiques.

Enfin, dans certains cas la létalité ou la stérilité hybride peut être liée à la présence de bactéries symbiotiques intestinales (Brucker et Bordenstein 2013). Il a été mis en évidence la présence d'un signal phylosymbiotique entre les guêpes du genre *Nasonia* et leurs bactéries intestinales, (Brucker et Bordenstein 2011) pouvant être une conséquence d'une évolution rapide des gènes de l'immunité en réponse à l'adaptation du microbiome, dans un cas classique de « course à l'armement ». Dans ce cas, l'interaction entre le microbiome d'une

espèce *a* interagit négativement avec le génome nucléaire ou cytoplasmique d'une espèce *b*, conduisant à la mort des hybrides.

1.4.1.2. Isolement post-zygotique

L'isolement post-zygotique a lieu entre deux populations différentes quand les individus issus d'un croisement hétérosécifique ont une valeur sélective inférieure à celle des parents, on parle alors de dépression hybride pour qualifier ces individus. L'isolement peut s'exprimer à différents moments du cycle de développement du zygote : des stades précoces de l'embryon jusqu'au stade adulte et même jusqu'à la prochaine génération. Il existe trois niveaux d'observations de l'isolement post-zygotique (Dobzhansky 1970) : la létalité des hybrides de première génération (F1), la stérilité des hybrides de première génération (F1) et la dégénérescence hybride, qui correspond à une réduction de la valeur sélective chez les individus de la génération hybride suivante (F2) ou chez les individus obtenus en *back cross* (croisement d'un hybride F1 avec un individu possédant l'un des génotypes parentaux). Cela a été particulièrement bien étudié chez les drosophiles (Coyne et Allen Orr 1998). Cependant lorsqu'il y a, même en très faible nombre, présence d'individus F2 ou d'hybrides obtenus en *back cross*, cela signifie qu'un échange de gènes a été possible et donc que l'isolement post-zygotique n'est pas total. L'isolement est également incomplet quand la létalité ou la non-viabilité ne touche qu'une partie de la population hybride, comme par exemple lorsqu'il y a un « effet Haldane ». La règle d'Haldane prédit que la létalité et la stérilité des hybrides affecte plus particulièrement le sexe hétérogamétique (Haldane 1922). Cette règle a depuis été vérifiée au travers de nombreuses études sur une grande diversité de modèles (oiseaux, mammifères, insectes). Chez les lépidoptères par exemple, Presgraves et collaborateurs montrent (Presgraves 2002) que dans 96 % des cas de croisements intersécifiques, les femelles hybrides (sexe hétérogamétique chez les lépidoptères) sont plus touchées par la stérilité ou moins viables que les mâles. Il montre également que plus les populations sont différentes plus l'effet Haldane sera fort. Il existe plusieurs explications génétiques à cet effet, qui seront abordées plus tard (chapitre 1.4.2.).

La sélection négative contre les hybrides peut être due à des barrières intrinsèques, c'est-à-dire indépendantes de l'environnement. C'est par exemple le cas d'incompatibilité génétique entre génomes provenant de deux populations différentes. Des gènes différents

entre les populations vont créer une interaction négative s'ils sont en présence chez l'hybride. Ces incompatibilités vont alors réduire le succès reproductif de l'hybride (Coyne et Orr 2004). La sélection négative contre les hybrides peut également être causée par des barrières extrinsèques, lorsque le phénotype hybride n'est pas en adéquation avec l'environnement et a une valeur sélective inférieure à celle des parents (Rundle et Whitlock 2001; Rundle et Nosil 2005). Il est difficile de faire la distinction entre la non-viabilité des hybrides due à des barrières intrinsèques de celle liée à des barrières extrinsèques. Chez les insectes phytophages, des études de transplantation réciproque d'hybrides F2 obtenus en *back cross* pour les deux populations parentales, permettent de contrôler l'incompatibilité intrinsèque génétique et de mesurer le succès des hybrides en fonction de l'environnement (Rundle et Whitlock 2001).

Plusieurs études ont montré que ces barrières reproductrices qu'elles soient pré- ou post-zygotiques sont relativement communes lors de processus de spéciation écologique chez les insectes phytophages, avec une occurrence plus grande de certaines, comme la sélection contre les migrants ou l'isolement par l'habitat (Funk *et al.* 2002; Matsubayashi *et al.* 2010). Cependant pour un système particulier, il est fréquent de constater la présence de plusieurs barrières reproductrices différentes (Tableau 1) (Funk *et al.* 2002; Nosil *et al.* 2005; Matsubayashi *et al.* 2010).

Tableau 1 : Cas d'études détaillées d'isolement reproducteur (IR), où plusieurs barrières sont détectées dans le même système d'étude (d'après Nosil, 2012)

Study system	Components of RI observed	reference
Z- and E-pheromone forms of <i>Ostrinia nubilalis</i> corn borers	Temporal isolation, sexual isolation, postmating-prezygotic isolation, some forms of intrinsic postzygotic isolation (e.g., F1 male behavioral dysfunction)	(Dopman <i>et al.</i> 2010)
<i>Henosepilachna vigintioctomaculata</i> and <i>H. pustulosa</i> lady beetles	Habitat isolation, sexual isolation, postmating-prezygotic isolation (specifically, conspecific sperm precedence), intrinsic postmating isolation	(Matsubayashi and Kanadura 2009)
<i>Gryllus furmus</i> and <i>G. pennsylvanicus</i> field crickets	Temporal isolation, immigrant inviability, sexual isolation, intrinsic postmating isolation	(Majora <i>et al.</i> 2009)
<i>Heliconius erato</i> and <i>H. mimeris</i> mimetic butterflies	Immigrant inviability, sexual isolation, extrinsic postmating isolation	(McMillan <i>et al.</i> 1997)
<i>Timema cristinae</i> host ecotypes	Habitat isolation, immigrant inviability, sexual isolation, postmating-prezygotic isolation	(Nosil 2007)
<i>Mimulus lewisii</i> and <i>M. cardinalis</i> monkeyflowers	Habitat isolation, immigrant inviability, sexual isolation	(Ramsey <i>et al.</i> 2003)
<i>Loxia curvirostra</i> crossbill call types	Habitat isolation, immigrant inviability, sexual isolation	(Smith and Beckman 2007)
Lake Victoria cichlide fish	Habitat isolation, temporal isolation, sexual isolation	(Seehausen <i>et al.</i> 1998)

L'étude des deux espèces voisines de coccinelles *Henosepilachna vigintioctopunctata* et *H. pustulosa* montre qu'il y a un effet cumulatif de la présence de plusieurs barrières reproductrices même si elles n'induisent pas un isolement reproducteur total. En effet pour ce complexe d'espèces, six barrières reproductrices de nature différente ont été mises en évidence. Prises une à une, les barrières n'induisent pas un isolement reproducteur complet, cependant l'action conjointe de ces barrières conduit à un isolement reproducteur entre les deux espèces quasi-total (Matsubayashi et Katakura 2009).

1.4.2. Signatures génétiques et génomiques de la spéciation

Les bases génétiques de l'isolement reproducteur post-zygotique, qui constitue une étape ultime dans la spéciation, ont été particulièrement bien étudiées chez les hybrides (Coyne et Orr 2004; Noor et Feder 2006). Dans le contexte de la spéciation écologique, des études se sont intéressées à rechercher le lien génétique pouvant exister entre isolement reproducteur et trait sous sélection, lien qui conduit à la spéciation (Nosil 2012b).

Les mécanismes génétiques de l'incompatibilité hybride

Une explication à l'incompatibilité génétique identifiée chez les hybrides proposée pour la première fois par Dobzhansky (Dobzhansky 1937) et Müller (Muller 1942) a pris le nom de modèle de Dobzhansky-Muller (DMI). Ce modèle considère que lorsque des populations divergent de façon isolée (en allopatrie par exemple), des mutations peuvent s'accumuler et créer de nouveaux allèles, qui restent avantageux dans le fond génétique ancestral, mais deviennent délétères lorsqu'ils sont mélangés aux allèles d'une autre espèce (Figure 10).

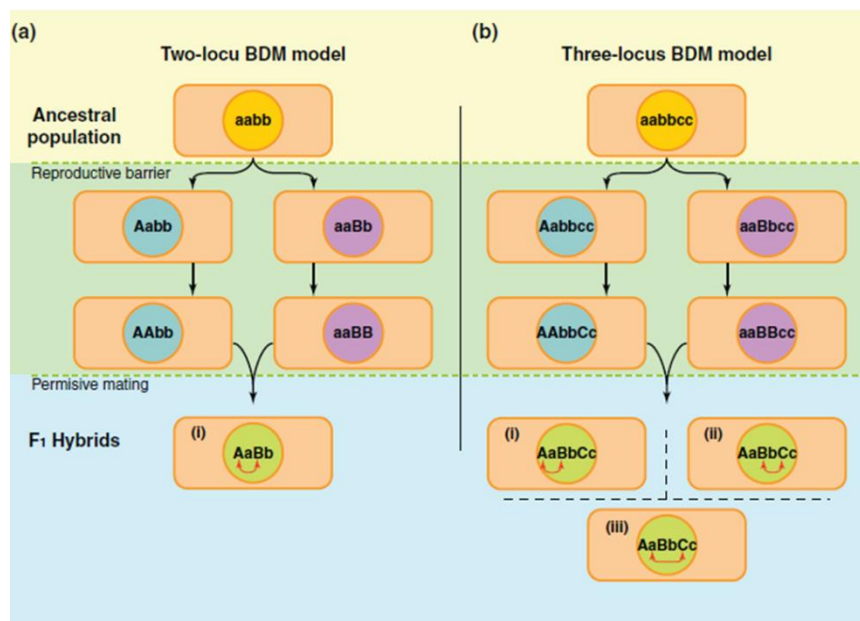


Figure 10 : Représentation schématique de l'incompatibilité de type Dobzhansky–Muller (DMI) (d'après Brucker et Bordenstein, 2012). (a) cas classique de DMI à deux loci quand une substitution au locus *a* et au locus *b* interagissent négativement dans les hybrides (les flèches rouge indiquent l'épistasie négative qui peut potentiellement exister) et qui engendre de l'incompatibilité chez les hybrides. Dans un modèle de DMI à trois loci (b), le nombre d'interaction d'épistasie négative potentiel passe de un à trois.

Pour expliquer le fait que la dépression hybride n'apparaît parfois qu'en deuxième génération ou sur une partie de la descendance F1 (effet Haldane par exemple), il faut prendre en compte la notion de dominance entre allèles (Turelli et Orr 2000). En effet, si les substitutions (allèles *a* et *b*) sont récessives par rapport aux allèles ancestraux (*A* et *B*) alors en F1 le génotype unique *Aa/Bb* peut être viable. Par contre en deuxième génération, les combinaisons génétiques parentales sont brisées par la recombinaison et il y a formation de génotypes homozygotes pour les allèles d'un parent au locus 1 et/ou homozygotes pour les allèles de l'autre parent au locus 2. Les génotypes présentant une homozygotie pour un

allèle nouveau, à un des deux locus seront alors incompatibles. L'effet Haldane est un exemple de DMI asymétrique bien étudié (Wu et Davis 1993; Orr 1995; Laurie 1997; Presgraves 2002). Une des hypothèses, qui explique ces mécanismes génétiques, stipule que les allèles responsables de la stérilité et de la non-viabilité des hybrides sont récessifs. Les hybrides hétérogamétiques sont alors plus touchés car ils expriment totalement le génotype délétère contrairement aux hybrides homogamétiques (Presgraves 2007).

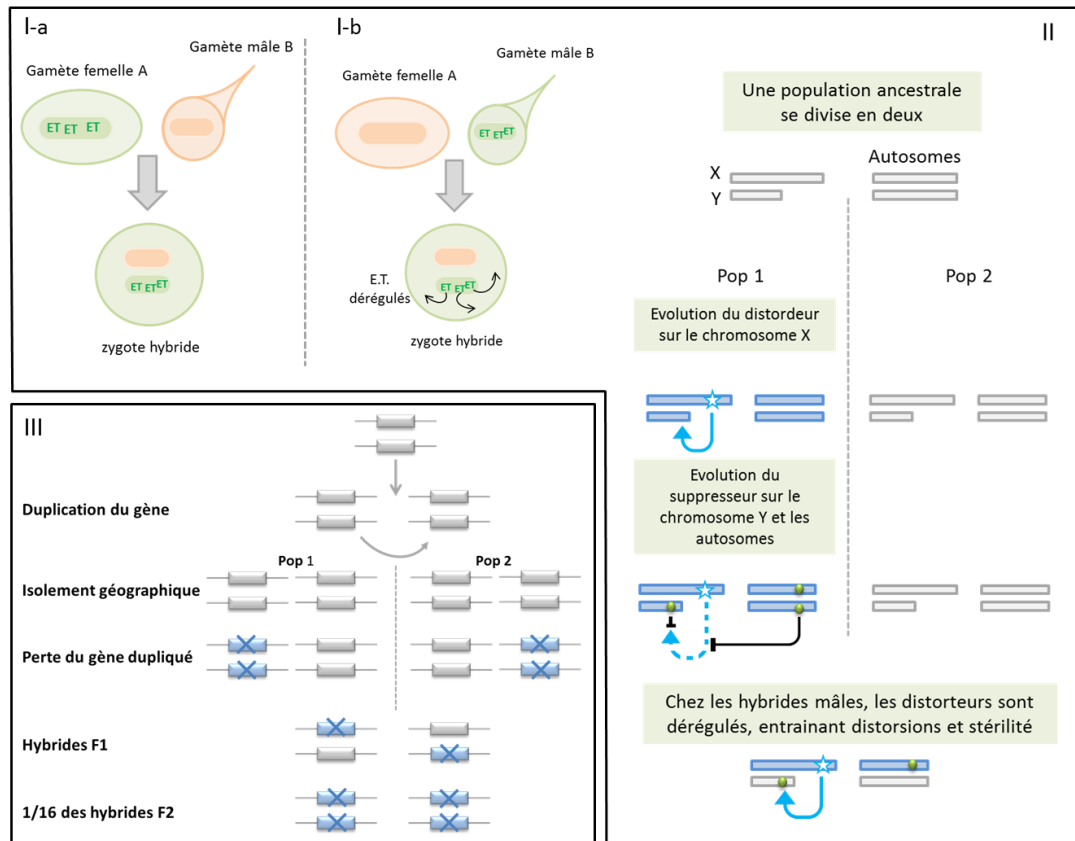


Figure 11 : Evolution des incompatibilités génétiques qui peuvent induire la non-viabilité et la stérilité des hybrides (d'après Presgraves, 2010). *I-Incompatibilité due à la dérégulation des éléments transposables (ET) chez les hybrides.* (a) Lors du croisement d'une femelle d'une espèce A (possédant des ET) avec un mâle de l'espèce B (n'en possédant pas), il n'y aura alors pas d'incompatibilité chez les hybrides car la machinerie de régulation des ET est transmise maternellement. A l'inverse, (b) si c'est le mâle B qui possède des ET, alors la machinerie de régulation des ET n'est pas présente dans l'hybride et il y aura dérégulation de l'élément transposable chez les hybrides. *II-Incompatibilité due à un conflit génétique.* Dans une population donnée, les distorateurs méiotiques peuvent évoluer sur le chromosome X (étoile) ce qui induira l'évolution des suppresseurs de distorsion sur les autosomes afin de permettre la régulation de ces éléments. Chez les hybrides, les distorateurs d'une espèce arrivent dans un fond génétique naïf ne comprenant pas les suppresseurs de distorsion adaptés. Les éléments alors dérégulés pourront induire des distorsions ou même de la stérilité. *III-Incompatibilité due à la présence de gènes dupliqués.* Quand un gène dupliqué est établi dans une population, l'une ou l'autre des copies de ce gène peut être inactivée par la fixation neutre de mutations dégénératives selon les populations. Lors de l'hybridation entre deux populations différentes, et si les gènes ont une fonction essentielle dans la viabilité ou la stérilité, alors les hybrides doubles homozygotes pour les gènes non fonctionnels seront non viables ou stériles.

La dépression hybride peut être également liée à la présence d'éléments égoïstes (*selfish element*) qui « parasitent » les génomes (Presgraves 2007; McDermott et Noor 2010; Presgraves 2010). Ils regroupent les éléments transposables (ET), les supprimeurs de distorsion méiotique (Figure 11 a et b). Ces éléments égoïstes sont en conflit perpétuel avec leur hôte, qu'ils tentent de manipuler pour faciliter leur propre transmission. En réaction, l'hôte quant à lui, tend à supprimer et contrôler ces éléments afin de limiter leurs effets délétères. Occasionnellement cette lutte constante entre les éléments égoïstes et leur hôte peut mener à des dépressions d'hybridation, soit lorsqu'il y a incompatibilité dans l'hybride entre les gènes censés réprimer ou diminuer l'expression des éléments égoïstes, soit quand les éléments égoïstes d'une espèce arrivent dans un fond génomique naïf d'une autre espèce (Johnson 2010; Presgraves 2010). Un dysfonctionnement peut se produire également chez les hybrides quand un gène dupliqué évolue différemment dans deux sous-populations pour donner des copies non fonctionnelles différentes. Les hybrides de deuxième génération peuvent hériter des deux copies paralogues non fonctionnelles et être affectés si le gène est un gène essentiel (Figure 11 c, (Presgraves 2010)).

Enfin des études récentes ont identifié des gènes dit de spéciation (tableau 2), notamment au travers des études sur les drosophiles (Orr *et al.* 2004; Noor et Feder 2006), qui seraient sous-jacents à l'incompatibilité hybride. Ces gènes ont des fonctions très différentes et certains provoqueraient chez les hybrides des « conflits génomiques » (Johnson, 2010), comme par exemple les incompatibilités cyto-nucléaires. Ces incompatibilités correspondent à un dysfonctionnement entre des gènes cytoplasmiques provenant d'une espèce (mitochondriaux la plupart du temps) et des gènes nucléaires provenant d'une autre espèce (Budar *et al.* 2003). Chez les hybrides, les gènes mitochondriaux hérités de l'espèce maternelle A se trouveraient dérégulés dans le cytoplasme de l'espèce paternelle B (Ellison et Burton 2008; Ellison *et al.* 2008).

Tableau 2 : gènes d'incompatibilité hybride (d'après Johnson, 2010).

Table 1. Hybrid incompatibility genes

Gene/gene pair	Gene function	Incompatibility type	Species pair ^a	Selection type	Evidence of conflict
<i>nad6/PPR</i> gene	Respiration	Male sterility	<i>Mgut/Mnast</i>	?	Strong
<i>OLI1/Aep2</i>	Respiration/sporulation	Sterility	<i>Scerv/Sbay</i>	Equivocal (relaxed constraints or positive)	Unlikely
<i>CYC/COX</i>	Respiration	F2 breakdown	<i>Tcal</i>	?	Unlikely
<i>DM1/DM2</i>	Pathogen defense	Necrosis	<i>Atha</i>	Balancing	Very unlikely
<i>Nup96</i>	Nucleoporin (nuclear transport?)	Inviability	<i>Dmel/Dsim</i>	Positive	Suggestive
<i>Nup160</i>	Nucleoporin (nuclear transport?)	Inviability	<i>Dmel/Dsim</i>	Positive	Suggestive
<i>OdsH</i>	DNA binding	Male sterility	<i>Dsim/Dmau</i>	Positive	Suggestive
<i>Prdm9</i>	DNA binding/recombination modification/chromatin modification	Male sterility	<i>Mmus</i>	Positive	Suggestive
<i>Hmr</i>	DNA binding	Male inviability	<i>Dmel/Dsim</i>	Positive	Suggestive
<i>Lhr</i>	Protein-protein binding	Male inviability	<i>Dmel/Dsim</i>	Positive	Suggestive
<i>Zhr</i>	Repetitive DNA	Female inviability	<i>Dmel/Dsim</i>	?	Suggestive
<i>hlx</i>	Heterochromatin (unknown)	Inviability	<i>Dmau/Dsec</i>	?	?
<i>tmy</i>	Unknown	Male sterility	<i>Dsim/Dmau</i>	?	Strong
<i>Ovd</i>	DNA binding?	Male sterility	<i>Dpsd</i>	?	Strong
<i>Xmrk/IR</i> gene	Oncogene/oncogene repressor	F2 inviability	various <i>Xip</i> species	Negative?	?
<i>JY1alpha</i>	Transmembrane ion exchange	Inviability	<i>Dmel/Dsim</i>	None	Very unlikely
<i>S5</i>	Protease	Embryo sac sterility	<i>Osat</i>	?	?
<i>SaM/SaF</i>	Ubiquitin-like modifier/ F-box protein	Male sterility	<i>Osat</i>	?	?

^aAbbréviation des espèces : *Mgut* = *Mimulus guttas*, *Mnas* = *Mimulus nasutus*, *Scerv* = *Saccharomyces cerevisiae*, *Sbay* = *Saccharomyces bayanus*, *Tcal* = *Tigriopus californicus*, *Atha* = *Arabidopsis thaliana*, *Dmel* = *Drosophila melanogaster*, *Dsim* = *Drosophila simulans*, *Dmau* = *Drosophila mauritiana*, *Mmus* = *Mus musculus*, *Dsec* = *Drosophila sechellia*, *Dpsd* = *Drosophila pseudoobscura*, *Xip* = *Xiphophorus*, *Osat* = *Oryza sativa*. Lorsqu'il n'y a qu'un seul nom d'espèce donné, c'est que les hybrides sont issus de croisement entre races géographiques ou entre sous-espèces.

Les mécanismes génétiques de l'isolement reproducteur en présence de flux de gènes

Les mécanismes génétiques précédemment abordés pour expliquer l'isolement post-zygotique entre deux populations sont fondés, pour la plupart, sur l'hypothèse de l'existence d'un isolement de la population ancestrale par des barrières extrinsèques puis d'un contact secondaire révélant l'incompatibilité post-zygotique entre les populations. Cependant, d'autres études montrent que lors de la spéciation avec flux de gènes, des mécanismes génétiques existent qui permettent d'associer la sélection divergente des traits écologiques avec les traits responsables de l'isolement reproducteur (Nosil 2012b). Deux processus génétiques sont envisagés. Le premier correspond à la pléiotropie et le second correspond au déséquilibre de liaison (Rice et Hostert 1993; Kirkpatrick et Ravigné 2002).

Dans le premier cas, un gène unique a un effet sur deux traits phénotypiques : un « trait » écologique (sous sélection) et un « trait » qui entraîne un isolement reproducteur (Figure 12). Il existe chez les insectes phytophages quelques exemples de pléiotropie pour

adaptation et isolement reproducteur. Chez le puceron du pois, par exemple, des études de cartographie QTL (locus de trait quantitatif) ont montré que les allèles conférant un taux de fertilité élevé sur luzerne avaient l'effet inverse sur trèfle et *vice versa* (Via 2002). La sélection peut également agir sur des gènes impliqués dans la phénologie (Feder *et al.* 1998) menant alors à un isolement temporel ou encore sur des gènes impliqués dans la morphologie (Kronforst *et al.* 2006) conduisant à un isolement par reconnaissance sexuelle.

Dans le second cas, les gènes sous sélection divergente sont physiquement différents de ceux induisant un isolement reproducteur. L'évolution de l'isolement reproducteur va en fait être liée à l'appariement non aléatoire (déséquilibre de liaison) des gènes d'isolement reproducteur avec ceux des traits écologiques. L'existence d'un déséquilibre de liaison entre deux gènes peut avoir lieu qu'ils soient ou non liés génétiquement (Figure 12).

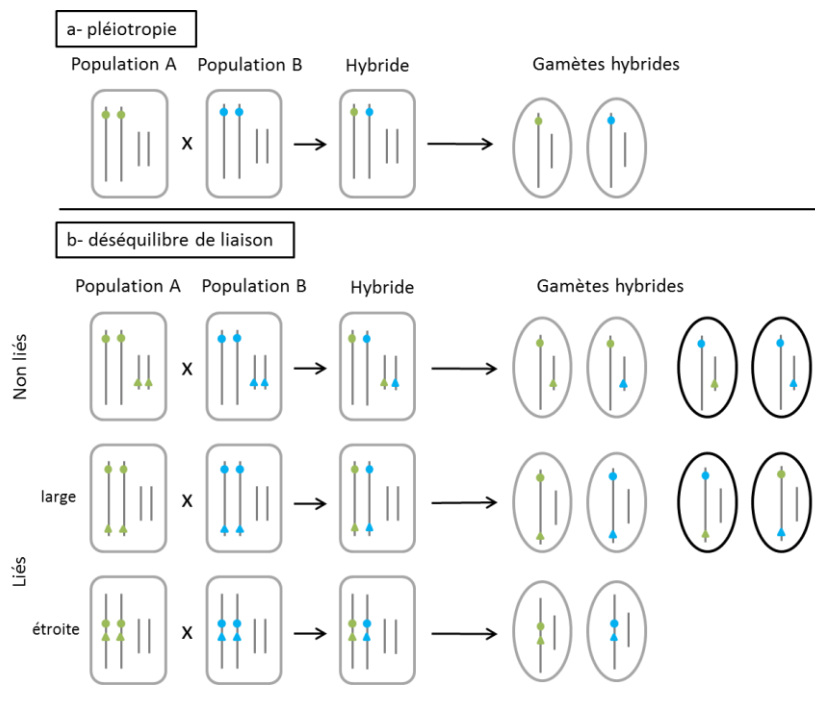


Figure 12 : Mécanismes génétiques qui lient la sélection divergente à l'isolement reproducteur (d'après Matsubayashi *et al.*, 2010). Deux loci sont représentés pour chaque chromosome. Les cercles représentent les loci soumis à la sélection divergente et les triangles sont les loci de l'isolement reproducteur. Les cercles verts sont les loci qui confèrent un avantage au sein de l'environnement A et les cercles bleus sont ceux qui confèrent un avantage au sein de l'environnement B. En cas de pléiotropie (a), le gène sélectionné et le gène de l'isolement reproducteur sont les mêmes et les deux traits seront co-transmis (b) Lorsque le gène sous sélection divergente et celui qui est responsable de l'isolement reproducteur sont différents, ils peuvent être co-transmis à cause d'un déséquilibre de liaison. Une grande proximité génétique entre les gènes peut maintenir un déséquilibre de liaison dans la descendance et mime les effets de la pléiotropie. Des gènes non liés ou bien liés de manière trop distante laissent place à la recombinaison entre les loci, ce qui peut rompre le déséquilibre de liaison dans une partie de la descendance (gamètes hybrides cerclés en noir).

En effet, si les gènes ne sont pas liés (sur des chromosomes différents) ou si la liaison est éloignée (sur le même chromosome mais éloignés l'un de l'autre), le déséquilibre de liaison peut décliner du fait de la recombinaison car des gènes éloignés recombinent plus facilement. A l'inverse une liaison étroite des deux gènes peut « mimer » la pléiotropie et induire un fort déséquilibre de liaison et donc induire l'évolution de l'isolement reproducteur (Ortiz-Barrientos *et al.* 2002; Servedio *et al.* 2009). Certains facteurs peuvent contribuer au maintien du déséquilibre de liaison, comme par exemple les inversions chromosomiques (Noor *et al.* 2001; Rieseberg 2001). Les inversions permettraient le maintien de combinaisons génétiques favorables en empêchant l'introgression de larges régions du génome et en évitant que ces régions soient « cassées » par la recombinaison (Butlin 2005; Hoffmann et Rieseberg 2008). Un exemple d'inversion favorisant le couplage entre gènes sous sélection et isolement reproducteur a été décrit par Joron et ses collaborateurs (Joron *et al.* 2011). Chez le papillon de jour *Heliconius numata*, le patron alaire est contrôlé par un supergène (groupement de loci) localisé au locus P (à une position génomique orthologue au complexe HmYb–HmSb–HmN d'*Heliconius melpomene*). Il existe au sein de la même espèce sept patrons alaires différents dans des sous populations en sympatrie qui miment les patrons alaires d'autres espèces différenciées du genre *Melinae*. Chaque morphe est contrôlé par un allèle spécifique au locus P. Le supergène de *H. numata* couvre 400kb dans lesquels il y a absence de recombinaison. Une zone d'un mégabase autour de cette région a été séquencée et comparée avec celle de *H. melpomene*. Deux segments de 400 kb (18 gènes) et de 180 kb (13 gènes) diffèrent chez *H. numata* par rapport à la référence dans l'ordre des gènes alors que le reste de la région est colinéaire. L'analyse de cette région génomique dans d'autres individus de populations naturelles montre qu'à quatre morphes différents correspond un type de réarrangement de la région, et les réarrangements sont des inversions.

Plus récemment, la théorie de la « divergence par auto-stop » (*divergence hitchhiking*) se base sur l'idée que la sélection divergente va engendrer une diminution du flux de gènes et ainsi diminuer le taux de recombinaison entre les populations ayant des habitats différents (Nosil *et al.* 2005; Via et West 2008; Via 2009; Kobayashi et Telschow 2011). La réduction de la recombinaison permettrait alors la formation de plus grandes régions de différenciation autour des gènes sur lesquels s'applique la sélection divergente.

Ce principe a été imaginé par Smadja *et al.* (Smadja *et al.* 2008) comme correspondant à de « l'auto-stop ». Le gène sous sélection étant dans le « siège du conducteur » et les loci physiquement liés à ce gène seraient les auto-stoppeurs vers la différenciation. Plus les régions empêchant le flux de gènes seront grandes et plus il est probable que les gènes de cette région seront impliqués dans l'isolement reproducteur. Cependant la taille des régions impliquées dans ce mécanisme semble être variable avec des études qui suggèrent que les régions sont grandes (Emelianov *et al.* 2004; Turner *et al.* 2005; Via et West 2008) et d'autres qui suggèrent l'inverse (Dopman *et al.* 2004; Noor *et al.* 2007; Mäkinen *et al.* 2008; Scascitelli *et al.* 2010).

La génomique de la spéciation

Avec le développement assez récent des nouvelles techniques de séquençages (*Next-Generation Sequencing*) offrant une plus grande accessibilité à des génomes entiers, de nouveaux axes de recherche sont apparus. Une des questions majeures concerne la répartition des zones divergentes le long du génome des populations en cours de différenciation supposées contenir les gènes de spéciation. L'hypothèse de départ est que la divergence n'est pas homogène le long du génome et varierait en fonction du niveau de différenciation des populations comparées. Ces patrons de divergence génomique sont illustrés par la métaphore de l'île ou du continent génomique de différenciation ((Nosil *et al.* 2009; Nosil et Feder 2012a), Figure 13). Une « île génomique » correspond à n'importe quelle région génétique, allant d'un nucléotide à un chromosome entier et pour laquelle il y a significativement plus de différenciation qu'attendu sous l'hypothèse de neutralité (représenté par le niveau de l'océan). Une étude expérimentale sur la mouche du pommier (*Rhagoletis pomonella*) a mis en évidence que la divergence entre les deux races d'hôtes sympatriques était répartie tout au long du génome et donc que la divergence est engendrée par la sélection sur plusieurs régions génomique indépendantes (Michel *et al.* 2010). Les auteurs de cette étude proposent l'hypothèse selon laquelle les « continents » constitués de plusieurs loci différenciés pourraient caractériser les stades précoces de la spéciation. Les « îles » et les « continents » génomiques seraient considérés plus comme des fins différentes d'un continuum où les continents seraient de très grandes îles présentant des reliefs différents.

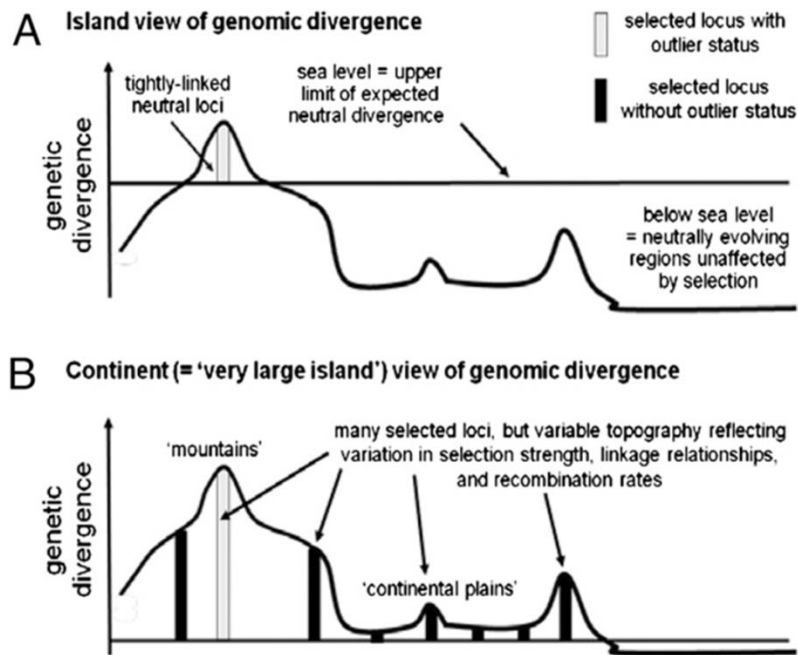


Figure 13 : Représentation schématique de la vision en île (A) ou en continent (B) de la divergence génomique d'après Michel *et al.*, (2010).

Chez les lépidoptères, des études de génomique populationnelle commencent à voir le jour. Dans le genre *Heliconius*, le transfert d'une espèce à l'autre (*H. melpomene amaryllis* à *H. timareta* ssp nov) du patron alaire par mimétisme fait intervenir l'échange entre espèces de larges régions génomiques portant un complexe d'allèles à trois loci HmYb–HmSb–HmN sur le chromosome 15, ce qui permet l'acquisition d'une architecture génétique qu'il serait difficile d'acquérir par accumulation séquentielle de mutations. Le reséquençage ciblé de cette région chez *H. melpomene* a mis en évidence une zone de 150 kb divergente entre des races de cette espèce variant par le patron alaire rouge (Nadeau *et al.* 2012). Les études d'expression génique de cette région génomique ont permis d'identifier le gène *optix* comme le seul gène dont le profil d'expression soit consistant avec un rôle dans la formation du patron alaire chez *H. melpomene* (Reed *et al.* 2011) mais aussi chez *H. erato* qui présente également différentes races variant par la couleur des ailes. Entre toutes ces races, la séquence du gène *optix* ne varie que par des mutations silencieuses et les différences d'expression viendraient de variations dans la région régulatrice. En effet, une étude sur la région inter-génique proche du gène *optix* a mis en évidence chez différentes populations hybrides de *H. melpomene* et *H. erato*, une région divergente conservée et comprenant plusieurs sites directement associés au phénotype au sein de chaque espèce (Supple *et al.*

2013). Ces deux exemples semblent correspondre à des cas où des îlots de divergence génomique constituent le support moléculaire d'une variation phénotypique adaptative.

2. Le modèle biologique : *Spodoptera frugiperda*

2.1. Généralités

Spodoptera frugiperda (J.E. Smith, 1797) appartient à la famille des Noctuidae, qui constitue la plus grande famille de lépidoptères avec plus de 40 000 espèces connues parmi les 174 250 potentielles (Kitching et Rawlins 1998). Les noctuelles regroupent un grand nombre d'espèces dont les larves sont des ravageurs des cultures. *Spodoptera frugiperda* fait partie d'un genre bien connu pour les dommages que ses chenilles occasionnent sur les cultures agricoles : le genre *Spodoptera* comprend une majorité d'espèces appelées *armyworm* (larves « légionnaires », qui lors de pullulations évoquent une armée en marche). Ce genre regroupe une trentaine d'espèces réparties mondialement (Figure 14 (Pogue 2002)) et dont la moitié, au moins, correspond à des ravageurs aux stades larvaires. *Spodoptera frugiperda* couramment appelé *fall armyworm* (FAW) en anglais (légionnaire d'automne) tient son nom de la capacité qu'ont les larves phytophages à se regrouper pour attaquer en « masse » des cultures à l'automne, dans l'hémisphère Nord.

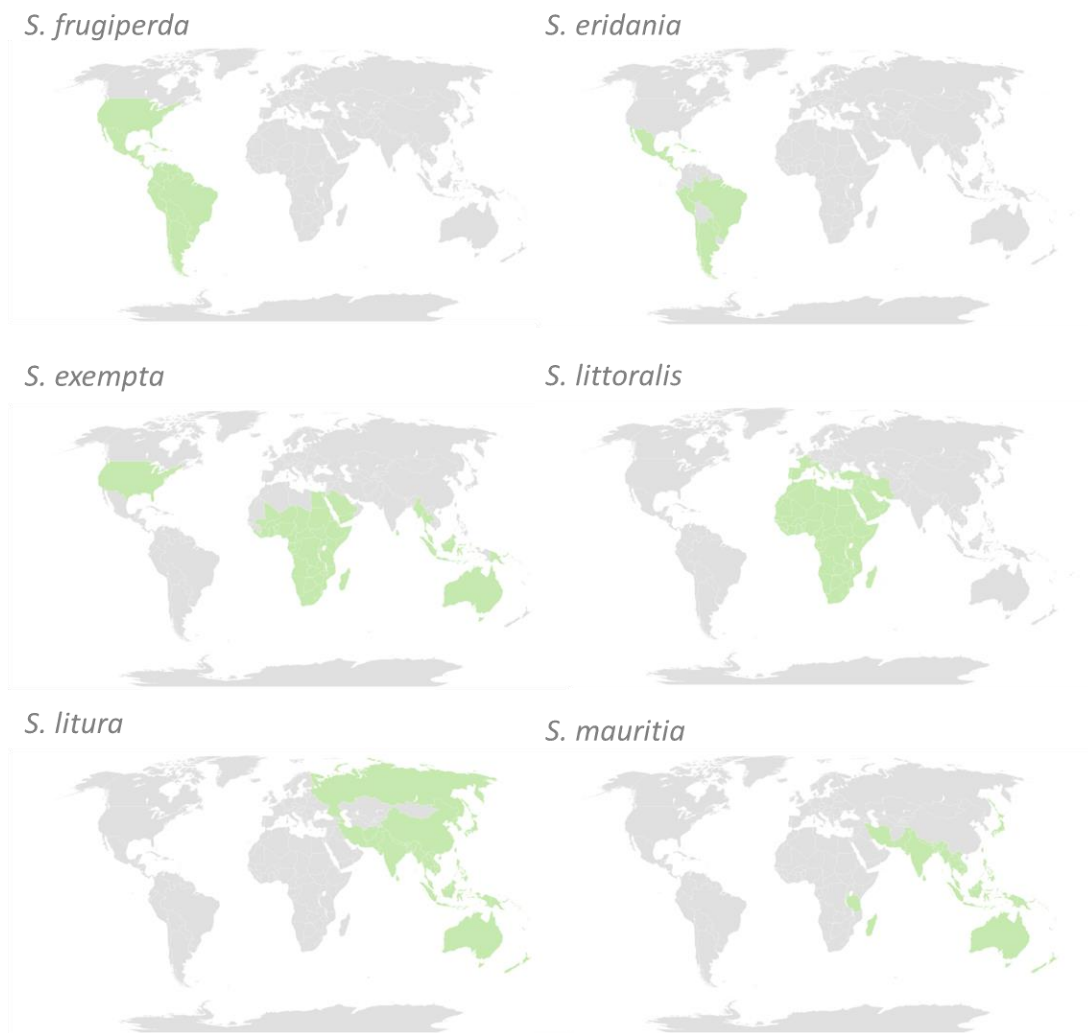


Figure 14 : Aires de répartition des six plus grandes espèces de ravageurs du genre *Spodoptera*

L'espèce *Spodoptera frugiperda* est répartie sur l'ensemble des Amériques (Figure 14). Les larves de *S. frugiperda* ont été retrouvées sur plus de 98 espèces de plantes hôtes différentes, appartenant à plus de 27 familles botaniques différentes (Capinera 1999; Pogue 2002), faisant de lui une espèce hautement polyphage. Une des caractéristiques du FAW est qu'il regroupe deux variants morphologiquement identiques mais présentant des différences d'hôtes. En effet c'est en 1985, que Pashley identifie pour la première fois les deux variants de *S. frugiperda*. Le premier est appelé « variant maïs » et le deuxième « variant riz » car les larves sont retrouvées préférentiellement sur maïs et sur riz respectivement (Pashley *et al.* 1985). Cependant la très grande polyphagie de cette espèce permet d'identifier une série de plante hôtes majoritairement consommées : le maïs (*Zea mays*), le sorgho (*Sorghum spp.*) et le coton (*Gossypium hirsutum*) pour le variant maïs et le riz (*Oriza sativa*), le chiendent

(*Cynodon spp.*), la luzerne (*Medicago sativa*) et le soja (*Glycine max*) pour le variant riz (Pashley *et al.* 1985; Pashley 1986; Nagoshi *et al.* 2007; Juárez *et al.* 2012). Les deux variants sont retrouvés en sympatrie sur toute l'aire de répartition de *S. frugiperda* (Pair *et al.* 1986; Pashley 1986; Machado *et al.* 2008).

Des données génomiques sont connues pour *S. frugiperda* (Gerbal *et al.* 2000). La taille du génome haploïde de *Spodoptera frugiperda* est de 400 Mb réparti en 31 chromosomes. Aucune constriction des chromosomes n'est observée en microscopie, ce qui est caractéristique des génomes holocentriques. En effet comme d'autres espèces de Lépidoptères, d'Hémiptères ou d'Hétéroptères ((Melters *et al.* 2012) pour revue), *S. frugiperda* est une espèce holocentrique avec un kinetochore diffus réparti le long du chromosome (Gerbal *et al.* 2000).

2.2. Biologie générale et cycle de vie

Les cycles biologiques des deux variants de *S. frugiperda* sont identiques et durent de 30 à 40 jours (Figure 15). Les œufs sont pondus la nuit sur les feuilles de la plante-hôte, collés à la face inférieure de la feuille et recouverts d'une couche protectrice de soies abdominales et d'écailles. L'éclosion nécessite 2 à 3 jours. Une fois écloses, les jeunes larves (jusqu'au stade L4) se nourrissent de façon grégaire de la face inférieure des jeunes feuilles créant un aspect caractéristique de « fenêtré ». A partir du stade L5, les larves deviennent cannibales et commencent alors à se disperser sur l'hôte pour éviter la compétition (Chapman *et al.* 1999), c'est à ce moment qu'elles entrent dans la phase dite de « légionnaires » et causent d'important dommages aux cultures. La durée du cycle de vie et des six stades larvaires dépend des conditions climatiques et de leur régime alimentaire, mais il dure environ 15 jours. Le stade larvaire est suivi d'un stade pupa qui se déroule à environ 15 cm dans le sol et qui dure à peu près 12 jours (Capinera 1999). La durée de vie des adultes est d'environ sept jours après émergence pour les femelles et six jours pour les mâles. *Spodoptera frugiperda* est une espèce tropicale typique, adaptée à la vie dans les zones de climat chaud du continent américain : la température de développement larvaire optimale est de 28°C.

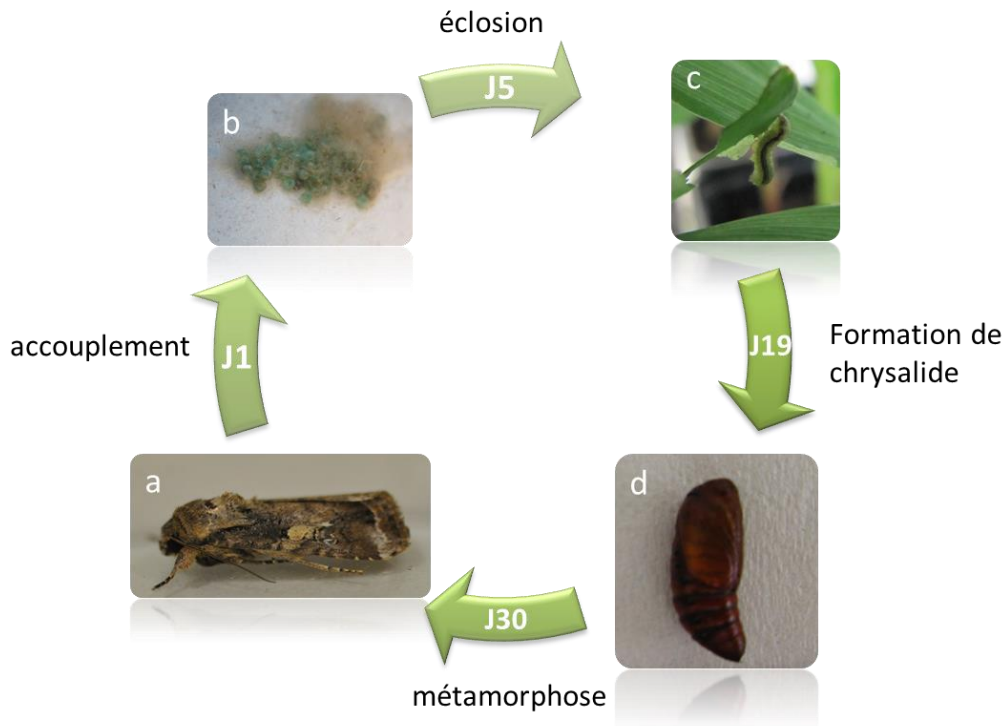


Figure 15 : Cycle de vie de *Spodoptera frugiperda*. Le stade adulte (a) dure environ 5 à 6 jours, après l'accouplement les femelles pondent des œufs (b) qui éclosent au bout de 3 jours environ pour donner des larves (c) qui passeront six stades successifs sur environ 15 jours. Puis les larves se laisseront tomber sur le sol pour s'enfouir dans la terre et se transformer en chrysalide (d). Au bout d'une dizaine de jours, les adultes émergent pour recommencer un nouveau cycle.

Sous les tropiques, la reproduction peut être continue avec quatre à six générations par an. À l'inverse, seulement deux générations peuvent se développer dans les régions plus nordiques. À de plus faibles températures, l'activité et le développement cessent, et si des températures trop basses (<0°C) sont atteintes, aucun stade ne survit généralement. Le *fall armyworm* n'étant pas capable de supporter une exposition prolongée au froid, les adultes présents au Nord du continent américain hibernent en Amérique Centrale et au Sud de l'Amérique du Nord (Floride et Texas aux États Unis) (Nagoshi *et al.* 2011). Cela fait de *S. frugiperda* un excellent migrateur, étant capable de voler du sud du Canada jusqu'au Sud des États Unis.

2.3. Différences existant entre les deux variants

Les deux variants de *S. frugiperda* sont classiquement définis comme des races d'hôtes car ils présentent une série de différences écologiques, comportementales et génétiques.

2.3.1. Ecologiques

Le statut de race d'hôte des deux variants de *S. frugiperda* est relativement difficile à justifier si l'on prend en compte la grande diversité de plantes hôtes identifiées pour chacun des variants (Pashley *et al.* 1985; Pashley 1986; Pashley Prowell et Martin 1987; Nagoshi *et al.* 2007). De plus, plusieurs études menées sur différentes populations de FAW ont montré qu'un pourcentage relativement élevé (19 %) d'individus du variant riz se trouve sur des plants de maïs, alors qu'un pourcentage plus bas (5%) d'individus maïs se trouve sur des herbacées (Prowell *et al.* 2004). Cependant des études portant sur d'autres populations à des périodes différentes présentent des résultats contradictoires, avec la totalité des individus du variant riz retrouvés sur des herbacées et 83% des individus maïs retrouvés sur maïs (Machado *et al.* 2008). Des résultats similaires ont été retrouvés également en Colombie, où 100% des individus maïs ont été retrouvés sur coton (Barros *et al.* 2010). Ainsi bien que l'on puisse distinguer des plantes préférentiellement utilisées par chacun des variants, il semble que le choix de l'hôte ne soit pas si catégorique que cela dans le cas de *S. frugiperda*. De plus, des études sur la survie larvaire et la contre-sélection des migrants ont été menées et ont donné lieu à une série de résultats contradictoires. En effet certaines études mettent en évidence un effet négatif (poids des larves diminué et temps de développement allongé) sur les individus d'un variant élevé sur la plante hôte de l'autre variant (Pashley Prowell et Martin 1987; Whitford F *et al.* 1988; Pashley Prowell *et al.* 1995), alors que d'autres études montrent l'inverse (Meagher *et al.* 2004; Groot *et al.* 2010). Enfin deux études sur le choix de la plante hôte lors de l'oviposition suggèrent que le variant riz choisit préférentiellement les herbacées lors de l'oviposition alors que les résultats sont plus contrastés pour le variant maïs bien qu'une tendance se dégage vers une préférence pour le coton ou le sorgho (Whitford F *et al.* 1988; Meagher *et al.* 2011).

2.3.2. Comportementales

Il existe entre les deux variants de *S. frugiperda* des différences comportementales de nature différente. La première correspond à un décalage dans le temps, ou allochronisme, du comportement de reproduction entre les deux variant maïs et riz (Pashley et Ke 1992; Schöfl *et al.* 2009). En effet le variant maïs serait actif en début de nuit contrairement au variant riz qui lui serait plus actif en fin de nuit, cependant ces deux périodes d'activité sexuelle se chevauchent. Le deuxième correspond à une différence dans la proportion des composés présents dans les phéromones sexuelles (Groot *et al.* 2008). En effet, l'étude des quatre phéromones femelles connues pour leur pouvoir attractif sur les mâles (Z9-14:Ac, Z7-12:Ac, Z11-16:Ac et Z9-12:Ac) montre que le cocktail de phéromones des femelles du variant maïs contient plus de Z11-16:Ac et significativement moins de Z7-12:Ac et de Z9-12:Ac que le cocktail de phéromone des femelles du variant riz (Figure 16).

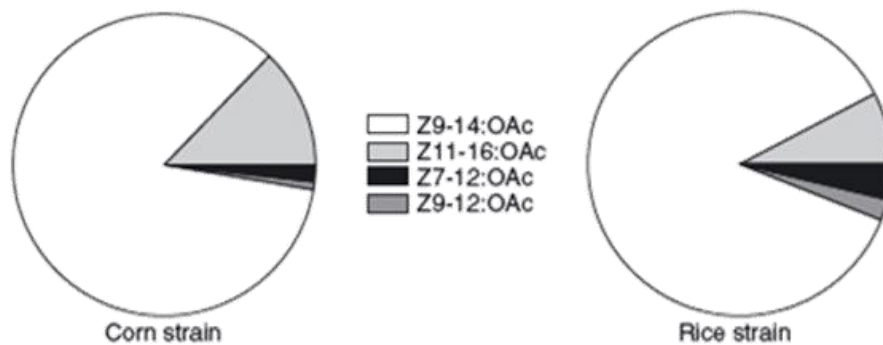


Figure 16 : Diagramme de la composition des différents composés présents dans le cocktail de phéromones sexuelles du variant maïs et riz de *S. frugiperda* (d'après Groot *et al.*, 2008).

2.3.3. Génétiques

Les deux variants bien que morphologiquement identiques sont identifiables génétiquement. En effet, une série de marqueurs a permis de mettre en évidence un polymorphisme génétique entre les deux variants. Les individus du variant maïs et du variant riz sont différents au niveau de la séquence ADN mitochondrial codant pour la sous unité I de la cytochrome oxydase (COI) et de la sous unité I de la nicotinamide adénine dinucléotide déshydrogénase (NADH I) (Pashley 1989; Pashley et Ke 1992; Levy *et al.* 2002; Meagher et

Gallo-Meagher 2003; Nagoshi *et al.* 2006). En ce qui concerne l'ADN nucléaire, un certain polymorphisme a été mis en évidence grâce à des marqueurs AFLP (*Amplified Fragment Length Polymorphism*) (McMichael et Pashley Prowell 1999; Prowell *et al.* 2004). De plus les individus du variant riz sont caractérisés par la présence, en grande quantité, d'une séquence répétée sur le chromosome sexuel Z chez les femelles (Lu *et al.* 1994; Lu et Adang 1996). Cette séquence est appelée FR pour *Fall armyworm Repeat* et permet une identification rapide des individus (Nagoshi et Meagher 2003; Nagoshi et Meagher 2003).

Des études menées en laboratoire sur la capacité des deux variants à s'hybrider ont conduit là encore à des résultats contradictoires. Dans une première étude (Pashley Prowell et Martin 1987), l'accouplement entre les deux variants n'était possible qu'entre une femelle « riz » et un mâle « maïs », le croisement réciproque n'étant pas fertile. De plus en back cross les femelles hybrides ne s'accouplent que difficilement avec leurs frères et pas du tout avec tout autre mâle (Pashley Prowell et Martin 1987; Whitford F *et al.* 1988; Groot *et al.* 2010). A l'inverse des études mettent en évidence qu'il n'y a aucun biais d'accouplement entre les deux variants (Quisenberry 1991; Meagher *et al.* 2011). De plus, des études réalisées à partir de données issues de populations naturelles de *S. frugiperda*, ont montré qu'il y avait un fort taux d'hybridation dans la nature comme au Brésil (14% d'hybrides) et aux Etats-Unis et Amérique Centrale (16% d'hybrides) (Prowell *et al.* 2004; Nagoshi *et al.* 2009). Enfin, une dernière étude récente a mis en évidence une diminution de la valeur sélective pour les hybrides issus de croisement entre les deux variants (Velásquez-Vélez *et al.* 2011). En effet, les auteurs montrent que le nombre de femelles hybrides (sexe hétérogamétique) est moins important que celui des mâles hybrides et qu'il y a une diminution générale de la fertilité des hybrides. Ces résultats suggèrent donc la possibilité d'un effet Haldane pour les deux variants de *Spodoptera frugiperda*.

2.4. Un ravageur des cultures

Spodoptera frugiperda est connu comme étant l'un des plus grand ravageurs de l'hémisphère Ouest, notamment des cultures de maïs, de coton et de soja (Sparks 1979; Capinera 1999; Capinera 2008; Barros *et al.* 2010). L'importance des dommages causés par les larves de *S. frugiperda* dépend également du stade de développement de la plante. Par exemple pour le maïs, les larves colonisent tout d'abord les plantes lors du stade précoce de

développement : le stade verticilles (Figure 17). A ce moment les larves sont protégées par la formation de la spirale de la feuille dans le cornet des feuilles (Barros *et al.* 2010).

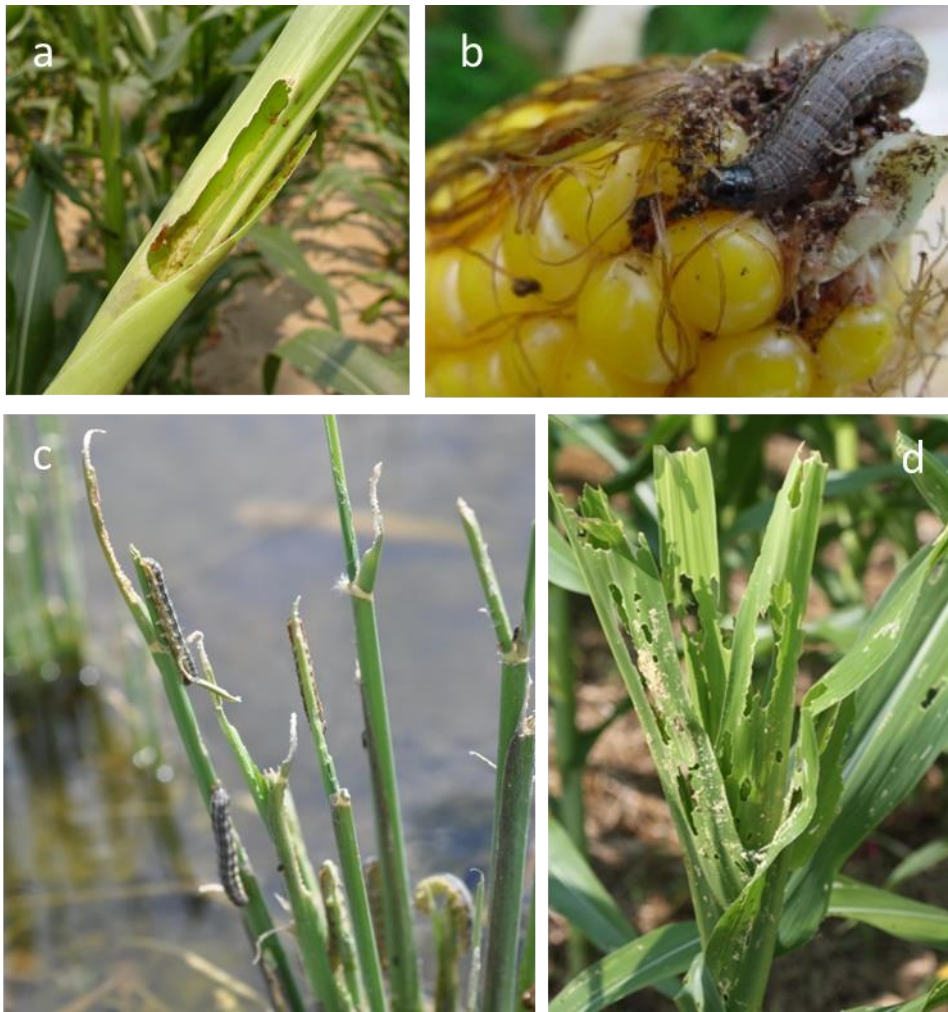


Figure 17 : Photographies de dommages engendrés par les larves de *S. frugiperda* sur maïs (a, b et d) et sur riz(c).

En ce qui concerne les plantes dicotylédones, les larves ne pouvant pas se protéger de la même manière, les structures reproductrices de la plante sont attaquées et ainsi pour le coton par exemple, ce sont les bourgeons et les fleurs qui sont détruits (All *et al.* 1996). Une étude sur l'impact des dommages causés par le FAW sur le développement de la plante (Marenco *et al.* 1992) montre que le stade verticille tardif est beaucoup plus sensible aux attaques que les stades verticilles plus précoces. Les larves de stade plus avancés (stade L6) sont extrêmement défoliatrices et les rendements de production peuvent décroître jusqu'à 57% pour les cultures de maïs en cas d'attaque du FAW (Cruz *et al.* 1999). Dans les régions tropicales, le nombre de générations de *S. frugiperda* peut aller jusqu'à huit si les conditions

climatiques sont adéquates, faisant de lui un ravageur des cultures redoutable en Amérique du Sud et Amérique Centrale (Busato *et al.* 2004).

3. Principaux objectifs de la thèse

Bien que de nombreux points apparaissent controversés dans la littérature, un certain nombre de différences existent entre les deux variants de *S. frugiperda* et amènent à se poser des questions sur le niveau réel de différenciation présent entre ces deux variants. L'objectif général de cette thèse est donc de clarifier le statut d'espèce des variants de *Spodoptera frugiperda*. Cette thèse a été réalisée dans un contexte intégratif, à l'interface entre deux disciplines de la biologie de l'évolution - la génétique des populations et la génomique comparative - et a été réalisée à différentes échelles, soit sur des populations naturelles, soit sur des individus de laboratoire.

Le premier niveau d'analyse de la thèse vise à replacer l'espèce *frugiperda* au sein du genre *Spodoptera* et à clarifier le statut évolutif des deux variants de *S. frugiperda* au travers d'une analyse de délimitation d'espèce. Le deuxième niveau d'analyse a pour objectif de regarder comment les populations des variants maïs et riz de *S. frugiperda* sont structurées sur leur aire de répartition. Ces points seront abordés dans le premier chapitre de la partie résultats.

Le troisième niveau d'analyse consistera à étudier le taux de polymorphisme génétique existant entre les deux variants et les conséquences éventuelles que cela peut avoir sur l'évolution d'un isolement reproducteur entre ces deux variants. Ce chapitre permettra d'aller jusqu'à la comparaison génomique de certaines régions du génome des variants de *S. frugiperda*. Cette étude sera présentée dans le deuxième chapitre de la partie résultats.

Partie 2 : Matériel et Méthode Général

PARTIE 2 : Matériel et Méthode Général

Déroulement chronologique de la thèse

L'ensemble de cette thèse s'est articulée autour de deux thèmes de recherches différents et complémentaires, représentés d'une part par le laboratoire Diversité, Génomes & Interactions Microorganismes-Insectes (DGIMI), au sein duquel toutes les analyses sur populations d'élevages ont été menées et d'autre part par le Centre de Biologie pour la Gestion des Populations (CBGP) où ont été conduite les études sur populations naturelles. Afin de mieux comprendre l'organisation des résultats (Partie 3) il apparaît nécessaire de décrire le déroulement des différentes étapes de la thèse (Figure 18). En effet, les résultats seront présentés non pas dans l'ordre chronologique de leur obtention mais de l'échelle d'analyse la plus large (analyses du genre *Spodoptera* et de délimitation d'espèces : partie 3.1.) à la plus fine (étude génétique et génomique de l'isolement post-zygotique : partie 3.2.).

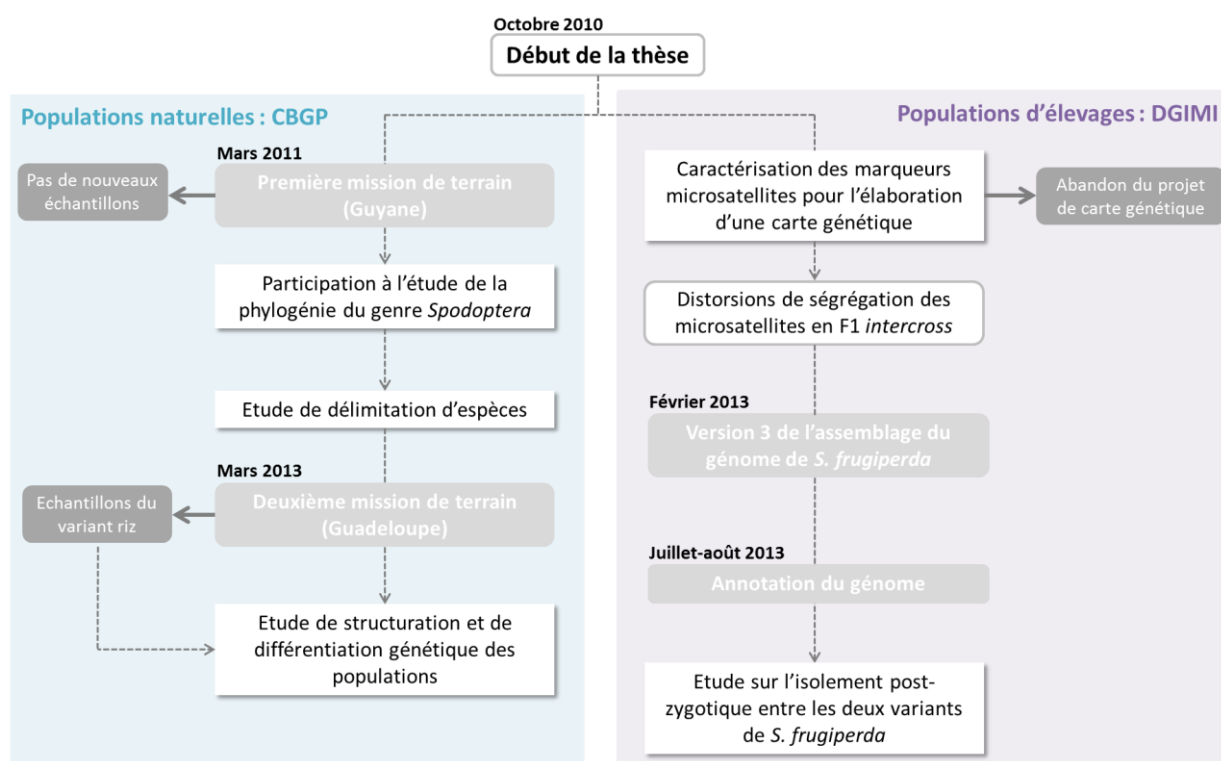


Figure 18 : Organigramme présentant chronologiquement les différentes étapes de la thèse.

1. Les élevages

1.1. *Spodoptera frugiperda*, insecte de quarantaine NS2

L'utilisation de *Spodoptera frugiperda* à des fins scientifiques nécessite un agrément des locaux en vue d'introduire, de détenir et de manipuler certains organismes nuisibles, végétaux, produits végétaux et autres objets énumérés aux annexes I à V de la directive européenne 2000/29/CE. Ceci pour des travaux à des fins d'essai ou à des fins scientifiques (dernière version 2008/61/CE), transcrite en droit français sous le nom « d'arrêté du 22 novembre 2002 relatif aux exigences sanitaires des végétaux, produits végétaux et autres objets ». L'agrément initial des locaux de l'UMR1333 et de l'insectarium de l'Université Montpellier 2 avait été obtenu le 13 octobre 2006 pour une durée de 5 ans, il a été renouvelé par le préfet de la région Languedoc-Roussillon pour 5 ans sous le numéro 2012080-0001 le 20 mars 2012. *Spodoptera frugiperda* est également répertorié sur la liste de quarantaine A1 de l'Organisation Européenne et Méditerranéenne pour la Protection des Plantes (OEPP). Ce ravageur est intercepté occasionnellement en Europe sur du matériel végétal importé (Seymour *et al.* 1985), il est donc évident qu'il pourrait s'introduire dans la région OEPP par les échanges commerciaux. Ainsi la manipulation de ces insectes et son élevage doivent respecter les règles de confinement NS2 (niveau de sécurité 2 sur une échelle allant de 1 à 3, 3 étant le niveau maximal de confinement), qui sont : (i) la manipulation des insectes doit se faire dans des pièces sécurisées ; (ii) les manipulateurs doivent éviter de disperser l'insecte par le port de blouses dédiées dans les salles d'expérimentation, ainsi que de pantalons de sécurité, de masques et de sur-chaussures à l'insectarium.

1.2. Les conditions d'élevage

Les élevages de *S. frugiperda* effectués au cours de cette thèse ont été réalisés à l'insectarium du laboratoire DGIMI à l'Université Montpellier II. Les deux variants de *S. frugiperda* y sont élevés sur milieu artificiel (« Poitout » composé de farine de maïs, d'agarose, de conservateurs et de compléments alimentaires (Poitout et Bues 1974)) dans des cellules indépendantes avec une photopériode (jour/nuit) de 16h / 8h, une température moyenne de 24°C et une hygrométrie moyenne à 40%. L'élevage *S. frugiperda* variant maïs est présent au laboratoire depuis plus de 10 ans et a été créé à partir d'une cinquantaine de

pupes récoltées en Guadeloupe. Le variant riz, quant à lui, est présent à l'insectarium depuis octobre 2010. Il a été élaboré à partir d'une cinquantaine de pupes récoltées dans le comté de Hardee en Floride, près de la ville d'Ona.

Les individus issus des différents croisements intra et inter variant réalisés lors de ma thèse ont été élevés selon le même protocole. Les pupes ont été sexées afin d'isoler les mâles des femelles pour ensuite procéder à l'accouplement d'un mâle naïf avec une femelle naïve. Les pontes de chacun des couples isolés (Figure 19) ont été récupérées et les larves juste écloses ont été ensuite isolées une par une dans des boîtes de culture cellulaire 24 puits, 12 puits et enfin 6 puits au fur et à mesure du développement larvaire (Figure 19). Les larves de *S. frugiperda* étant cannibales aux stades larvaires tardifs (Chapman *et al.*, 1999), l'isolement de chaque larve était nécessaire. Au total, 120 larves hybrides de génération F2 (50 issues du croisement d'une femelle « riz » et d'un mâle « maïs » et 80 issues du croisement réciproque) ont été génotypées. De la même manière en intraspécifique, un total de 45 couples de génération G0 (23 couples « maïs » et 22 couples « riz »), 63 couples de génération F1 (34 et 29 couples formés à partir de la descendance des croisements « maïs » et « riz » précédents formés) et 188 larves de génération F2 ont été génotypées afin de servir de contrôle.

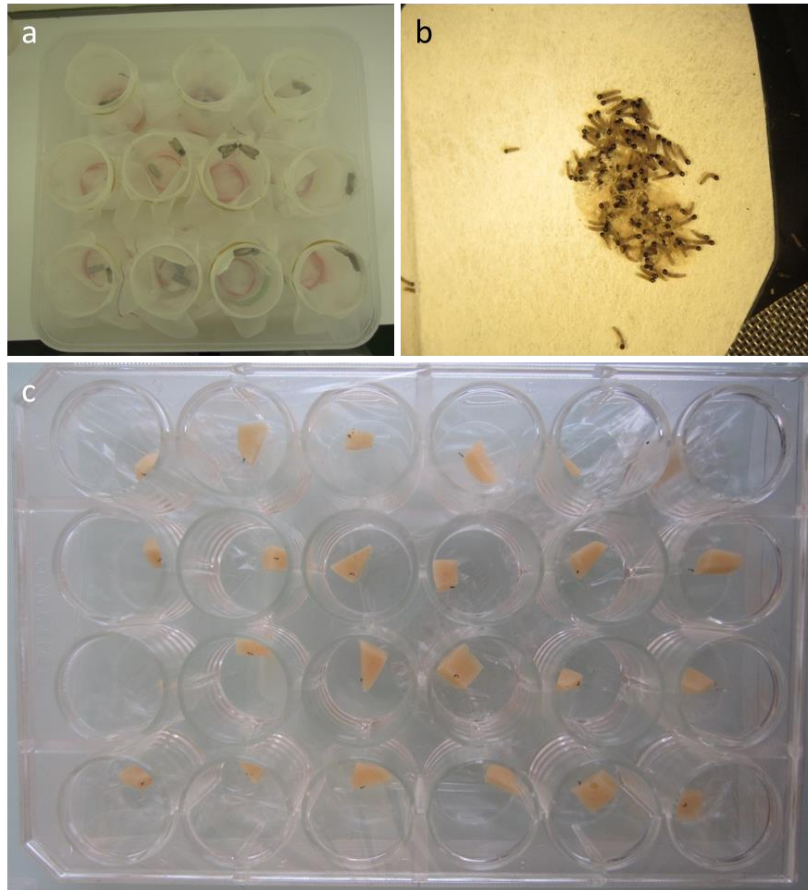


Figure 19 : Photographie du protocole expérimental utilisé pour réaliser des croisements intra et inter variants. a) une série de couples isolés réalisés avec une femelle naïve et un mâle naïf. b) observation sous loupe binoculaire des larves juste écloses. c) Répartition de chaque larve néonate d'une même ponte dans une boîte de culture cellulaire 24 puits.

2. La caractérisation moléculaire des variants de *S. frugiperda*

2.1. Sur population d'élevage

Utilisation de la séquence d'ADN en tandem FR

Avant de procéder aux manipulations d'accouplement, les individus d'élevage ont été identifiés grâce à une séquence répétée en tandem (intitulée séquence FR) présente en grande quantité sur le chromosome Z pour les femelles du variant riz (Nagoshi et Meagher 2003). Le protocole d'amplification du fragment est présenté en Annexe.

Utilisation du site polymorphique de restriction *MspI*

Les individus ont également été caractérisés grâce au polymorphisme qu'il y a pour les séquences d'ADN mitochondrial du gène codant pour la sous-unité du cytochrome c oxydase I (COI) (Lu et Adang 1996). Un fragment spécifique est ainsi amplifié avec les amorces JM76 et JM77 (Levy *et al.* 2002; Nagoshi *et al.* 2006) puis digéré par l'enzyme de restriction *MspI* (Promega). La séquence du fragment COI amplifiée pour les individus appartenant au variant maïs présente un site de restriction pour *MspI* (Figure 20). A l'inverse, la séquence des individus du variant riz ne possède pas cette séquence, le fragment ADN ne sera donc pas digéré.

```

      *      *      *      *      *      *      *      *
1  CGAAAATGACTTTATTCAACAATCATAAAGATATTGGAACATTATATTTTATTTTGGAAATTTGAGCAGGGATAGTAGG
   .....A.....
81  TACTTCITTTAAGTTTATTAAATTCGAGCTGAATTAGGGACTCCAGGATCTTTAATTGGAGATGATCAAATTTATAACTA
   .....A.....
161 TGTAAACAGCTCATGCTTTTATTATAATTTTTTTTATAGTTATACCTATTATAAATTGGAGGATTGGAAATTTGACTTGTA
   .....C.....A.....
241 CCTTTAATATTAGGAGCCCTGATATAGCTTTCCCACGTATAAATAAATAAGTTTTTGACTTTTACCCCATCTTTAAC
   .....T.....
321 TTTATTAATTTCTAGTAGCATTGTAGAAAATGGAGCAGGAAC TGATGAACAGTTTACCCCCCTCTCCTCTAATATTG
   .....
401 CTCATGGTGGTAGTTCAGTAGATTTAGCTATTTTCTCAGTTTATTTAGCTGGAATTTTCATCTATTTTAGGAGCTATTAAC
   .....
481 TTTATTACTACTATTATTAATATACGATTAATAATTTATCATTGTGATCAAATACCTTTTATTATTGAGCTGTAGGTAT
   .....C.....
561 TACTGCATTCCTTATTATTATTATCTTTACCTGTTTTAGCCGGAGCTATTACTATATTACTTACTGATCGAAATTTAAATA
   .....C.....T.....T.....C.....
641 CATCATTTTTCGATCCTGCAGGTGGAGGTGATCCTATTCCTTA 683
   .....A.....

```

←

Figure 20 : Comparaison entre variant maïs et variant riz des séquences d'un fragment du gène COI. La séquence du haut correspond au variant maïs, les similarités sont indiquées par un point. Le site de restriction *MspI* est surligné en jaune. Les séquences des amorces sont surlignées.

Marqueurs microsatellites

La caractérisation des marqueurs microsatellites a été réalisée au début de cette thèse dans l'objectif premier d'élaborer une carte génétique de *Spodoptera frugiperda*. Le principe repose sur l'analyse de la ségrégation des gènes dans les descendances hybrides. L'analyse de la ségrégation de deux marqueurs se base sur la fréquence de recombinaison retrouvée dans la descendance hybride. La recombinaison est liée au processus de crossing-over lors de la méiose (Figure21)

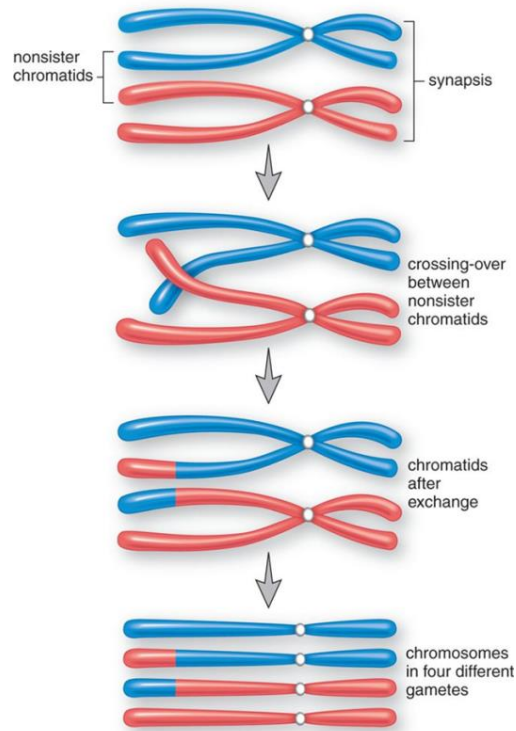


Figure 21 : Représentation schématique du processus de crossing-over lors de la formation des gamètes.

Nous disposons d'une banque de microsatellites obtenue par séquençage nouvelle génération par la technologie 454 ayant permis de définir des couples d'amorces pour 377 marqueurs microsatellites et nous en avons testé 110 pour obtenir 17 marqueurs microsatellites polymorphes entre les deux variants de *S. frugiperda*. Pour ce faire, nous avons utilisé une technique d'électrophorèse capillaire avec amorces marquées. Les amplicons correspondants aux microsatellites sont révélés grâce à l'électrophorèse et seuls les microsatellites polymorphes sont amplifiés avec des amorces fluorescentes qui permettront la réalisation de PCR multiplexes. La technique des amorces queues M13 a été mise au point et adaptée (Annexe) à notre modèle d'après Schuelke (Schuelke 2000). Douze marqueurs microsatellites ont été utilisés pour étudier la ségrégation des gènes chez les individus obtenus en génération F2 pour les différents croisements. Cependant les analyses préliminaires sur la ségrégation de ces marqueurs dans la descendance hybride présentant un fort taux de distorsion, l'élaboration de la carte génétique a été abandonnée pour se concentrer sur l'étude des incompatibilités génétiques existant potentiellement entre les deux variants de *S. frugiperda*.

2.2. Sur populations naturelles

Les analyses menées sur les populations naturelles reposent sur l'exploitation d'échantillons obtenus soit grâce à des collaborations internationales, soit directement collectés sur le terrain lors de missions (Figure 22). Le jeu de données a également été enrichi d'un certain nombre de séquences disponibles dans des bases de données internationales comme GenBank.



Figure 22 : Carte du continent américain présentant l'échantillonnage de *Spodoptera frugiperda*. En vert les échantillons réunis grâce à des collaborations, en gris les séquences récupérées sur GenBank et en rose les localités échantillonnées dans le cadre de la thèse.

L'échantillonnage sur le terrain a été réalisé au cours de trois missions : une en Guyane et deux en Guadeloupe entre 2011 et 2013. Les individus adultes ont été capturés

grâce à des pièges phéromonaux. Des capsules de phéromone de synthèse spécifique de *S. frugiperda* ont été placées dans des pièges deltas (Figure 23) afin d'attirer les mâles qui vont alors s'engluer sur la plaque de colle positionnée au fond du piège. Les larves, elles, ont été échantillonnées à la main directement sur les plantes hôtes.



Figure 23 : Photographies prises lors de l'échantillonnage sur le terrain de FAW adultes, grâce à des pièges à phéromone (piège delta en vert) et la récolte de larves à la main sur plante hôte (*Pennisetum purpureum*).

Les échantillons ont ensuite été génotypés pour deux fragments du gène COI avec les amorces Ron/Nancy (Simon *et al.* 1994) ainsi que Tonya/Hobbes (Monteiro et Pierce 2001) par réaction de PCR (Annexe).

Ces fragments ont permis de réaliser les analyses de délimitation d'espèces (partie 3.1.2.) ainsi qu'une analyse de *barcoding* (Hebert *et al.* 2003), qui correspond au calcul du pourcentage de divergence, pour un fragment du gène COI, entre plusieurs individus. Classiquement (Hebert *et al.* 2004), la divergence entre des séquences d'individus appartenant à une même espèce est supposée être inférieure à la divergence entre les séquences d'individus appartenant à des espèces différentes, ce qui se traduit normalement par un *barcode gap*. A l'aide d'un histogramme qui met en avant les distances intra et interspécifique, on s'attend ainsi à avoir un écart quantitatif important entre les deux classes de distances.

Le jeu de données COI obtenu pour le genre *Spodoptera* a permis de mener une analyse de *barcoding*. Tout d'abord, si le variant riz et le variant maïs de *S. frugiperda* sont considérés comme une seule et même espèce les distances génétiques entre les individus sont relativement élevées et correspondent plus à des distances interspécifiques (Figure 24 a). Cependant, lorsque les analyses sont menées en considérant le variant riz et le variant maïs de *S. frugiperda* comme deux espèces différentes, il apparaît un seuil plus net entre les distances intra et interspécifiques (Figure 24 b). Cependant ces résultats ne permettent pas clairement de mettre en évidence un *barcode gap*, ce qui confirme les résultats de plusieurs articles qui remettent en question la pertinence de l'approche de DNA barcoding pour délimiter les espèces (Hurst et Jiggins 2005; Whitworth *et al.* 2007; Wiemers et Fiedler 2007).

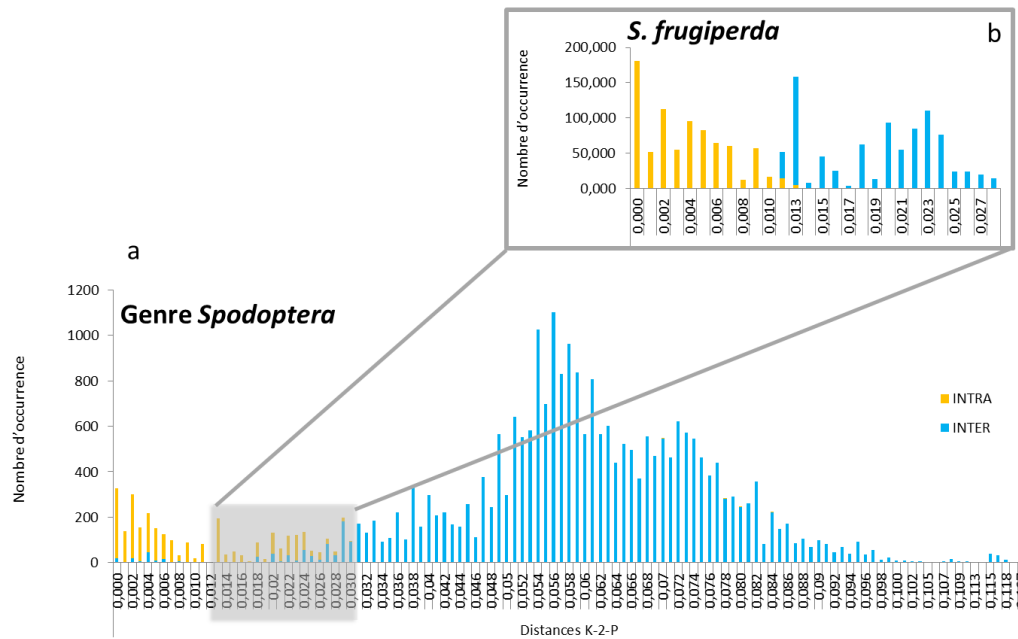


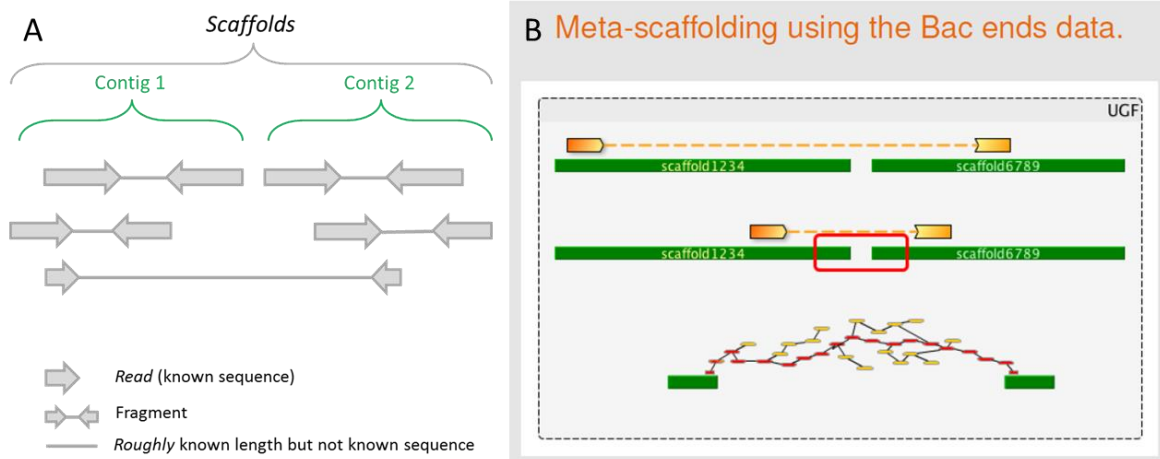
Figure 24 : Analyse de *barcode* pour une région du gène COI : fréquence des distances entre 247 taxa répartis en 25 espèces de *Spodoptera* (a) et parmi 60 taxa de l'espèce *S. frugiperda* (b) ; distances intraspécifiques en jaune, distances interspécifiques en bleu.

3. Le génome de *Spodoptera frugiperda*

Avant le début de cette thèse, un projet de séquençage du génome de *S. frugiperda* variant maïs avait été initié au laboratoire DGIMI grâce à un projet Génoscope coordonné par Philippe Fournier et une partie des résultats a été obtenue au cours de la thèse, nous permettant d'utiliser certains outils génomiques.

3.1. Assemblage du génome

L'assemblage du génome a en partie été réalisé par le Génoscope qui a proposé jusque-là trois versions d'assemblage du génome avec des outils publics de type Velvet et ALLPATHS, à partir de banques de type *paired ends* (séquences chevauchantes) de 170 pb (couverture 165x) et des banques de différentes taille de type *Mate pairs* (extrémité de séquences appariées) réalisées à partir de l'ADN génomique d'un minimum de larves issues d'une même fratrie (2) obtenue à partir d'un couple parental isolé afin de limiter le polymorphisme, plus une larve issue d'un autre croisement pour les banques de fragments les plus larges (10 kb). Les statistiques de la dernière version de l'assemblage du génome (version 3) sont présentées dans le tableau 3.



¹Figure 25 : Schéma de la méthode utilisée pour assembler A) les *contigs* correspondent à des séquences continues et les *scaffolds*¹ correspondent à l'assemblage de plusieurs *contigs*². B) Certains *scaffolds* ont pu être réunis entre eux grâce aux séquences des extrémités des BACs (d'après F. Legeai).

Tableau 3 : Statistiques de l'assemblage du génome de *S. frugiperda* version 3

Taille de l'assemblage	528,182,211
Nombre de <i>scaffolds</i> > 500bp	48,723
Taille moyenne	10,942
Taille N50	40,066
Nombre de <i>contigs</i> > 500bp	85,242
Taille moyenne	5,681
Taille N50	10,017

L'assemblage du génome a été perfectionné par Fabrice Legeai (UMR IRISA de Rennes) grâce notamment à l'utilisation de la banque de données BACs (*Bacterial artificial chromosome*) de *S. frugiperda* (d'Alençon *et al.* 2004). Les extrémités des BACs séquencés (73000 séquences de 750 pb) ont permis d'assembler les *scaffolds* entre eux (Figure 25) afin de créer des très grand *scaffolds*.

3.2. Annotation du génome et alignement

L'annotation du génome a été réalisée par le Génoscope. A partir du génome assemblé de *S. frugiperda* (version 3, statistiques sur Figure précédente) plusieurs étapes ont

¹ Un *scaffold* est une portion de séquence du génome, reconstruite à partir des séquences des extrémités des clones de banques génomiques représentatives du génome entier. Les *scaffolds* sont composés de contigs et de zones de séquence inconnue, « trous » ou « gaps » en anglais dont la taille est connue approximativement.

² Un contig est une séquence contiguë du génome dans laquelle l'ordre des bases est connu avec un niveau de confiance élevé.

été réalisées pour obtenir un génome annoté et pouvant être ensuite utilisé. Tout d'abord les séquences répétées du génome ont été « masquées » grâce à des logiciels comme *Tandem Repeat Finder* (TRF) ou encore *Repeat Masker* en choisissant le taxon *Insecta* dans la base de données d'éléments répétés Rebase <http://www.girinst.org/repbase/index.html> comme base de référence. Puis, une banque de séquences protéiques de lépidoptères extraites de la base de données protéique *Universal Protein Resource* (UniProt) ainsi qu'une banque de contigs d'ADN complémentaire (ADNc) de *S. frugiperda* ont été alignées sur l'assemblage génomique. Pour chaque protéine, seul les séquences ayant obtenu le meilleur score de correspondance *best match* (BM) ou dont le score est supérieur ou égal à 90% du score du BM ont été retenus. Enfin une prédiction des gènes a été réalisée grâce à la méthode *ab initio* entraînée sur les gènes de *Bombyx mori*. La méthode *ab initio* de prédiction des gènes correspond à une méthode de prédiction intrinsèque, qui ne repose que sur la séquence du gène elle-même et non sur des informations obtenues expérimentalement. L'ensemble de ces données collectées (alignement des protéines, des ADNc et les prédictions *ab initio* des gènes) ont alors prises en compte selon l'importance de chacune (ADNc > Protéines > *ab initio*) et ont permis de construire 24 447 modèles de gènes pour *S. frugiperda* (prédictions codant pour un CDS > 30nt, avec un score > 0).

Cependant ces informations sur les prédictions de gènes ne sont arrivées que tardivement au cours de la thèse et nous avons donc utilisé la méthode *ab initio* avec pour référence des gènes de différents organismes (diptère du genre *Drosophila*, coléoptère du genre *Tribolium* et lépidoptère *Bombyx mori*) pour l'annotation des *scaffolds* comprenant les régions flanquantes des microsatellites analysés. Nous avons alors utilisé l'algorithme de FGNSH de la base de données d'outils pour la génomique : softberry (www.softberry.com). FGNSH se base sur le modèle HMM (*Hidden Markov Model*) qui utilise des chaînes de Markov, pour établir la probabilité que la séquence analysée corresponde à un des « états » donnés caractérisés au préalable sur des sets de gènes connus pour un organisme donné. Les « états » pouvant correspondre à un « état » intron ou exon par exemple. Nous avons pris en compte les prédictions des gènes lorsqu'ils semblaient bien conservés chez les différents organismes utilisés pour paramétrer la prédiction. L'annotation des régions flanquantes des microsatellites a ensuite été réalisée en suivant un pipeline d'analyses bio-informatiques (Figure 26).

3.3. Comparaison des *scaffolds S. frugiperda* variant riz et maïs

Au vu des résultats préliminaires sur la divergence des deux variants obtenus au cours de la thèse, le séquençage du variant riz a été réalisé en plus du séquençage du génome de *S. frugiperda* variant maïs à partir d'une banque de type *paired ends* de 380 bp avec une couverture de >100 x (Genoscope). Ces séquences du variant riz ont été alignées sur la séquence de référence (variant maïs) avec le logiciel Mummer par Fabrice Legeai. En parallèle, un assemblage a été réalisé par deux méthodes (logiciels Minia et Soap, N50 3.1 kb pour les deux méthodes) par l'équipe de Claire Lemaître à Rennes. La taille estimée du génome haploïde est de 455 Mb Ceci nous a permis de comparer les séquences du variant maïs et du variant riz dans certaines régions génomiques d'intérêt (Figure 26).

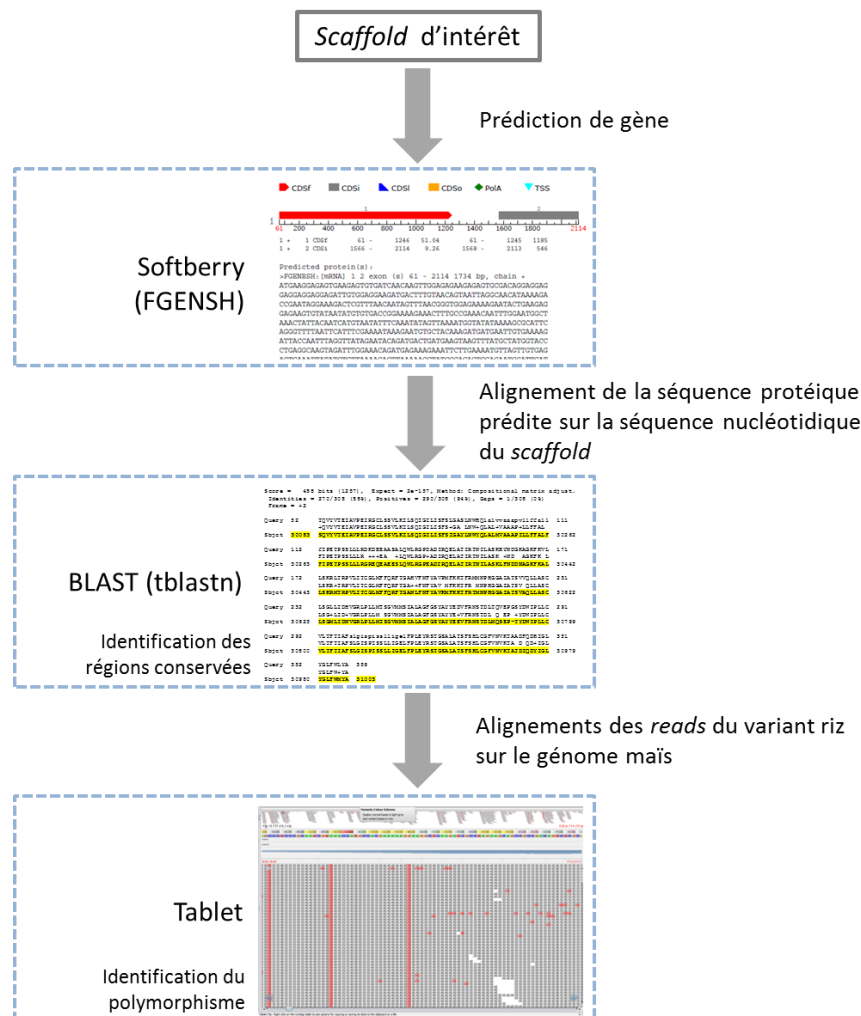


Figure 26 : Pipeline d'analyses bio-informatiques réalisées pour annoter les régions flanquantes des microsatellites étudiés et comparer les séquences du variant riz et du variant maïs de *S. frugiperda*.

Partie 3 : Principaux Résultats

PARTIE 3 : Principaux résultats

Dans cette partie sont présentés les principaux résultats obtenus au cours de la thèse en allant d'une échelle d'analyse large vers une échelle plus fine. L'ensemble de ces résultats visent à évaluer le niveau de différenciation génétique existant entre les deux variants de *Spodoptera frugiperda*.

1. Premier niveau d'analyse : études sur populations naturelles

Ce chapitre se concentre sur l'étude de la divergence des deux variants en s'appuyant sur l'analyse des populations naturelles de *S. frugiperda*.

Le premier paragraphe est consacré à l'article sur la phylogénie du genre *Spodoptera*, réparti mondialement et permet de situer la position phylogénétique de l'espèce *frugiperda* au sein du genre *Spodoptera*. Les différentes analyses de datation suggèrent que les deux variants de *S. frugiperda* auraient divergé entre 2.59 (intervalle de confiance à 95% de 1.87 - 3.61 Ma) et 3.74 (intervalle de confiance à 95% de 1.57 - 8.89 Ma) millions d'années. Les analyses phylogénétiques correspondantes mettent également en évidence une différenciation génétique marquée entre les deux variants, avec une distance génétique moyenne de 2.09% pour le COI (distance de Kimura à deux paramètres).

Le second paragraphe est dédié à l'étude de délimitation d'espèce afin de clarifier, à partir d'un échantillonnage beaucoup plus dense, le statut des deux variants de *S. frugiperda*. Cet article vérifie les hypothèses émises dans le paragraphe précédent sur le niveau élevé de différenciations génétiques entre les deux variants en identifiant deux clusters d'espèce potentiels correspondant aux deux variants de *S. frugiperda*. Cette étude utilise le modèle GMYC (*General Mixed Yule-Coalescent*) pour délimiter les espèces putatives et s'intéresse également à la sensibilité de cette approche aux choix des *a priori* en inférence Bayésienne.

Enfin le dernier paragraphe de ce chapitre s'intéresse à la structure de populations naturelles de *S. frugiperda* à partir d'un échantillonnage représentatif de leur aire de répartition. Ces analyses – bien que préliminaires - de génétique des populations et de

phylogéographie mettent en évidence un niveau de différenciation génétique élevé entre les populations du variant maïs et les populations du variant riz. En outre, au sein du variant riz on met en évidence un niveau de structuration significatif entre différentes populations séparées géographiquement.

1.1. Phylogénie et évolution du genre *Spodoptera*

Article n°1 publié dans *Molecular Phylogenetics and Evolution*. *Disentangling dispersal, vicariance and adaptive radiation patterns: A case study using armyworms in the pest genus Spodoptera (Lepidoptera: Noctuidae)*.



Disentangling dispersal, vicariance and adaptive radiation patterns: A case study using armyworms in the pest genus *Spodoptera* (Lepidoptera: Noctuidae)

Gael J. Kergoat^{a,*}, Dorothy P. Prowell^b, Bruno P. Le Ru^c, Andrew Mitchell^d, Pascaline Dumas^{a,e}, Anne-Laure Clamens^a, Fabien L. Condamine^f, Jean-François Silvain^g

^a INRA, UMR 1062 CBGP, IRD, CIRAD, Montpellier SupAgro, Campus de Baillarguet, 34988 Montferrier/Lez, France

^b Department of Entomology, Louisiana State University, Baton Rouge, LA 70803-1710, USA

^c Unité de Recherche IRD 072, ICIPE, PO Box 30772, Nairobi, Kenya

^d Research and Collections, The Australian Museum, 6 College Street, Sydney, NSW 2010, Australia

^e INRA, UMR 1333 DGIMI, UMII, Place Eugène Bataillon, Bat 24, 34095 Montpellier, France

^f Centre de Mathématiques Appliquées, UMR 7641 (CNRS, École Polytechnique), Route de Saclay, 91128 Palaiseau, France

^g Unité de Recherche IRD 072, Laboratoire Evolution, Génomes et Spéciation, UPR 9034, CNRS, 91198 Gif/Yvette, France and Université Paris-Sud 11, 91405 Orsay, France

ARTICLE INFO

Article history:

Received 16 March 2012

Revised 6 August 2012

Accepted 10 August 2012

Available online 23 August 2012

Keywords:

Bayesian relaxed clock

DEC analyses

Grasslands expansion

Long distance dispersal

Historical biogeography

Spodoptera

ABSTRACT

Thanks to the recent development of integrative approaches that combine dated phylogenies with models of biogeographic evolution, it is becoming more feasible to assess the roles of dispersal and vicariance in creating complex patterns of geographical distribution. However, the historical biogeography of taxa with good dispersal abilities, like birds or flying insects, still remains largely unknown because of the lack of complete phylogenies accompanied by robust estimates of divergence times. In this study, we investigate the evolution and historical biogeography of the globally distributed pest genus *Spodoptera* (Lepidoptera: Noctuidae) using complete taxon sampling and an extensive set of analyses. Through the analysis of a combined morphological and molecular dataset, we provide the first robust phylogenetic framework for this widespread and economically important group of moths. Historical biogeography approaches indicate that dispersal events have been the driving force in the biogeographic history of the group. One of the most interesting findings of this study is the probable occurrence of two symmetric long-distance dispersal events between the Afrotropical and the Neotropical region, which appear to have occurred in the late Miocene. Even more remarkably, our dated phylogenies reveal that the diversification of the clade that includes specialist grass feeders has followed closely the expansion of grasslands in the Miocene, similar to the adaptive radiation of specialist grazing mammals during the same period.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

In this study, we investigate the evolutionary history of the armyworm genus *Spodoptera* Guenée (Lepidoptera, Noctuoidea, Noctuidae, Amphipyriinae). The majority of *Spodoptera* species are commonly named armyworms due to the often-exhibited gregarious behavior of their larvae, which sometimes move in great numbers from one location to another when looking for food resources. About half the 30 known *Spodoptera* species are important/major pests of cultivated plants such as cotton, maize, rice, sorghum, sugarcane or tomato (Pogue, 2002). High fecundity and short life cycles under tropical climates explain the high rate of population increase and subsequent population outbreaks observed in many species (Harvey and Mallya, 1995; Day et al., 1996), which are the main causes of their pest status. Several species (especially

S. exigua, *S. frugiperda* and *S. littoralis*) have also been selected for use as biological model systems and are commonly used in research as diverse as the study of baculovirus interactions (Herrero et al., 2007), host-plant use (Barros et al., 2010), resistance to pesticides (Ahmad et al., 2007), or protein production (Sandhu et al., 2007). The fall armyworm, *S. frugiperda*, is becoming a model organism in genomic research (Nègre et al., 2006; d'Alençon et al., 2011). Complete genome sequencing of several *Spodoptera* species is under progress (P. Fournier pers. com). Finally, the importance of the genus *Spodoptera* is also illustrated by the result of a recent search on the PubMed database, which recovers nearly 8000 articles with the word *Spodoptera* in it. Despite a growing interest in the genus, little information is available on *Spodoptera* species relationships, and the status of several closely related species (*S. cosmiodes* and *S. descoinsi*; *S. marima* and *S. ornithogalli*) or subspecies (*S. mauritia mauritia* and *S. mauritia acronyctoides*) is still ambiguous (Brown and Dewhurst, 1975; Todd and Poole, 1980; Pogue, 2002). This is also the case for the fall armyworm,

* Corresponding author. Fax: +33 499 623 345.

E-mail address: kergoat@supagro.inra.fr (G.J. Kergoat).

S. frugiperda, which appears to be an assemblage of two closely related host-races referred to as the corn and rice strains (Pashley et al., 2004).

There are several fundamental questions about *Spodoptera* that can be addressed using a phylogenetic framework, such as unraveling the historical biogeography of the genus. *Spodoptera* species are distributed worldwide, mostly in tropical or subtropical areas, and are generally found in open-land habitats (Lalanne-Cassou et al., 1994; Monti et al., 1995). The genus presents a clearly disjunct distribution, as one can easily separate a distinct Western Hemisphere from an ‘Old World’ (i.e., the Afrotropical, Australasian, Oriental and Palearctic regions) fauna (Brown and Dewhurst, 1975; Todd and Poole, 1980; Pogue, 2002). There are no taxa historically shared by the two faunas although a single species of Old World origin, *S. exigua*, colonized the Western Hemisphere as a result of an accidental introduction in North America in 1876 (Wilson, 1932). Dispersal over long distances is facilitated by good flying abilities, and several species, such as *S. exempta*, *S. exigua*, *S. frugiperda* or *S. litura*, can migrate hundreds to thousands of kilometers in several weeks, flying at night in the downwind direction (Pair et al., 1986; Johnson, 1987; Gunn and Gatehouse, 1993; Gatehouse, 1997; Saito, 2001; Westbrook, 2008). The disjunct distribution of *Spodoptera* is interesting from an historical biogeography perspective, as it raises not only the question of the center of origin for the genus (either Old World or Western Hemisphere), but also of the timing and the nature (dispersal and/or vicariance) of the biogeographic events that have shaped the diversification of *Spodoptera* armyworms. Finally, a phylogenetic hypothesis for the genus will also allow us to study the evolution of host-plant associations in *Spodoptera* species, especially in the light of known ecosystem changes in the Cenozoic (Zachos et al., 2001). Although most species are polyphagous and able to feed on several plant families, a clear trend is discernable, as they generally are either specialized on monocots or dicots (Brown and Dewhurst, 1975; King and Saunders, 1984; Pogue, 2002). Interestingly, within the monocots feeders, larvae of species that feed mostly on grasses (Poaceae) and sedges (Cyperaceae) possess specialized mandibles with a chisel-like edge that is more adapted to the consumption of silica-rich leaves (Brown and Dewhurst, 1975; Pogue, 2002). Silica is well known for its adverse effects on lepidopteran larvae, and sometimes leads to complete abrasion of the mandibles (Djain and Pathak, 1967; Dravé and Laugé, 1978; Smith, 2005). In contrast to the grass and sedge feeders, the species that feed mostly on dicots or on silica-free monocots harbor serrate-like processes (Godfrey, 1987; Pogue, 2002). *Spodoptera frugiperda* is the only species known to feed regularly on grasses (maize, rice, sorghum and sugarcane) without possessing specialized chisel-like mandibles (Pogue, 2002). In several groups of phytophagous insects (Orthoptera: Acrididae, Lepidoptera: Crambidae, Lepidoptera: Noctuidae: Apameini), comparable adaptations have also occurred, such as the development of enlarged mandibles adapted to the consumption of silica-rich grasses (Zilli et al., 2005; Massey et al., 2006). If the mandibles with a serrate-like processes are derived in *Spodoptera*, we can hypothesize that their appearance coincided with biotic shifts from woodlands to open grasslands that accompanied the major global climate change in the Miocene (Zachos et al., 2001; Osborne, 2008).

To tackle both historical biogeography and evolution of host-plant associations of *Spodoptera*, we aim at reconstructing the densest species tree possible. Though it is desirable to obtain a complete molecular dataset for all species (with several specimens per species), such task is complicated by frequent non-availability of adequately preserved specimens for all species. To overcome this problem, we reconstruct the phylogenetic relationships of species using a combined approach, which associates a morphological dataset encompassing all species (Pogue, 2002) with a molecular

dataset for which data are missing for several species. In this study, we combine phylogenetic information from four mitochondrial genes and two nuclear genes, which are known for their phylogenetic usefulness in insects (Simon et al., 1994, 2006; Mitchell et al., 1997; Kim et al., 2000). The use of a combined mitochondrial and nuclear dataset is also aimed at maximizing information on species histories (Simon et al., 2006; Edwards, 2009) and overcoming possible bias such as mitochondrial introgression (Funk and Omland, 2003).

In summary, we intend to: (1) provide a first complete and robust species-level phylogenetic framework for the genus *Spodoptera* and, in so doing, clarify the status of several species, (2) investigate the historical biogeography of the genus to reconstruct their origin, infer their geographic range evolution in order to distinguish between dispersal and vicariance as the main process that shaped their current distribution pattern, (3) examine the evolution of host-plant associations and more specifically determine whether the diversification of grass and sedge feeders in *Spodoptera* can be connected with the spread of grasslands in the Miocene.

2. Materials and methods

2.1. Taxon sampling

2.1.1. *Spodoptera* sampling

Adult armyworms were collected from various locations by means of light or pheromone trapping, obtained from laboratory colonies, or reared from larvae (final stage larvae were also stored in ethanol for morphological studies). In addition, dried specimens were obtained from various collections (specimens used in this study are housed in the CBGP laboratory and in the Museum National d’Histoire Naturelle, Paris). Out of 30 known species, 27 species were sampled. Whenever possible, several specimens from distinct localities/countries were included for each species. A thorough sampling was obtained for the two possible members of the *frugiperda* species complex for which it had been possible to rear larvae from known hosts (either maize or rice). We also managed to obtain specimens from the two subspecies of *S. mauritia* (*S. m. mauritia* and *S. m. acronyctoides*), which can be differentiated by variation in forewing patterns. Distinguishing between the two is relevant in a historical biogeography context because the two subspecies have a disjunct area of distribution (*S. m. mauritia* is known from Madagascar, Mauritius and the Comoro Islands in the Afrotropical region whereas *S. m. acronyctoides* is found in the Oriental and Australasian regions; Brown and Dewhurst, 1975). Identification of specimens was mostly conducted by J.-F. Silvain and B. Lalanne-Cassou, who have an expertise on *Spodoptera* taxonomy (Lalanne-Cassou et al., 1994; Silvain and Lalanne-Cassou, 1997).

2.1.2. Outgroup sampling

Six distinct outgroup taxa were selected. To root the tree we chose to use a tiger moth species, *Hyphantria cunea* (Noctuoidea, Erebidae, Arctiinae). This choice was made because recent molecular phylogenetic analyses have indicated that the clade encompassing the families Erebidae and Nolidae is the sister clade of the Noctuidae (Mutanen et al., 2010; Zahiri et al., 2011). The comprehensive study of Mitchell et al. (2006) was used to determine five other outgroups belonging to the family Noctuidae. Three of them were chosen because they belong to subfamilies that are closely related to the subfamily Amphipyriinae (a Condidinae: *Condica* sp.; a Noctuinae: *Agrotis ipsilon*; and an Heliothinae: *Helicoverpa zea*.) whereas two other taxa were chosen as more distant outgroup taxa (an Acontiinae: *Amyna octo*; and a Cuculiinae: *Neogalea sunia*).

2.2. Molecular and combined datasets

2.2.1. Molecular datasets

Total DNA was extracted from hind legs and/or thoracic muscles using phenol–chloroform protocols or Qiagen DNeasy® tissue kits (Qiagen, Inc.), with 154 specimens being processed (Table 1). Despite multiple attempts (with distinct protocols), we were unable to recover suitable DNA templates for three species (*S. apertura*, *S. malagasy* and *S. pulchella*), for which only old dried specimens were available. This also was the case for six specimens of other *Spodoptera* species (mostly dried specimens). Polymerase chain reaction (PCR) amplifications were conducted for four mitochondrial (12S rRNA '12S', 16S rRNA '16S', cytochrome *b* 'Cyt *b*', and cytochrome *c* oxidase subunit I 'COI') and two nuclear (elongation factor-1 α 'EF-1 α ', and 28S rDNA '28S') genes, using standard settings (see Table A1 of the supplementary material for additional information). The resulting PCR products were processed by the French sequencing center Genoscope using a BigDye v3.1 sequencing kit and Applied 3730xl sequencers. Both strands were sequenced for all specimens to minimize PCR artefacts and ambiguities. Sequences of complementary strands were edited and reconciled using Geneious v3.8 software (Drummond et al., 2007). All sequences were deposited in GenBank (see Table 1 for accession numbers). Unlike the sequences of coding genes (Cyt *b*, COI and EF-1 α), the sequences of ribosomal genes (12S, 16S and 28S) contained some variations in length. Their alignment was accomplished using ClustalX (Thompson et al., 1997) with default option settings. For all protein-coding genes, we used Mesquite v2.75 (Maddison and Maddison, 2011) to check the coding frame for possible errors or stop codons. The resulting matrix ('141-taxa dataset') encompasses 135 individuals from 24 *Spodoptera* species plus the six outgroup taxa and 5080 aligned characters. In this study, this dataset is used to assess the phylogenetic relationships of all sampled specimens, and to investigate possible cases of paraphyly.

2.2.2. Reduced datasets

We chose to use species-level datasets to avoid biases in both dating and character optimization analyses. For instance, the inclusion of multiple specimens from the same species in an interspecific phylogeny is known to lead to a systematic overestimation of recent divergence times (Ho et al., 2005; Ho and Larson, 2006). Similarly, for each taxon of interest, it is preferable to summarize all the information on geographic distribution and host use rather than scattering the information among multiple individuals that do not summarize the ecology and geographic distribution of their species.

To build the reduced datasets, the original dataset ('141 taxa dataset') was modified under Mesquite to generate a molecular dataset with only one representative per *Spodoptera* species. Instead of randomly picking up a specimen per species, we used consensus sequences to have a better representation of species genetic variations (*consensus sequence strip* option, with default values). Specific consensus sequences were also reconstructed for the two geographic races of *S. mauritia* and for the two *S. frugiperda* strains in order to provide more insights on their evolutionary history. The corresponding dataset consists of 26 *Spodoptera* 'taxa' plus six outgroups ('reduced molecular dataset'). This dataset is used to implement the Bayesian relaxed clock (BRC) analyses, which generally rely on molecular-only matrices. A second 'reduced dataset' was then built to encompass a more comprehensive taxon sampling. This 'reduced combined dataset' was constructed by combining the 'reduced molecular dataset' with a morphology dataset corresponding to the matrix listed in Appendix I of Pogue (2002). This morphological dataset includes all *Spodoptera* species (plus two outgroups), and comprised 24 unordered characters based on adult

morphology. We used similar character states for the two host strains of *S. frugiperda* and for the two subspecies of *S. mauritia*. Due to the differences in the choice of outgroups used by Pogue (2002) and our study, question marks were used to code outgroups in the morphology dataset. The resulting 'reduced combined dataset' encompasses 32 *Spodoptera* 'taxa' plus six outgroups, and it is used to perform host-plant optimizations, dating analyses under penalized likelihood (PL) and historical biogeography analyses.

2.3. Phylogenetic analyses

2.3.1. Parsimony analyses of the '141-taxa dataset'

Parsimony analyses were carried out using the software TNT v1.1 (Goloboff et al., 2008). To assess congruence in phylogenetic signal from the mitochondrial and nuclear genes, we used the incongruence length difference test (ILD test; Farris et al., 1994), as implemented in PAUP* v4.0b10 (Swofford, 2003). Since the result of the test was not significant ($P = 0.65$), the whole dataset was directly analyzed through TNT. Heuristic searches were conducted using sectorial searches with Random Sectorial Searches and Consensus-based Sectorial Searches (Goloboff, 1999), with the algorithms of tree ratchet, tree drifting and tree fusing (Goloboff, 1999) selected ('new technology search'), 100 random-addition replicates and a MaxTrees value of 1000. Gaps were treated as a fifth character in all parsimony analyses. Support of trees was assessed using non-parametric bootstrap values (1000 replicates were used). Nodes supported by bootstrap values >70% were considered as strongly supported following Hillis and Bull (1993). For all taxa, intraspecific and interspecific distances (Kimura 2-parameter distance; Kimura, 1980) were assessed for the COI gene.

2.3.2. Bayesian inference analyses of the '141-taxa dataset'

We also analyzed the '141-taxa dataset' under Bayesian inference (BI), using MrBayes v3.12 (Ronquist and Huelsenbeck, 2003). To increase the fit of the models with the data, we conducted partitioned analyses, thus dividing our dataset into multiple subsets that possess their own specific sets of parameters and models of substitution (Nylander et al., 2004; Brown and Lemmon, 2007). Subsets of data were defined based on gene identity, codon positions, or origin (mitochondrial or nuclear). Best-fit models of evolution for each subset of the combined dataset were selected with jModelTest (Posada, 2008) using both the corrected Akaike information criterion (AICc) and the Bayesian information criterion (BIC). For this study, six competing partitioning strategies (summarized in Table 2) were compared using Bayes factors (B_F , Brown and Lemmon, 2007). For each partitioning strategy, two independent analyses were carried out to avoid being stuck in a local optimum (Huelsenbeck et al., 2002). Distinct analyses were also conducted whenever AICc and BIC suggested different best-fit models of evolution for a given data subset. Each analysis consisted of two independent runs with eight Markov Chains Monte Carlo (MCMC, one cold and seven incrementally heated) that ran for 10^7 generations. A conservative burn-in of 25% was applied after checking for stability on the log-likelihood curves and the split-frequencies of the runs. Estimates of harmonic mean of the likelihood values (obtained with the *sump* command in MrBayes) were further used to approximate the B_F , using twice the difference between the harmonic means as an estimate ($B_F = 2 \times |H_0 - H_1|$). We considered that B_F values > 10 significantly favor one model over another (Kass and Raftery, 1995). Support of nodes for BI analyses was provided by clade posterior probability (CPP) estimates. Clade posterior probabilities ≥ 0.95 were considered strong support (Erixon et al., 2003). Whenever species were recovered paraphyletic, we used B_F to test whether there was statistical support for their non-monophyly. To do so, specific analyses (in which selected species are constrained to be monophyletic) were carried

Table 1
Taxon sampling.

Genus species	Voucher no.	Genbank accession no.						Origin
		12SS	16S	Cyt b	COI	28S	Ef1a	
<i>Spodoptera</i>								
<i>albula</i> (Walker, 1857)	B20 ^a	HQ178019	HQ178248	HQ177614	HQ177285	HQ178554	PCR failure	French Guiana
<i>albula</i> (Walker, 1857)	B53 ^a	HQ178020	HQ178249	HQ177615	HQ177286	HQ178555	HQ177099	Honduras
<i>albula</i> (Walker, 1857)	LSU1 ^a	HQ178021	HQ178250	HQ177616	HQ177287	HQ178556	HQ177100	Guadeloupe
<i>albula</i> (Walker, 1857)	LSU16 ^a	HQ178022	HQ178251	HQ177617	HQ177288	HQ178557	HQ177101	French Guiana
<i>albula</i> (Walker, 1857)	LSU24 ^a	HQ178023	HQ178252	HQ177618	HQ177289	HQ178558	PCR failure	USA – Louisiana
<i>androgea</i> (Stoll, 1782)	B1 ^a	HQ178024	HQ178253	HQ177619	HQ177290	HQ178559	HQ177102	Guadeloupe
<i>androgea</i> (Stoll, 1782)	LSU26 ^a	HQ178025	HQ178254	HQ177620	HQ177291	HQ178560	HQ177103	French Guiana
<i>apertura</i> (Walker, 1865)	S5 ^b	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	Comoro Islands
<i>apertura</i> (Walker, 1865)	S6 ^b	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	Madagascar
<i>cilium</i> Guenée, 1852	GK123 ^a	HQ178026	HQ178255	HQ177621	HQ177292	PCR failure	HQ177104	Kenya
<i>cilium</i> Guenée, 1852	GK124 ^a	PCR failure	PCR failure	HQ177622	HQ177293	PCR failure	PCR failure	Kenya
<i>cilium</i> Guenée, 1852	GK125 ^a	PCR failure	PCR failure	HQ177623	HQ177294	PCR failure	PCR failure	Kenya
<i>cilium</i> Guenée, 1852	S7 ^b	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	Kenya
<i>cilium</i> Guenée, 1852	S8 ^b	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	Madagascar
<i>cilium</i> Guenée, 1852	S9 ^b	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	Madagascar
<i>cilium</i> Guenée, 1852	S10 ^b	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	Morocco
<i>cosmiodes</i> (Walker, 1858)	B19 ^a	HQ178027	HQ178256	HQ177624	HQ177295	HQ178561	HQ177105	French Guiana
<i>cosmiodes</i> (Walker, 1858)	B27 ^a	HQ178028	HQ178257	HQ177625	HQ177296	HQ178562	HQ177106	French Guiana
<i>cosmiodes</i> (Walker, 1858)	B100 ^a	HQ178029	HQ178258	HQ177626	HQ177297	HQ178563	HQ177107	French Guiana
<i>cosmiodes</i> (Walker, 1858)	B150 ^a	HQ178030	HQ178259	HQ177627	HQ177298	HQ178564	HQ177108	Brazil
<i>cosmiodes</i> (Walker, 1858)	B151 ^a	HQ178031	HQ178260	HQ177628	HQ177299	HQ178565	HQ177109	Brazil
<i>cosmiodes</i> (Walker, 1858)	B166 ^a	HQ178032	HQ178261	HQ177629	HQ177300	HQ178566	HQ177110	Brazil
<i>cosmiodes</i> (Walker, 1858)	LSU12 ^a	HQ178033	PCR failure	HQ177630	HQ177301	HQ178567	HQ177111	French Guiana
<i>descoinsi</i> L.C.&Silvain, 1994	B17 ^a	HQ178035	HQ178262	HQ177631	HQ177302	HQ178568	HQ177112	French Guiana
<i>descoinsi</i> L.C.&Silvain, 1994	B32 ^a	HQ178036	HQ178263	HQ177632	HQ177304	HQ178569	HQ177113	French Guiana
<i>descoinsi</i> L.C.&Silvain, 1994	B38 ^a	HQ178037	HQ178264	HQ177633	HQ177305	HQ178570	HQ177114	French Guiana
<i>descoinsi</i> L.C.&Silvain, 1994	B102 ^a	HQ178038	HQ178265	HQ177634	HQ177306	HQ178571	HQ177115	French Guiana
<i>descoinsi</i> L.C.&Silvain, 1994	B157 ^a	HQ178039	HQ178266	HQ177635	HQ177307	HQ178572	HQ177116	French Guiana
<i>descoinsi</i> L.C.&Silvain, 1994	LSU30 ^a	HQ178040	HQ178267	HQ177636	HQ177308	HQ178573	HQ177117	French Guiana
<i>descoinsi</i> L.C.&Silvain, 1994	LSU31 ^a	HQ178041	HQ178268	HQ177637	HQ177309	HQ178574	HQ177118	French Guiana
<i>depravata</i> (Butler, 1879)	GK206 ^a	HQ178034	PCR failure	PCR failure	HQ177302	PCR failure	PCR failure	Japan
<i>dolichos</i> (Fabricius, 1794)	B2 ^a	HQ178044	HQ178271	HQ177640	HQ177312	HQ178577	HQ177121	French Guiana
<i>dolichos</i> (Fabricius, 1794)	B35 ^a	HQ178045	HQ178272	HQ177641	HQ177313	HQ178578	HQ177122	French Guiana
<i>dolichos</i> (Fabricius, 1794)	B37 ^a	HQ178046	HQ178273	HQ177642	HQ177314	HQ178579	HQ177123	Guadeloupe
<i>dolichos</i> (Fabricius, 1794)	B104 ^a	HQ178047	HQ178274	HQ177643	HQ177315	HQ178580	HQ177124	Guadeloupe
<i>dolichos</i> (Fabricius, 1794)	B122 ^a	HQ178048	HQ178275	HQ177644	HQ177316	HQ178581	PCR failure	USA – Georgia
<i>dolichos</i> (Fabricius, 1794)	LSU3 ^a	HQ178049	HQ178276	HQ177645	HQ177317	HQ178582	PCR failure	Guadeloupe
<i>dolichos</i> (Fabricius, 1794)	LSU14 ^a	HQ178050	HQ178277	HQ177646	HQ177318	HQ178583	PCR failure	French Guiana
<i>dolichos</i> (Fabricius, 1794)	LSU27 ^a	HQ178051	HQ178278	HQ177647	HQ177319	HQ178584	HQ177125	USA – Florida
<i>dolichos</i> (Fabricius, 1794)	LSU29 ^a	HQ178052	HQ178279	HQ177648	HQ177320	HQ178585	PCR failure	French Guiana
<i>eridania</i> (Stoll, 1782)	B16 ^a	HQ178053	HQ178280	HQ177649	HQ177321	HQ178586	PCR failure	USA – Louisiana
<i>eridania</i> (Stoll, 1782)	B47 ^a	HQ178054	HQ178281	HQ177650	HQ177322	HQ178587	PCR failure	Guadeloupe
<i>eridania</i> (Stoll, 1782)	B70 ^a	HQ178055	HQ178282	PCR failure	HQ177323	HQ178588	PCR failure	USA – Louisiana
<i>eridania</i> (Stoll, 1782)	B71 ^a	HQ178056	HQ178283	HQ177651	HQ177324	HQ178589	HQ177126	USA – Louisiana
<i>eridania</i> (Stoll, 1782)	B72 ^a	HQ178057	HQ178284	HQ177652	HQ177325	HQ178590	PCR failure	USA – Louisiana
<i>eridania</i> (Stoll, 1782)	B73 ^a	HQ178058	HQ178285	HQ177653	HQ177326	HQ178591	HQ177127	USA – Louisiana
<i>eridania</i> (Stoll, 1782)	B75 ^a	HQ178059	PCR failure	HQ177654	HQ177327	HQ178592	HQ177128	Guadeloupe
<i>eridania</i> (Stoll, 1782)	B87 ^a	HQ178060	HQ178286	PCR failure	PCR failure	HQ178593	HQ177129	Peru
<i>eridania</i> (Stoll, 1782)	B88 ^a	HQ178061	HQ178287	PCR failure	PCR failure	HQ178594	HQ177130	Peru
<i>eridania</i> (Stoll, 1782)	B108 ^a	HQ178062	HQ178288	PCR failure	HQ177328	HQ178595	HQ177131	Peru
<i>eridania</i> (Stoll, 1782)	B109 ^a	HQ178063	HQ178289	HQ177655	HQ177329	HQ178596	HQ177132	Peru
<i>eridania</i> (Stoll, 1782)	LSU5 ^a	HQ178064	HQ178290	HQ177656	HQ177330	HQ178597	HQ177133	USA
<i>evanida</i> Schaus, 1914	B155 ^a	HQ178042	HQ178269	HQ177638	HQ177310	HQ178575	HQ177119	French Guiana
<i>evanida</i> Schaus, 1914	B156 ^a	HQ178043	HQ178270	HQ177639	HQ177311	HQ178576	HQ177120	French Guiana
<i>exempta</i> (Walker, 1857)	B24 ^a	HQ178065	HQ178291	HQ177657	HQ177331	HQ178598	HQ177134	Belgium (Laboratory)
<i>exempta</i> (Walker, 1857)	B25 ^a	HQ178066	HQ178292	HQ177658	HQ177332	HQ178599	PCR failure	Belgium (Laboratory)
<i>exempta</i> (Walker, 1857)	B93 ^a	HQ178067	HQ178293	HQ177659	HQ177333	HQ178600	HQ177135	Belgium (Laboratory)
<i>exempta</i> (Walker, 1857)	B162 ^a	HQ178068	HQ178294	HQ177660	HQ177334	HQ178601	PCR failure	Kenya
<i>exempta</i> (Walker, 1857)	B163 ^a	HQ178069	HQ178295	HQ177661	HQ177335	HQ178602	PCR failure	Kenya
<i>exigua</i> (Hübner, 1808)	B22 ^a	HQ178070	HQ178296	PCR failure	HQ177336	HQ178603	PCR failure	USA – California
<i>exigua</i> (Hübner, 1808)	B34 ^a	HQ178071	HQ178297	HQ177662	HQ177337	HQ178604	HQ177136	France (Laboratory)
<i>exigua</i> (Hübner, 1808)	B67 ^a	HQ178072	HQ178298	HQ177663	HQ177338	HQ178605	HQ177137	USA – California
<i>exigua</i> (Hübner, 1808)	B68 ^a	HQ178073	HQ178299	PCR failure	HQ177339	HQ178606	HQ177138	France (Laboratory)
<i>exigua</i> (Hübner, 1808)	B69 ^a	HQ178074	HQ178300	PCR failure	HQ177340	HQ178607	HQ177139	France (Laboratory)
<i>exigua</i> (Hübner, 1808)	G2178 ^a	PCR failure	PCR failure	PCR failure	HQ177343	HQ178610	PCR failure	Egypt
<i>exigua</i> (Hübner, 1808)	G2182 ^a	PCR failure	PCR failure	PCR failure	HQ177344	HQ178611	PCR failure	Egypt
<i>exigua</i> (Hübner, 1808)	G2185 ^a	PCR failure	PCR failure	PCR failure	HQ177345	HQ178612	PCR failure	Egypt
<i>exigua</i> (Hübner, 1808)	LSU11 ^a	HQ178075	HQ178301	HQ177664	HQ177341	HQ178608	PCR failure	USA
<i>exigua</i> (Hübner, 1808)	LSU15 ^a	HQ178076	HQ178302	HQ177665	HQ177342	HQ178609	HQ177140	USA – California
<i>frugiperda</i> (Smith, 1797)	B90C ^a	HQ178084	HQ178310	HQ177673	HQ177351	HQ178620	PCR failure	Peru (ex. maize)
<i>frugiperda</i> (Smith, 1797)	B91C ^a	HQ178085	HQ178311	PCR failure	HQ177352	HQ178621	PCR failure	Peru (ex. maize)
<i>frugiperda</i> (Smith, 1797)	C5.15.43B ^a	HQ178087	HQ178313	PCR failure	PCR failure	HQ178623	PCR failure	Guadeloupe (ex. maize)

Table 1 (continued)

Genus species	Voucher no.	Genbank accession no.						Origin
		12SS	16S	Cyt b	COI	28S	Ef1a	
<i>frugiperda</i> (Smith, 1797)	LSU21C ^a	HQ178086	HQ178312	HQ177674	HQ177353	HQ178622	HQ177144	Guadeloupe (ex. maize)
<i>frugiperda</i> (Smith, 1797)	B7R ^a	HQ178077	HQ178303	HQ177666	HQ177346	HQ178613	HQ177141	French Guiana (ex.rice)
<i>frugiperda</i> (Smith, 1797)	B8R ^a	HQ178078	HQ178304	HQ177667	HQ177347	HQ178614	PCR failure	Guadeloupe (ex. rice)
<i>frugiperda</i> (Smith, 1797)	B10R ^a	HQ178079	HQ178305	HQ177668	HQ177348	HQ178615	HQ177142	French Guiana (ex.rice)
<i>frugiperda</i> (Smith, 1797)	B11R ^a	HQ178080	HQ178306	HQ177669	PCR failure	HQ178616	PCR failure	French Guiana (ex.rice)
<i>frugiperda</i> (Smith, 1797)	B12R ^a	HQ178081	HQ178307	HQ177670	HQ177349	HQ178617	HQ177143	French Guiana (ex.rice)
<i>frugiperda</i> (Smith, 1797)	LSU22R ^a	HQ178082	HQ178308	HQ177671	HQ177350	HQ178618	PCR failure	Guadeloupe (ex. rice)
<i>frugiperda</i> (Smith, 1797)	R.5.20.16 ^a	HQ178083	HQ178309	HQ177672	PCR failure	HQ178619	PCR failure	Guadeloupe (ex. rice)
<i>latifascia</i> (Walker, 1856)	B3 ^a	HQ178088	HQ178314	HQ177675	HQ177354	HQ178624	HQ177145	USA – Louisiana
<i>latifascia</i> (Walker, 1856)	B18 ^a	HQ178089	HQ178315	HQ177676	HQ177355	HQ178625	HQ177146	Guadeloupe
<i>latifascia</i> (Walker, 1856)	B29 ^a	HQ178090	HQ178316	HQ177677	HQ177356	HQ178626	HQ177147	Guadeloupe
<i>latifascia</i> (Walker, 1856)	B31 ^a	HQ178091	HQ178317	HQ177678	HQ177357	HQ178627	HQ177148	Barbados
<i>latifascia</i> (Walker, 1856)	B33 ^a	HQ178092	HQ178318	HQ177679	HQ177358	HQ178628	HQ177149	USA – Louisiana
<i>latifascia</i> (Walker, 1856)	B36 ^a	HQ178093	HQ178319	HQ177680	HQ177359	HQ178629	HQ177150	Barbados
<i>latifascia</i> (Walker, 1856)	B39 ^a	HQ178094	HQ178320	HQ177681	HQ177360	HQ178630	HQ177151	USA – Louisiana
<i>latifascia</i> (Walker, 1856)	B101 ^a	HQ178095	HQ178321	PCR failure	PCR failure	HQ178631	PCR failure	Barbados
<i>latifascia</i> (Walker, 1856)	B123 ^a	HQ178096	PCR failure	HQ177682	HQ177361	HQ178632	PCR failure	USA – Georgia
<i>latifascia</i> (Walker, 1856)	LSU6 ^a	HQ178097	HQ178322	HQ177683	HQ177362	HQ178633	HQ177152	Guadeloupe
<i>latifascia</i> (Walker, 1856)	LSU17 ^a	PCR failure	HQ178323	HQ177684	HQ177363	HQ178634	HQ177153	USA
<i>littoralis</i> (Boisduval, 1833)	ADN0 ^a	PCR failure	PCR failure	PCR failure	HQ177371	HQ178642	PCR failure	Egypt
<i>littoralis</i> (Boisduval, 1833)	B14 ^a	HQ178098	HQ178324	HQ177685	HQ177364	HQ178635	HQ177154	France (laboratory)
<i>littoralis</i> (Boisduval, 1833)	B98 ^a	HQ178099	HQ178325	HQ177686	HQ177365	HQ178636	HQ177155	Egypt
<i>littoralis</i> (Boisduval, 1833)	B99 ^a	HQ178100	HQ178326	HQ177687	HQ177366	HQ178637	HQ177156	Egypt
<i>littoralis</i> (Boisduval, 1833)	B164 ^a	HQ178101	HQ178327	PCR failure	HQ177367	HQ178638	PCR failure	Congo
<i>littoralis</i> (Boisduval, 1833)	B165 ^a	HQ178102	HQ178328	HQ177688	HQ177368	HQ178639	PCR failure	Congo
<i>littoralis</i> (Boisduval, 1833)	GK118 ^a	HQ178105	HQ178331	HQ177691	HQ177372	HQ178643	HQ177159	Mali
<i>littoralis</i> (Boisduval, 1833)	GK119 ^a	HQ178106	HQ178332	HQ177692	HQ177373	HQ178644	PCR failure	Mali
<i>littoralis</i> (Boisduval, 1833)	GK120 ^a	HQ178107	HQ178333	HQ177693	HQ177374	HQ178645	HQ177160	Mali
<i>littoralis</i> (Boisduval, 1833)	LSU7 ^a	HQ178103	HQ178329	HQ177689	HQ177369	HQ178640	HQ177157	France (laboratory)
<i>littoralis</i> (Boisduval, 1833)	LSU35 ^a	HQ178104	HQ178330	HQ177690	HQ177370	HQ178641	HQ177158	France (laboratory)
<i>litura</i> (Fabricius, 1775)	B15 ^a	HQ178108	HQ178334	HQ177694	HQ177375	HQ178646	HQ177161	Indonesia
<i>litura</i> (Fabricius, 1775)	B113 ^a	HQ178109	HQ178335	HQ177695	HQ177376	HQ178647	HQ177162	Australia
<i>litura</i> (Fabricius, 1775)	B114 ^a	HQ178110	HQ178336	HQ177696	HQ177377	HQ178648	HQ177163	Australia
<i>litura</i> (Fabricius, 1775)	B153 ^a	HQ178111	HQ178337	PCR failure	HQ177378	HQ178649	PCR failure	Indonesia
<i>litura</i> (Fabricius, 1775)	B154 ^a	HQ178112	HQ178338	HQ177697	HQ177379	HQ178650	PCR failure	Philippines
<i>litura</i> (Fabricius, 1775)	B160 ^a	HQ178113	HQ178339	PCR failure	HQ177380	HQ178651	PCR failure	Indonesia
<i>litura</i> (Fabricius, 1775)	B161 ^a	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	Indonesia
<i>litura</i> (Fabricius, 1775)	GK126 ^a	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	New Caledonia
<i>malagasy</i> Viette, 1967	S1 ^b	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	Madagascar
<i>malagasy</i> Viette, 1967	S2 ^b	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	Madagascar
<i>malagasy</i> Viette, 1967	S3 ^b	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	Madagascar
<i>malagasy</i> Viette, 1967	S4 ^b	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	Madagascar
<i>marima</i> (Schaus, 1904)	B145 ^a	HQ178114	HQ178340	HQ177698	HQ177381	HQ178652	PCR failure	French Guiana
<i>marima</i> (Schaus, 1904)	LSU19 ^a	HQ178115	HQ178341	HQ177699	HQ177382	HQ178653	PCR failure	French Guiana
<i>mauritia</i> (Boisduval, 1833)	B65 ^a	HQ178116	HQ178342	PCR failure	HQ177382	HQ178654	HQ177165	Reunion
<i>mauritia</i> (Boisduval, 1833)	B66 ^a	HQ178117	HQ178343	HQ177700	HQ177383	HQ178655	HQ177166	Reunion
<i>mauritia</i> (Boisduval, 1833)	B94 ^a	HQ178118	HQ178344	HQ177701	HQ177384	HQ178656	PCR failure	Reunion
<i>mauritia</i> (Boisduval, 1833)	B95 ^a	HQ178119	HQ178345	PCR failure	HQ177385	HQ178657	PCR failure	Reunion
<i>m. subsp. acronyctoides</i>	B152 ^a	PCR failure	PCR failure	HQ177702	HQ177386	PCR failure	PCR failure	Papua New Guinea
<i>ochrea</i> (Hampson, 1909)	B84 ^a	HQ178120	HQ178346	HQ177703	HQ177387	HQ178658	HQ177167	Peru
<i>ochrea</i> (Hampson, 1909)	B85 ^a	HQ178121	HQ178347	HQ177704	HQ177388	HQ178659	HQ177168	Peru
<i>ochrea</i> (Hampson, 1909)	B86 ^a	HQ178122	HQ178348	HQ177705	HQ177389	HQ178660	HQ177169	Peru
<i>ochrea</i> (Hampson, 1909)	B110 ^a	HQ178123	HQ178349	HQ177706	HQ177390	HQ178661	HQ177170	Peru
<i>ornithogalli</i> Guenée, 1852	B4 ^a	HQ178124	HQ178350	HQ177707	HQ177391	HQ178662	HQ177171	USA – Louisiana
<i>ornithogalli</i> Guenée, 1852	B28 ^a	HQ178125	HQ178351	HQ177708	HQ177392	HQ178663	PCR failure	USA – Louisiana
<i>ornithogalli</i> Guenée, 1852	B54 ^a	HQ178126	HQ178352	HQ177709	HQ177393	HQ178664	HQ177171	USA – Louisiana
<i>ornithogalli</i> Guenée, 1852	B59 ^a	HQ178127	HQ178353	PCR failure	PCR failure	HQ178665	PCR failure	USA – Louisiana
<i>ornithogalli</i> Guenée, 1852	B106 ^a	HQ178128	HQ178354	HQ177710	HQ177394	HQ178666	HQ177172	USA
<i>ornithogalli</i> Guenée, 1852	B121 ^a	HQ178129	HQ178355	PCR failure	HQ177395	HQ178667	PCR failure	USA – Georgia
<i>ornithogalli</i> Guenée, 1852	B146 ^a	HQ178130	HQ178356	HQ177711	HQ177396	HQ178668	PCR failure	French Guiana
<i>ornithogalli</i> Guenée, 1852	LSU8 ^a	HQ178131	HQ178357	PCR failure	HQ177397	HQ178669	HQ177173	USA
<i>ornithogalli</i> Guenée, 1852	LSU34 ^a	HQ178132	HQ178358	HQ177712	HQ177398	HQ178670	PCR failure	USA – Louisiana
<i>pecten</i> Guenée, 1852	P1 ^a	HQ178133	PCR failure	HQ177713	HQ177399	HQ178671	HQ177174	Indonesia
<i>pecten</i> Guenée, 1852	P2 ^a	HQ178134	PCR failure	PCR failure	HQ177400	HQ178672	PCR failure	Indonesia
<i>pectinicornis</i> (Hampson, 1895)	B142 ^a	HQ178135	PCR failure	HQ177714	HQ177401	HQ178673	HQ177175	Australia
<i>pectinicornis</i> (Hampson, 1895)	B143 ^a	HQ178136	PCR failure	HQ177715	HQ177402	HQ178674	HQ177176	Australia
<i>pectinicornis</i> (Hampson, 1895)	B144 ^a	HQ178137	PCR failure	HQ177716	HQ177403	HQ178675	HQ177177	Australia
<i>picta</i> (Guérin-Méneville, 1838)	B158 ^a	HQ178138	PCR failure	HQ177717	HQ177404	HQ178676	HQ177178	Australia
<i>picta</i> (Guérin-Méneville, 1838)	B159 ^a	HQ178139	PCR failure	HQ177718	HQ177405	HQ178677	PCR failure	Australia
<i>praeifica</i> (Grote, 1875)	B5 ^a	HQ178140	HQ178359	HQ177719	HQ177406	HQ178678	HQ177179	USA – California
<i>praeifica</i> (Grote, 1875)	B30 ^a	HQ178141	HQ178360	HQ177720	HQ177407	HQ178679	HQ177180	USA – California
<i>praeifica</i> (Grote, 1875)	B105 ^a	HQ178142	HQ178361	HQ177721	HQ177408	HQ178680	PCR failure	USA – California
<i>praeifica</i> (Grote, 1875)	LSU32 ^a	HQ178143	PCR failure	HQ177722	HQ177409	HQ178681	HQ177181	USA – California

(continued on next page)

Table 1 (continued)

Genus species	Voucher no.	Genbank accession no.						Origin
		12SS	16S	Cyt b	COI	28S	Ef1a	
<i>pulchella</i> (H.-Shäffer, 1868)	S12 ^b	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	Haiti
<i>pulchella</i> (H.-Shäffer, 1868)	S13 ^b	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	PCR failure	Jamaica
<i>triturrata</i> (Walker, 1857)	GK121 ^a	HQ178144	HQ178362	HQ177724	HQ177411	HQ178683	HQ177183	Kenya
<i>triturrata</i> (Walker, 1857)	Tri ^a	PCR failure	PCR failure	HQ177723	HQ177410	HQ178682	HQ177182	Kenya
<i>Amyna octo</i> (Guenée, 1852)	B76 ^a	PCR failure	PCR failure	HQ177610	PCR failure	PCR failure	PCR failure	Guadeloupe
<i>Agrotis ipsilon</i> (Hufnagel, 1766)	B21 ^a	PCR failure	PCR failure	HQ177613	PCR failure	HQ178553	HQ177098	France
<i>Condica</i> sp.	LSU20 ^a	HQ178017	HQ178246	PCR failure	HQ177283	PCR failure	HQ177097	French Guiana
<i>Helicoverpa zea</i> (Boddie, 1850)	B51 ^a	HQ178018	HQ178247	HQ177612	HQ177284	PCR failure	PCR failure	France (Laboratory)
<i>Hyphantria cunea</i> (Drury, 1773)	39466	GU592049	GU592049	GU592049	GU592049	no data	U85671	Genbank data
<i>Neogalea sunia</i> (Guenée, 1852)	B44 ^a	PCR failure	HQ178245	HQ177611	HQ177282	PCR failure	PCR failure	Guadeloupe

^a CBGP laboratory.^b Museum National d'Histoire Naturelle, Paris.**Table 2**

Partitioning strategies used in this study.

Partitioning strategy ^a	Definition
P_A (1)	Unpartitioned data set
P_B (2)	1 partition for the mitochondrial genes/1 partition for the nuclear genes
P_C (4)	1 partition per codon position for the coding genes/1 partition for the noncoding genes
P_D (6)	1 partition per gene
P_E (7)	1 partition per codon position for the mitochondrial coding genes/1 partition per codon position for the nuclear coding gene/1 partition for the noncoding genes
P_F (9)	1 partition per codon position for the mitochondrial coding genes/1 partition per codon position for the nuclear coding gene/1 partition per gene for the noncoding genes

^a The number of partition is given into brackets.

out with MrBayes using the *constraint monophyly* command and the best partitioning strategy (as determined beforehand). The corresponding estimates of harmonic means were subsequently compared with those of the unconstrained analyses (using B_F) to determine whether they were statistically significantly less supported.

2.3.3. Analyses of the combined dataset

The 'reduced combined dataset' was analyzed under BI. The same subsets of data were used plus one for the morphological dataset. For the latter subset, we used a model of substitution analogous to the JC69 model (Jukes and Cantor, 1969) except that it has a variable number of states (one-parameter Markov k -state model; Mk1: Lewis, 2001), as implemented in MrBayes v3.1.2. The analytic pipeline that had been used on the '141-taxa dataset' was then implemented for the 'reduced combined dataset': (i) estimation of best-fit models of substitution for all subsets of data using both the AICc and the BIC; (ii) run of the competing partitioning strategies with MrBayes (same settings for the MCMC analyses); (iii) assessment of the best partitioning strategy through B_F comparisons.

2.4. Host-plant optimizations

Host-plant preferences for *Spodoptera* were compiled from the literature in a list of known host-plant families (Viette, 1967; Brown and Dewhurst, 1975; Todd and Poole, 1980; King and Saunders, 1984; Silvain and Thiberville, 1984; Holloway, 1989; Zhang, 1994; Pogue, 2002: Table A2 of the supplementary material). No host-plant information was available for the rare species *S. compta* (of which only three specimens are known; Pogue, 2002) and for three species host-plant data were partially speculative and likely incomplete (Table A2). Yet, for most species,

information from the literature allows us to determine whether taxa feed preferentially on monocots or dicots. It was difficult to categorize the feeding preference of the two *S. frugiperda* strains, which feed preferentially on cultivated grasses but are also able to develop on wild or cultivated dicots (Barros et al., 2010). Optimizations at the host-plant family level (or at a lower taxonomic level) were not performed because of the great number of multiple associations exhibited by numerous *Spodoptera* species. To map the evolution of host-plant associations in the genus, the following three character states were used (see Table 3): (0) dicots specialist; (1) monocots specialist; (2) ambiguous (to categorize the host-preferences of the *S. frugiperda* strains). Optimizations were conducted on the best-recovered phylogram as determined by B_F resulting from the analysis of the 'reduced combined dataset'. This option was favored because this dataset includes all *Spodoptera* species, and thus constitutes the best option to conduct character optimizations (Turner et al., 2009). This tree was further modified by removing outgroups and suppressing the branches leading to the terminal taxa with no host-plant data. Host-plant optimizations were carried out using the Mk1 model (Lewis, 2001), as implemented in Mesquite. We considered that the support of one state over another (at a given node) was significant if the difference between their log-likelihoods was greater than or equal to 2.0 (Schluter et al., 1997; Pagel, 1999).

2.5. Molecular dating

2.5.1. Fossil calibration

The fossil calibration is based on the two oldest unambiguous fossils of Noctuoidea, which allow setting a minimum age for the root node. A first fossil (specimen UWBM 66000) is a relatively complete specimen that is described in details in Douglas (1991). A discussion of its morphological traits supports its inclusion

Table 3
Categorization of host-plant preferences and distribution data used in this study.

Species	Main hosts	Native area of distribution
<i>S. albula</i>	Dicots	Nearctic, Neotropical
<i>S. androgea</i>	Dicots	Nearctic, Neotropical Specialist on dicots
<i>S. apertura</i>	Dicots	Afrotropical, Oriental, Sahul, Wallacea
<i>S. cilium</i>	Monocots	Afrotropical, Oriental, Palearctic (West), Wallacea
<i>S. compta</i>	Unknown	Neotropical
<i>S. cosmiodes</i>	Dicots	Neotropical
<i>S. depravata</i>	Monocots	Oriental, Palearctic (East)
<i>S. descoinsi</i>	Dicots	Neotropical
<i>S. dolichos</i>	Dicots	Nearctic, Neotropical
<i>S. eridania</i>	Dicots	Nearctic, Neotropical
<i>S. evanida</i>	Dicots	Neotropical
<i>S. exempta</i>	Monocots	Afrotropical, Oriental, Sahul, Wallacea
<i>S. exigua</i>	Dicots	Afrotropical, Oriental, Palearctic (West, East)
<i>S. frugiperda</i> (maize)	Ambiguous	Nearctic, Neotropical
<i>S. frugiperda</i> (rice)	Ambiguous	Nearctic, Neotropical
<i>S. latifascia</i>	Dicots	Nearctic
<i>S. littoralis</i>	Dicots	Afrotropical, Palearctic (West)
<i>S. litura</i>	Dicots	Oriental, Palearctic (East), Sahul, Wallacea
<i>S. malagasy</i>	Dicots	Afrotropical
<i>S. marima</i>	Dicots	Neotropical
<i>S. mauritia</i>	Monocots	Afrotropical
<i>S. m. acronyctoides</i>	Monocots	Oriental, Sahul, Wallacea
<i>S. ochrea</i>	Dicots	Neotropical
<i>S. ornithogalli</i>	Dicots	Nearctic, Neotropical
<i>S. pecten</i>	Monocots	Oriental, Wallacea
<i>S. pectinicornis</i>	Monocots	Oriental, Wallacea
<i>S. picta</i>	Monocots	Oriental, Sahul, Wallacea
<i>S. praefica</i>	Dicots	Nearctic
<i>S. pulchella</i>	Dicots	Nearctic
<i>S. roseae</i>	Dicots	Neotropical
<i>S. triturrata</i>	Monocots	Afrotropical
<i>S. umbraculata</i>	Monocots	Sahul

within the Arctiinae (Douglas and Stockey, 1996). This fossil was found embedded in tuffs from the Klondike Mountain formation (Washington State, Republic), which were radiometrically dated at approximately 48–49 Myr (Wolfe and Wehr, 1987). A second fossil (specimen UAPAL 459; Douglas and Stockey, 1996) is a hind-wing fragment that presents a combination of venation characters that is found in the trifold subfamilies of Noctuidae. This fossil was found embedded in gray shales from Quilchena Creek, whose ages have been correlated with potassium–argon estimates (48–50 Myr) from neighboring formations (Princeton area and McAbee localities). Because the assignment of a fossil to an extant lineage is often a tricky issue (Sauquet et al., 2012), we chose to remain conservative by considering these two fossils as unambiguous representatives of the noctuid superfamily rather than assigning them to a specific tribe or subfamily. Hence we set a minimum age of 48 Myr for the root node. This was made possible thanks to the inclusion of a tiger moth representative (*Hyphantria cunea*; Erebidae, Arctiinae) in all datasets.

2.5.2. Dating methods

In a preliminary way, the applicability of a molecular clock was investigated for all molecular datasets using PAUP*. Since the hypothesis of a molecular clock was not statistically supported by the corresponding likelihood ratio test ($P < 0.05$ for all molecular datasets), methods of dating that account for rate variation across lineages were used. Two subsequent sets of analyses were performed, using either BRC or PL approaches.

First, we have conducted BRC analyses on the ‘reduced molecular dataset’. In this study BRC analyses were carried out using the

BEAST v1.6.2 (Drummond and Rambaut, 2007). The constraint on the root age was enforced using either a uniform or an exponential distribution. The lower bound was set to 48 Myr while a maximum age of 200 Myr was enforced to optimize the search of the optimal root age by restraining the size of the parameters space to explore (see Sanderson, 2004 for more rationale on this issue). This upper bound is likely very conservative as it is more than four times the age of the oldest known noctuid fossil. Two distinct runs were carried out for each set of calibration (‘BRC uniform’ and ‘BRC exponential’), with one chain of 5×10^7 generations, a constant-rate Yule speciation process, default priors and trees sampled every 100 generations. After applying a conservative burn-in of 25% both the mean parameter estimates and the 95% highest posterior densities (95% HPD) were directly estimated using TreeAnnotator v1.6.2 (Drummond and Rambaut, 2007). Convergence of runs was assessed by examining the effective sample size (ESS) of parameters with Tracer v1.5 (Rambaut and Drummond, 2011).

The second set of analyses was designed to provide age estimates even for taxa for which molecular data were missing in the ‘reduced molecular dataset’. For doing so, we carried out PL analyses. In contrast with BRC approaches, PL only requires a given phylogenetic tree with its estimated branch lengths (hence no molecular matrix is needed for PL analyses). Therefore it was possible to use the phylogram resulting from the analysis of the ‘reduced combined dataset’, which includes all *Spodoptera* ‘taxa’. Because of the likely associated biases in branch length estimation (due to the lack of molecular data for several taxa), we chose to rely on secondary calibrations rather than only using the constraint on the root. To do so we constrained the nodes shared by both reduced datasets by using the minimum and maximum ages corresponding to the 95% HPD values from the BRC analyses. Two distinct sets of analyses were performed using the 95% HPD resulting from the two sets of BRC analyses (‘BRC uniform’ and ‘BRC exponential’). Penalized likelihood analyses were carried out with r8s v1.71 (Sanderson, 2003): in all analyses the default truncated Newton algorithm was used to better handle age constraints (Sanderson, 2004). Confidence intervals on parameters were estimated through the inference of additional phylograms with the same topology but different sets of branch lengths. To do so we conducted additional MrBayes analyses (same settings, best-fit partitioning strategy) on a species-level dataset. After randomly picking 1000 of the resulting trees, additional r8s analyses were carried out (one analysis per supplementary tree, same settings) to provide a median age and a confidence interval for each node (Lopez-Vaamonde et al., 2009).

2.6. Historical biogeography analyses

Information on the distribution of species was taken from several studies (Brown and Dewhurst, 1975; Todd and Poole, 1980; Pogue, 2002) and categorized into eight main biogeographic regions (Afrotropical, Nearctic, Neotropical, Oriental, West Palearctic, East Palearctic, Sahul and Wallacea; Table 3). Following the views of several authors (Hines, 2008; Mansion et al., 2008; Nylander et al., 2008), taxa with marginal distribution in an area were not assigned to it. We have also excluded records that correspond to the accidental introduction of *S. exigua* or *S. litura* in the Western Hemisphere (Pogue, 2002; Nagoshi et al., 2011).

To reconstruct the ancestral areas of distribution of *Spodoptera* species, ML inference of geographic range evolution was performed using the dispersal-extinction-cladogenesis (DEC) method (Ree et al., 2005; Ree and Smith, 2008). All DEC analyses were performed using the PL chronogram based on the BRC analyses with an exponential distribution. This chronogram was preferred over the chronogram based on the BRC analyses with a uniform distribution as it exhibits less age discrepancies. We used stratified

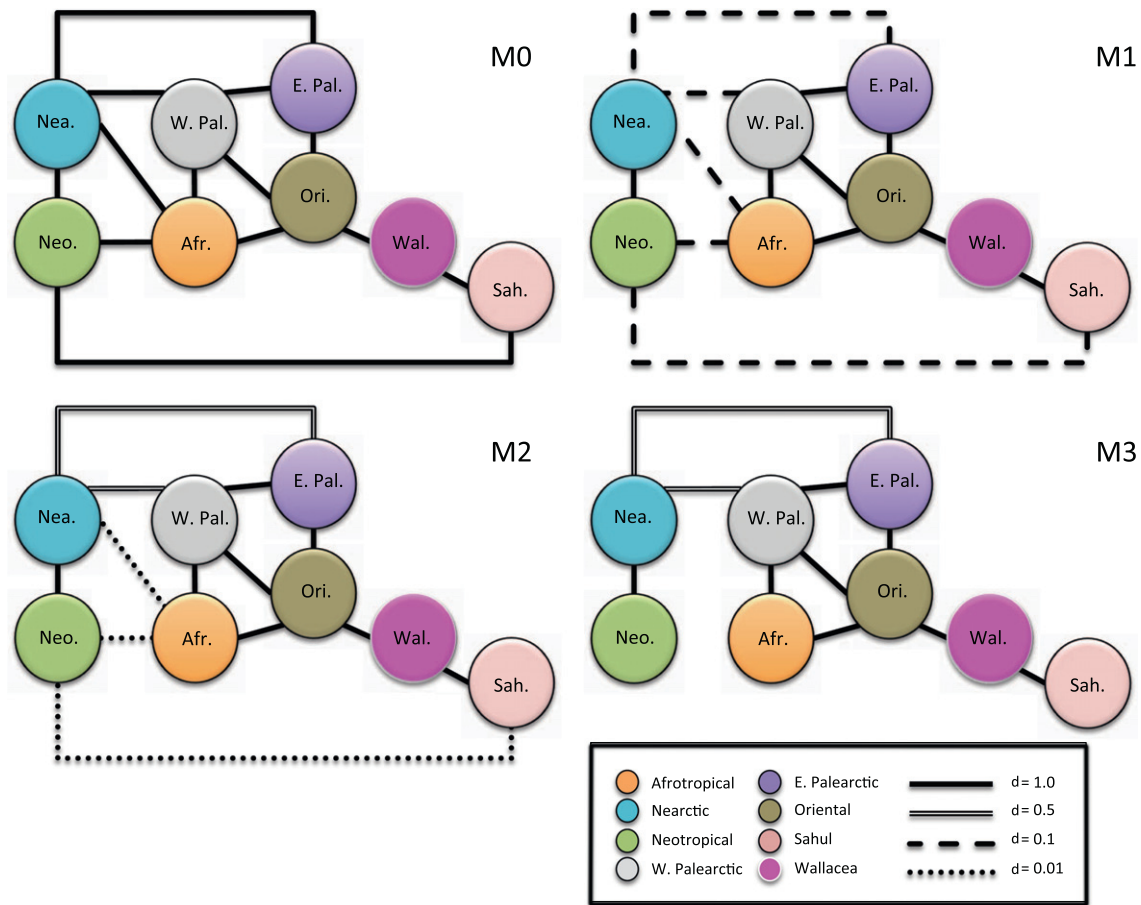


Fig. 1. Stratified DEC models. Representation of the four stratified DEC models used in this study. Two time slices are used, the sole difference being the absence of the Wallacea region in the first (40–20 Myr) time slice (the Oriental and Sahul regions are thus considered as non-adjacent during this period). Rates of dispersal are schematized using diverse categories of line as indicated in the figure's caption. For the model M0 all dispersals between the Western Hemisphere and the Old World are assigned a rate of 1.0 (continuous lines). For the model M1, all dispersals between the Western Hemisphere and the Old World are assigned a rate of 0.1 (dashed lines). The model M2 put higher penalties on transoceanic LDD (rate of 0.01, dotted lines) while favouring dispersal between the Nearctic and Palearctic regions (rate of 0.5, double lines). The model M3 prevents all transoceanic LDD and only authorizes dispersal between the Nearctic and Palearctic regions.

models with two time slices (0–20 Myr and 20–40 Myr) in order to account for the formation of Wallacea 20 Myr ago (Hall, 2002). To reduce the set of allowed arrangements (Ree and Smith, 2008), only combination of adjacent areas were taken into account for all models (e.g. Afrotropical + Oriental is valid whereas Afrotropical + E. Palearctic is not valid). The maximum number of ancestral areas was also set to four, which corresponds to the maximum number of areas observed for extant taxa (without considering human-mediated introductions). Four distinct stratified models (M0, M1, M2 and M3; see Fig. 1) were defined to estimate distribution ranges inherited by the descending *Spodoptera* lineages at each node of the trees. All models share the following characteristics: (i) the rate of dispersal between adjacent areas is set to 1.0; (ii) dispersal between distribution ranges that are separated by one intercalated area is considered as a LDD and is assigned a rate of dispersal of 0.1; (iii) dispersal between areas that are separated by two (or more) intercalated areas is assigned a rate of 0.0. These four models were designed to specifically investigate the possible pattern of LDD in armyworms, by progressively increasing the penalties associated with transoceanic LDD (Fig. 1). For instance, in the first model (M0) dispersals between the Old World and the Western Hemisphere are not penalized (rate of 1.0). In the model M1, all dispersals between the Old World and the Western Hemisphere are considered as LDD and assigned a rate of 0.1. In the model M2, transoceanic LDD between the Afrotropical (or Sahul) regions and the New World are more strongly penalized (rate of 0.01) while

dispersals between the Palearctic and the Nearctic regions (Trans-Atlantic and Trans-Beringian roads) are favored and assigned a rate of 0.5. Finally, the model M3 prevents all dispersal between the Afrotropical (or Sahul) regions and the New World (rate of 0.0), thus constraining the use of Trans-Atlantic and Trans-Beringian roads. Supplementary analyses were further conducted to determine the best-fit ancestral root area, using local optimizations conditional on the root state. These optimizations were conducted for all models and considering single and multiple area ranges. A 2-log likelihood unit's threshold was used to choose which area (or combination of areas) was better supported (Ree et al., 2005; Ree and Smith, 2008).

3. Results

3.1. Phylogenetic analyses

3.1.1. Parsimony analyses of the '141-taxa dataset'

Out of 5080 characters, 832 were parsimony-informative. The analysis of the '141-taxa dataset' yielded 42 most-parsimonious trees (3216 steps; CI = 0.463; RI = 0.851) that are mostly distinguishable by the position of several terminal individuals (Figure A1 of the supplementary material). Most of the non-intraspecific nodes are well supported by bootstrap values (bootstrap > 70% for 41 nodes; Fig. 2A). The genus *Spodoptera* is recovered monophyletic with a bootstrap support of 97%. Within *Spodoptera*, one

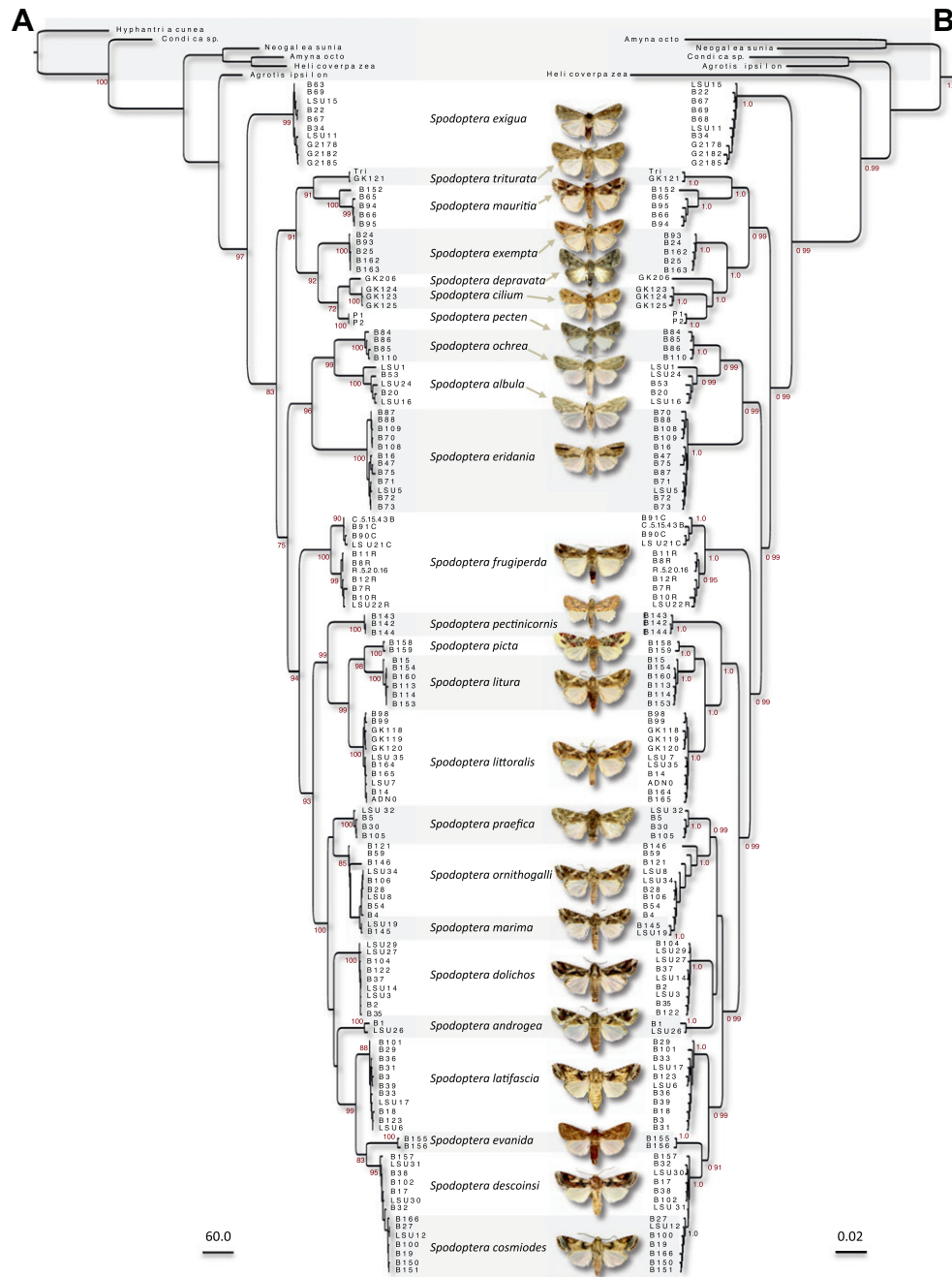


Fig. 2. Phylogenetic relationships of *Spodoptera* species ('141 taxa dataset'). Tree A is one of the 42 most-parsimonious trees (3216 steps; CI = 0.463; RI = 0.851) from parsimony analysis of the molecular data set. Numbers at nodes correspond to non-parametric bootstrap values $\geq 70\%$. Tree B results from partitioned Bayesian inference analyses conducted using the best partitioning strategy (P_F). Numbers at nodes indicate the CPP values ≥ 0.95 . Branch lengths are proportional in both trees. For all *Spodoptera* species, pictures of adult habitus are provided (pictures by B. Lalanne-Cassou).

species, *S. ornithogalli*, is consistently (all 42 most-parsimonious trees) found paraphyletic due to the placement of *S. marima* specimens. Likewise but to a lesser extent, *S. descoinsi* is paraphyletic in 20 out of the 42 most-parsimonious trees due to the placement of *S. cosmiodes* individuals. The level of mean sequence divergence (Kimura 2-parameter distance; COI gene) observed between individuals of *S. cosmiodes* and *S. descoinsi* is 0.46%, and between individuals of *S. marima* and *S. ornithogalli* is 0.38%. For two other species, *S. frugiperda* and *S. mauritia*, an interesting pattern is also recovered. Members of *S. frugiperda* are divided into two well-supported (90% and 99%) clades that are distinguishable by a mean sequence divergence of 2.09%. As expected, the two clades can be

explained by larval host-plant adaptation (maize or rice). Regarding *S. mauritia*, the sole individual of *S. m. acronyctoides* is found outside the cluster of *S. m. mauritia* individuals with a mean sequence divergence of 2.34%.

3.1.2. Bayesian inference analyses of the '141-taxa dataset'

The comparison of B_F values for the diverse partitioning strategies indicated that the best-fit strategy is the strategy P_F in combination with the selection of substitution models determined by the BIC ($H_0 = -22289.40$; see Table 4). The corresponding runs reached convergence as indicated by the log-likelihood curves and the split-frequencies of the runs (average split-frequencies

Table 4

Selection of substitution models and harmonic means of partitioning strategies. The best-fit partitioning strategy for each dataset is indicated using bold values.

Dataset	'141 Taxa dataset'		'All species dataset'	
	BIC	AICc	BIC	AICc
<i>Partition type</i>				
No partition	GTR+G	GTR+G	GTR+G	GTR+G
All mitochondrial genes	GTR+I+G	GTR+I+G	TrN+I+G	TrN+I+G
All nuclear genes	GTR+G	GTR+G	GTR+G	TrNef+G
All coding genes	GTR+I+G	GTR+I+G	TrN+I+G	TrN+I+G
All noncoding genes	GTR+G	GTR+G	GTR+G	GTR+G
Mitochondrial coding genes	GTR+I+G	GTR+I+G	TrN+I+G	TrN+I+G
Mitochondrial noncoding genes	TrN+I+G	TrN+I+G	TrN+I+G	TrN+I+G
12S rRNA	TrN+I+G	HKY+I+G	HKY+I+G	HKY+I+G
16S rRNA	TrN+I+G	TrN+I+G	GTR+G	GTR+I+G
CO1	GTR+I+G	GTR+I+G	TPM1uf+I+G	TPM1uf+I+G
Cyt b	TrN+I+G	TrN+I+G	TrN+I+G	TrN+I+G
EF1a	SYM+I+G	GTR+I+G	TrNef+I+G	GTR+I+G
28S rDNA (D2–D3)	HKY+G	TrN+G	HKY+G	TrN+G
Mitochondrial coding genes (1st pos.)	TrN+I+G	TrN+I+G	GTR+I+G	GTR+I+G
Mitochondrial coding genes (2nd pos.)	HKY+I+G	HKY+I+G	F81+I	HKY+I+G
Mitochondrial coding genes (3rd pos.)	TrN+I+G	TrN+I+G	TPM1uf+I+G	GTR+I+G
Nuclear coding gene (1st pos.)	TrN+I	F81+I	F81+I	TrN+I
Nuclear coding gene (2nd pos.)	F81	F81	F81	F81
Nuclear coding gene (3rd pos.)	GTR+G	GTR+G	GTR+G	GTR+G
<i>Harmonic means</i>				
Strategy P _A	–23926.17	–23926.17	–21913.00	–21913.00
Strategy P _B	–23314.29	–23314.29	–21188.97	–21188.97
Strategy P _C	–23898.39	–23898.39	–22002.87	–22002.87
Strategy P _D	–23149.85	–23150.79	–21076.68	–21056.44
Strategy P _E	–23340.48	–23136.54	–21156.07	–21072.30
Strategy P _F	–22289.40	–22502.80	–20518.52	–20464.52

values < 5%). In the resulting topology (Fig. 2B), the genus *Spodoptera* is recovered monophyletic with a CPP value of 0.99. Most of the non-intraspecific nodes are well supported by bootstrap values (CPP > 0.95 for 43 nodes; Fig. 2B). The branching order of all *Spodoptera* species is congruent with those resulting from the parsimony trees except for position of *S. depravata*. Under BI, *S. depravata* is not sister to *S. cilium* but to a clade grouping *S. cilium* and *S. pecten* instead. As in parsimony, two well-supported (CPP > 0.95) *S. frugiperda* clades are recovered, and *S. m. acronyctoides* is also found sister to the remaining members of *S. mauritia* (CPP of 1.0). The same two species, *S. descoinsi* and *S. ornithogalli*, are found paraphyletic due to the position of *S. cosmiodes* and *S. marima* specimens, respectively. To specifically address the paraphyly of *S. descoinsi* and *S. ornithogalli*, constrained trees (in which *S. cosmiodes* and *S. marima* were positioned as sister species) were built. The alternative hypothesis of monophyly for *S. ornithogalli* was rejected as indicated by the B_F value ($B_F = 2 \times |-22289.40 - 22312.05| = 45.3$) thus suggesting that the paraphyletic status of *S. ornithogalli* is statistically supported by our data. The alternative hypothesis of a monophyletic *S. descoinsi* was also rejected by a B_F value > 10 ($B_F = 2 \times |-22289.40 - 22283.88| = 11.04$).

3.1.3. Analyses of the combined dataset

The comparison of B_F values for the diverse partitioning strategies indicated that the best-fit strategy is the strategy P_F in combination with the selection of substitution models determined by the AICc ($H_0 = -20464.52$; see Table 4). The corresponding runs reached convergence as indicated by the log-likelihood curves and the split-frequencies of the runs (average split-frequencies values < 5%). In the resulting topology (Fig. 3), the genus *Spodoptera* is recovered as monophyletic with a CPP value of 0.99. The overall support for this topology is high, as 25 of the 37 nodes are supported by CPP values > 0.95. The branching order of the different species of *Spodoptera* is similar to those resulting from the analyses of the '141-taxa dataset' except for the position of *S. depravata*, which is found in a sister position with *S. pecten* (but with a CPP of 0.82).

3.2. Host-plant optimizations

The ML optimizations yield a clear pattern for ancestral character states of all but one node and are significantly supported by the comparison of likelihood scores (Fig. 3). A similar pattern is recovered even if the three species for which host-plant data are speculative are removed from the analyses. Character mapping reveals that specialization on dicots is likely the ancestral condition for the genus, and that specialization on monocots has evolved independently at least three times. On one occasion, this dietary shift was followed by the diversification of a clade of monocots specialists (encompassing eight taxa), which groups all the species with chisel-like edge mandibles together (type II, see Fig. 3).

3.3. Divergence time estimations

Overall both BRC and PL analyses provide similar estimates (Table 5). However, PL estimates based on the BRC analyses with an exponential distribution exhibit less age discrepancies with their BRC counterpart. Depending on the dating method the median age of the common ancestor of the genus *Spodoptera* ranges from 22 to 30.7 Myr. Age estimates ranging from 14.7 to 23.2 Myr were inferred for the clade of species that feed preferentially on grass and sedges.

3.4. Historical biogeography analyses

The comparison of the best likelihood scores recovered by the four DEC analyses reveals that the analysis with moderate constraints on LDD is statistically better supported (M1: $-\ln L = 97.82$) than analyses with less (M0: $-\ln L = 105.90$) or more constraints (M2: $-\ln L = 98.34$; M3: $-\ln L = 99.54$). All DEC analyses recover an origin in the Old World for *Spodoptera* spp. (Table A3). For all model but M3, the optimal ancestral area of distribution encompasses the Afrotropical and the Oriental regions. The DEC reconstructions of biogeographic history also suggest two major LDD events between

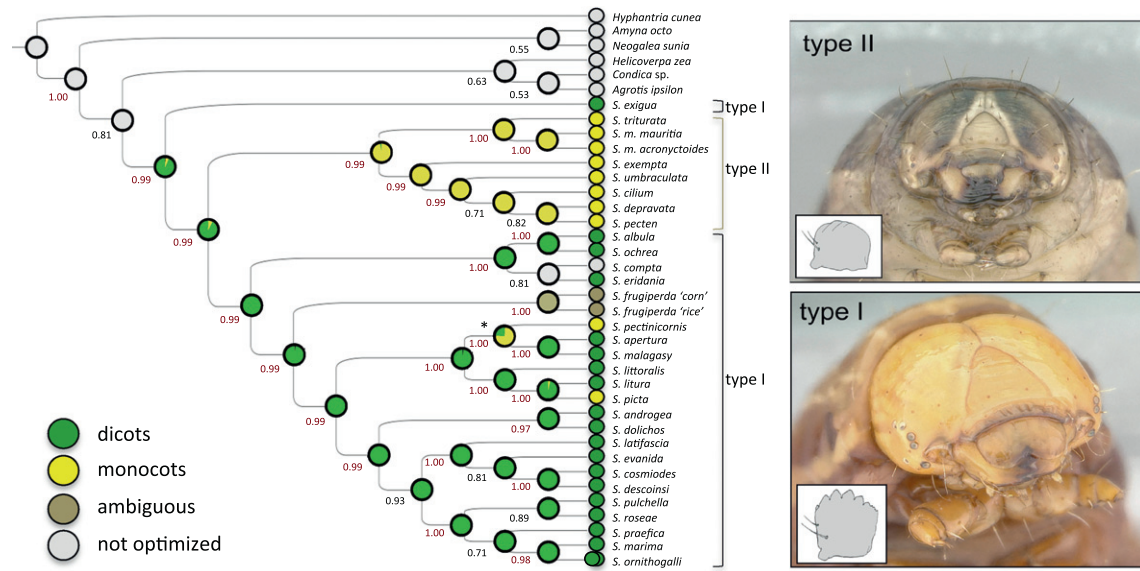


Fig. 3. Phylogenetics ('reduced combined dataset') and evolution of host-plant associations in *Spodoptera*. Tree resulting from the partitioned Bayesian inference analyses conducted using the best partitioning strategy (P_F). Numbers at nodes indicate the CPP values. On the same tree, the result of ML optimization of host-preferences is figured. Probabilities of character states are figured on nodes with pie charts. Only one node (marked with an asterisk) has non-significantly supported character states. Pictures of two final stage larvae are also provided to illustrate the differences between serrate-like (type I) and chisel-like edge mandibles (type II). On the top right is figured a larva of *S. depravata* and on the bottom right is figured a larva of *S. picta* (pictures by G.J. Kergoat). For more clarity, schematic views of the two types of mandibles (redrawn from Brown and Dewhurst, 1975) are also enclosed.

the Old World and Western Hemisphere (Fig. 4). For the first major dispersal event, the most likely scenario involves a LDD event from the Afrotropical region into the Neotropical region. Interestingly, the second major dispersal event appears to be a reverse colonization of the Old World (Afrotropical + Oriental regions according to the M1 model) from the Neotropical region. According to the optimal model M1, the Nearctic region was colonized by two distinct lineages between 6 and 10 Myr. In the Old World, all lineages (including the clade that originated from the Western Hemisphere) have progressively diversified eastward, and the Wallacea region was first reached between 12.8 and 14.8 Myr.

4. Discussion

4.1. Phylogenetic relationships and species status of *Spodoptera*

Results of the phylogenetic analyses are well supported and consistent across multiple inference methods (parsimony or Bayesian Inference). This strong phylogenetic signal is likely attributable to the good taxonomic sampling coverage and to the size and usefulness of the selected molecular dataset. The improved resolution obtained here allows us to discuss the results of a previous study based on morphological characters (Pogue, 2002). To a large extent, the phylogenetic hypotheses recovered in the two studies are in agreement, and *S. exigua* appears as the most basal species in both studies. This finding is consistent with the fact that this species exhibits several distinctive plesiomorphic morphological features (Pogue, 2002), especially with reference to wing maculation and male genitalia. *Spodoptera exigua* has the simplest wing maculation with a circular orbicular spot whereas in all other *Spodoptera* species there is a tendency towards a more or less elongated oval-shaped orbicular spot. Though the clades recovered by Pogue's study are comparable with those recovered in this study, their branching orders indicate discrepancies. The most noticeable difference between the two studies is the position of two clades. In Pogue's study the clade that groups *S. exempta*, *S. mauritia* and *S. triturrata* is in a more derived position whereas in our analyses it

is basal. The clade that groups *S. apertura*, *S. malagasy* and *S. pectinicornis* is in a more basal position in Pogue's analysis than in our combined analysis.

Our study adds further data on the status of several taxa that have received attention from systematists in recent years. Not surprisingly, our results confirm the existence of a marked level of genetic differentiation between the maize and the rice strain of *S. frugiperda* (Pashley et al., 2004; Nagoshi et al., 2011). The fact that both strains also exhibit limitations in interstrain mating in nature (Pashley and Martin, 1987; Nagoshi and Meagher, 2003; Groot et al., 2010) support the hypothesis that they may constitute separate species. Regarding *S. m. acronyctoides*, our study reveals a clear level of differentiation between the sampled individual of *S. m. acronyctoides* and other *S. mauritia* individuals. Additional samples are needed to resolve this preliminary finding, which is not clearly supported by morphological data (Brown and Dewhurst, 1975). Another significant result is the probable synonymy between *S. marima* and *S. ornithogalli*. These species are morphologically quite similar and can only be distinguished by differences in hindwing coloration and by the lack of sexual dimorphism in *S. marima* (Todd and Poole, 1980). The two species also occur in different geographic areas, as *S. marima* is only found in the Western and Eastern Neotropics whereas *S. ornithogalli* is found in the Northern Neotropics and in the South, Western and Eastern Nearctic (Pogue, 2002). In our phylogenetic analyses, *S. ornithogalli* is consistently (and with a high support) recovered paraphyletic with respect to *S. marima*. This result is also supported by the B_F value and by the low level of sequence divergence (0.38%) between representatives of *S. marima* and *S. ornithogalli*. Following the views of Todd and Poole (1980), these findings suggest that the status of *S. marima* should be reconsidered. Interestingly our study does not provide more support for the species status of *S. descoinsi*. Though nearly half of the parsimony trees recovered the reciprocal monophyly of *S. cosmiodes* and *S. descoinsi*, BI analyses consistently found *S. descoinsi* individuals in a derived position within a clade that encompasses *S. cosmiodes* specimens. The hypothesis of reciprocal monophyly for both taxa is also not supported by the corresponding B_F value. However, it seems difficult to put the two taxa into

Table 5
Age estimates (Myr) based on two distinct distributions for the root age (uniform or exponential). For each set a first analysis was carried out on the reduced molecular dataset using BRC. A second analysis was then performed on trees resulting from the analyses of the combined reduced dataset using PL (shared nodes – numbered a–f and 1–25 – were constrained using intervals based on the 95% HPD values recovered under BRC; additional nodes are labeled a1–a6). For each node the median age and 95% HPD are reported.

Node	Ancestor of	BRC (uniform)		PL (BRC uniform)		BRC (exponential)		PL BRC (exponential)	
		Median	95% HPD	Median	95% HPD	Median	95% HPD	Median	95% HPD
a	<i>Hyphantria cunea</i> – <i>S. ornithogalli</i>	65.52	48.00–140.72	77.17	48.00–100.77	58.39	48.00–102.02	58.74	48.00–74.62
b	<i>Neogalea sunia</i> – <i>Amyna octo</i>	35.83	16.69–79.75	54.41	38.27–74.35	31.74	15.56–59.20	42.23	30.42–53.14
c	<i>Neogalea sunia</i> – <i>S. ornithogalli</i>	45.20	22.83–98.40	48.58	35.94–69.68	40.06	23.19–73.19	37.79	28.88–48.16
d	<i>Helicoverpa zea</i> – <i>Condica</i> sp.	31.16	15.97–68.99	36.20	21.67–51.44	27.38	15.70–51.03	29.11	19.94–37.47
e	<i>Agrotis ipsilon</i> – <i>Condica</i> sp.	11.13	0.00–36.20	13.40	4.28–36.20	10.60	0.00–29.37	10.30	2.51–26.39
f	<i>Helicoverpa zea</i> – <i>S. ornithogalli</i>	37.56	19.03–82.52	42.22	24.92–56.17	33.22	19.07–60.68	32.52	24.67–42.61
1	<i>S. exigua</i> – <i>S. ornithogalli</i>	25.11	12.56–54.84	30.79	17.94–38.39	22.08	13.27–41.08	23.87	17.72–29.06
2	<i>S. triturrata</i> – <i>S. ornithogalli</i>	21.57	11.05–47.26	28.35	16.23–35.11	19.00	11.10–34.62	21.80	16.55–27.21
3	<i>S. triturrata</i> – <i>S. cilium</i>	16.84	8.46–36.84	23.21	13.97–29.93	14.72	8.42–27.03	18.11	12.87–22.02
4	<i>S. triturrata</i> – <i>S. m. mauritia</i>	12.61	6.30–28.28	14.68	8.61–18.93	11.10	6.17–21.04	11.25	7.81–14.23
5	<i>S. m. mauritia</i> – <i>S. m. acronyctoides</i>	4.62	1.82–10.41	3.69	2.43–5.22	4.08	1.80–8.20	2.81	1.89–3.94
6	<i>S. exempta</i> – <i>S. cilium</i>	11.22	5.20–24.98	19.46	12.42–24.98	9.84	5.15–18.38	14.87	10.99–18.38
7	<i>S. pecten</i> – <i>S. cilium</i>	8.18	3.72–18.52	11.95	6.25–15.94	7.20	3.58–13.78	9.08	5.93–12.69
8	<i>S. pecten</i> – <i>S. depravata</i>	7.09	3.15–16.35	9.37	4.70–12.72	6.26	2.97–12.39	7.14	4.46–10.08
9	<i>S. albula</i> – <i>S. ornithogalli</i>	19.60	10.41–43.18	27.02	5.40–12.92	17.20	10.12–31.43	20.48	15.32–24.84
10	<i>S. albula</i> – <i>S. eridania</i>	14.48	7.00–32.02	20.44	12.97–28.58	12.75	7.08–23.68	15.86	11.25–21.13
11	<i>S. albula</i> – <i>S. ochrea</i>	8.64	3.82–19.64	9.02	5.40–12.92	7.56	3.68–14.49	6.92	4.94–9.14
12	<i>S. frugiperda</i> 'rice' – <i>S. ornithogalli</i>	16.91	8.84–37.21	23.99	14.75–29.37	14.90	8.67–27.24	18.34	12.02–21.58
13	<i>S. frugiperda</i> 'rice' – <i>S. frugiperda</i> 'corn'	3.74	1.57–8.89	3.39	1.90–4.87	3.28	1.39–6.50	2.59	1.87–3.61
14	<i>S. pectinicornis</i> – <i>S. ornithogalli</i>	14.50	7.59–31.78	22.03	14.02–26.94	12.79	7.60–23.81	17.02	11.89–20.34
15	<i>S. pectinicornis</i> – <i>S. picta</i>	11.35	5.73–25.03	19.10	11.33–25.03	9.98	5.53–18.54	14.70	9.86–18.54
16	<i>S. littoralis</i> – <i>S. picta</i>	7.29	3.41–16.00	8.01	4.32–10.57	6.41	3.42–12.00	6.06	3.91–8.13
17	<i>S. litura</i> – <i>S. picta</i>	4.36	1.85–9.69	4.36	2.19–5.76	3.84	1.78–7.38	3.40	2.16–4.35
18	<i>S. androgea</i> – <i>S. ornithogalli</i>	9.77	4.82–21.25	15.37	8.72–21.25	8.55	4.91–16.15	12.05	6.92–15.96
19	<i>S. androgea</i> – <i>S. dolichos</i>	8.87	4.08–19.47	9.86	4.62–13.85	7.74	4.18–14.63	7.58	4.50–10.40
20	<i>S. latifascia</i> – <i>S. ornithogalli</i>	9.34	4.78–20.46	14.36	7.78–20.46	8.16	4.42–15.17	11.08	6.18–15.17
21	<i>S. latifascia</i> – <i>S. descoinsi</i>	6.36	3.26–14.38	6.72	3.74–9.47	5.56	3.11–10.44	5.11	3.11–6.83
22	<i>S. evanida</i> – <i>S. descoinsi</i>	4.90	2.37–11.12	4.95	2.55–6.44	4.27	2.21–7.96	3.79	2.40–4.78
23	<i>S. cosmiodes</i> – <i>S. descoinsi</i>	0.86	0.24–2.08	0.71	0.38–1.11	0.76	0.26–1.67	0.54	0.30–0.82
24	<i>S. praefica</i> – <i>S. ornithogalli</i>	7.45	3.54–16.87	9.22	6.46–13.70	6.54	3.31–12.32	7.12	5.28–10.22
25	<i>S. marima</i> – <i>S. ornithogalli</i>	0.77	0.20–2.00	0.96	0.38–1.75	0.67	0.16–1.55	0.73	0.31–1.33
a1	<i>S. umbraculata</i> – <i>S. pecten</i>	n/a	n/a	16.93	9.74–23.40	n/a	n/a	12.89	7.79–17.50
a2	<i>S. compta</i> – <i>S. eridania</i>	n/a	n/a	16.51	5.08–24.15	n/a	n/a	12.51	4.22–18.62
a3	<i>S. pectinicornis</i> – <i>S. apertura</i>	n/a	n/a	12.99	4.58–21.86	n/a	n/a	10.16	3.11–15.26
a4	<i>S. malagasy</i> – <i>S. apertura</i>	n/a	n/a	3.04	0.52–12.36	n/a	n/a	2.40	0.29–9.00
a5	<i>S. pulchella</i> – <i>S. ornithogalli</i>	n/a	n/a	13.10	6.19–20.08	n/a	n/a	10.00	4.96–14.89
a6	<i>S. pulchella</i> – <i>S. roseae</i>	n/a	n/a	8.15	1.48–15.71	n/a	n/a	6.01	0.89–11.68

synonymy because they exhibit marked differences in wing ornamentation and genitalia in addition to the differences in pheromones component ratio and calling period (Monti et al., 1995; Pogue, 2002).

4.2. Age and historical biogeography of *Spodoptera*

Molecular dating analyses suggest a middle Miocene origin for the genus *Spodoptera* (estimates ranging between 22 and 30.7 Myr depending on the analyses). These estimates are consistent with the fact that the genus *Spodoptera* belongs to a relatively derived group among the Noctuidae (Mitchell et al., 2006). The family has been suggested to have diversified during the mid-Cenozoic in response to diversification of many modern angiosperm groups (Mitchell et al., 2006). In other groups of Lepidoptera, such as geometrid moths (Yamamoto and Sota, 2007; Canfield et al., 2008) or butterflies (Wahlberg, 2006; Peña and Wahlberg, 2008; Condamine et al., 2012), diversification of lineages at the genus and species level mostly occurred during the late Cenozoic. From a biogeographic point of view, this dating indicates that diversification of *Spodoptera* took place well after major tectonic vicariance events associated with the Gondwanan continental breakup. Twenty-five million years ago, the world had already taken on a modern look, although Florida and parts of Eurasia were still underwater (Sanmartín et al., 2001; Sanmartín and Ronquist, 2004). Antarctica was covered by ice and the northern landmasses were cooling rapidly (Zachos et al., 2001).

This pattern thus clearly favors the role of dispersal over vicariance events to explain the present disjunct distribution of the *Spodoptera* fauna. This history limit the role of potential dispersal routes such as Trans-Atlantic and Trans-Beringian land bridges, because these bridges had disappeared before the middle Miocene or, if some existed, only cold-adapted organisms could use them (Sanmartín et al., 2001).

With reference to the question of the center of origin for the genus, the historical biogeography method used in this study explicitly favors the hypothesis of an Old World origin for the genus. According to the optimal model, the Afrotropical and Oriental regions are the most likely areas of origin for the genus. Though our most probable scenarios are only partially statistically supported (the threshold of 2 log-likelihood units is a very conservative measure; Pagel, 1999), we stress that our analyses rely on several sets of calibrations, a well-resolved phylogenetic hypothesis, and complete taxonomic coverage for the genus of interest. DEC biogeographic analyses support the hypothesis of a LDD between Afrotropical and Neotropical regions (Fig. 4), early in the history of *Spodoptera* diversification (between 20.4 and 23.8 Myr). In Lepidoptera, similar LDD events (between South America and Africa) have been recovered by biogeographic analyses of butterflies from the subfamily Nymphalinae (Wahlberg et al., 2005; Wahlberg, 2006). Transoceanic dispersal in *Spodoptera* can likely be accounted for by their excellent flying abilities (Johnson, 1987; Gunn and Gatehouse, 1993; Gatehouse, 1997), as they have been able to colonize almost all tropical islands or archipelagos

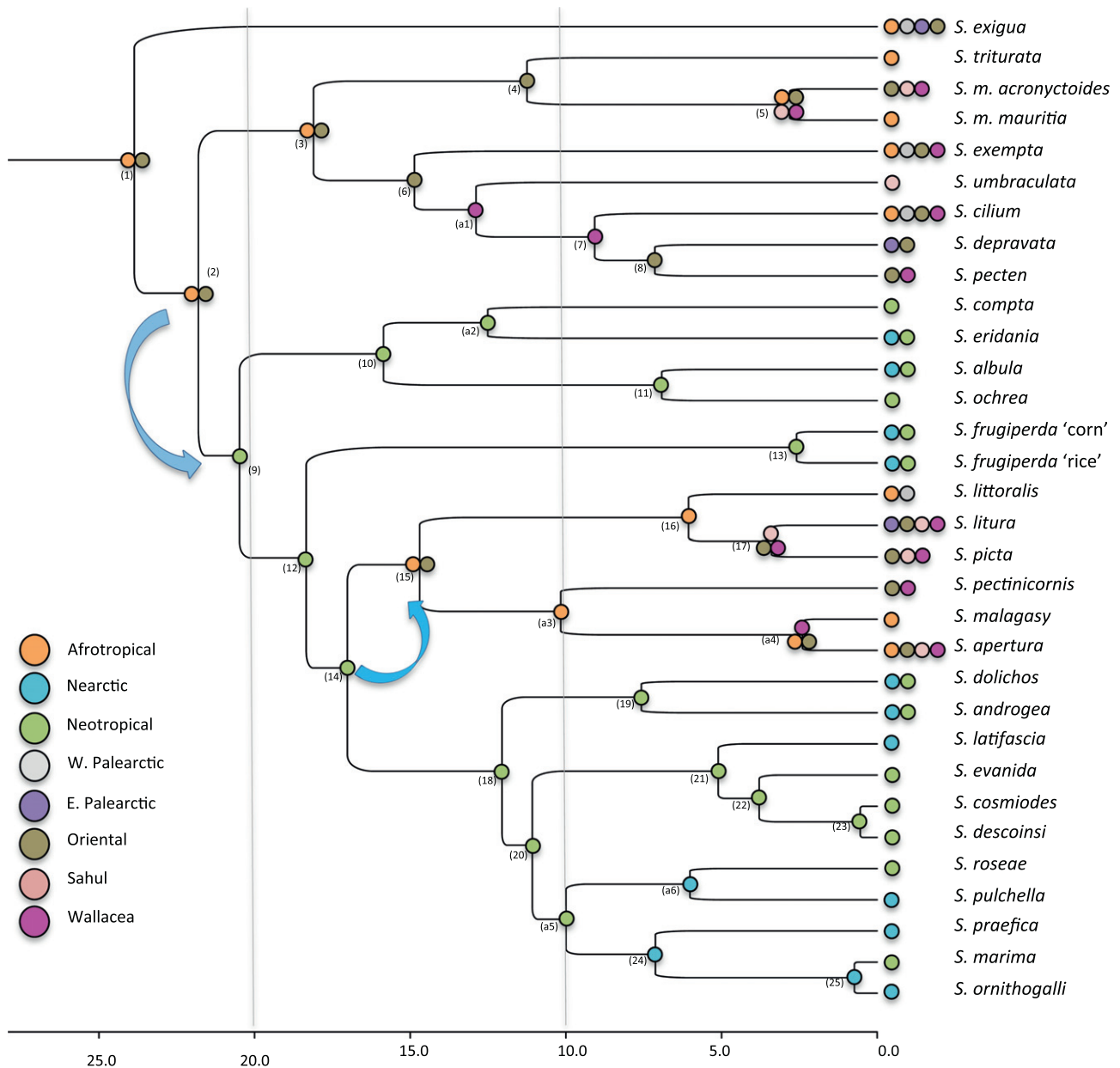


Fig. 4. Molecular dating and historical biogeography of *Spodoptera*. Chronogram corresponding to the result of the PL analysis based on the BRC analyses with an exponential distribution. The labels on nodes refer to Table 5. The result of the DEC analysis (M1 model) is also figured on the same tree. For each node optimal ML reconstructions of ancestral areas of distribution are figured. The two transoceanic long distance dispersal events are figured using large blue arrows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Pogue, 2002). It is also worth stressing that this result was even recovered with the M2 model, which put extremely high penalties to transoceanic LDD (dispersal rate of 0.01) while favoring dispersal between the Palearctic and the Nearctic regions (dispersal rates of 0.5). This first LDD event was further followed by a reverse colonization of the Old World by a Neotropical lineage, between 14 and 17 Myr. The latter constitutes one of the most interesting findings of this study because evidence for oceanic reverse-colonizations are scarce (Bellemain and Ricklefs, 2008), even when considering flying taxa. The ancestral area reconstructions carried out by the DEC method indicate that the distinct Old World lineages have progressively spread into the Wallacea and Sahul regions, a process that most likely has been favored by the fact that these regions were in close proximity in the mid-Miocene (Hall, 2002; Sanmartín and Ronquist, 2004). This pattern of coloni-

zation of the Sahul region corresponds to a theory referred as the 'northern dispersal hypothesis' (Braby et al., 2007). As indicated by our study, the two other alternative hypotheses ('southern vicariance hypothesis' and 'Indogondwanan hypothesis') classically proposed to explain the origin of Lepidoptera groups in Australia (see Braby et al., 2007) could be ruled out because Australia was no longer connected to South America through Antarctica at the time of *Spodoptera* diversification. In the Western Hemisphere, dispersal from the Neotropical to the Nearctic region occurred progressively, starting about 5 Myr before the formation of the Isthmus of Panama 3.5 Myr ago. According to DEC analyses, several independent colonizations of the Nearctic from the Neotropical region have occurred in the history of *Spodoptera*, in a way similar to the evolutionary history of *Nemoria* moths in the Western Hemisphere (Canfield et al., 2008). On the whole these reconstructions

indicate that diversification of *Spodoptera* species has followed a very dynamic pattern, with multiple episodes of independent colonizations and range expansions. As indicated by the biogeographic and molecular dating analyses, LDD have played an important role in this pattern, especially when considering the reverse colonization of the Afrotropical region. Following the view of Bellemain and Ricklefs (2008), we think that similar patterns are probably more frequent across a wide variety of taxa than has previously been realized (Clayton et al., 2009), and stress the importance of having more integrative approaches to historical biogeography studies in order to detect them (Gillespie et al., 2011).

4.3. Evolution and timing of host-plant associations

According to our age estimates, the genus *Spodoptera* started its diversification in the Miocene, when open habitats were dominant and filled with modern herbaceous angiosperm groups (Mitchell et al., 2006). This is in agreement with our results that suggest polyphagy on dicots is the likely ancestral feeding behavior in *Spodoptera*. This view is supported by the fact that most species (including the most basal species *S. exigua*) have mandibles with teeth-like processes, a character typical of herbaceous dicot feeding Lepidoptera (Bernays, 1991). The *S. frugiperda* complex appears somewhat anomalous unless considered in the context of human involvement with host plants. Although both strains have mandibles with teeth like processes and have been recorded and reared from a wide variety of wild and cultivated dicots, they are mostly known as important pests of several cultivated grasses (maize, rice, sorghum and sugarcane). The molecular dating analyses suggest that the two strains of *S. frugiperda* may have diverged more than 2 Myr ago. This means that their divergence took place well before the domestication or introduction of maize, rice, sorghum and sugarcane in the Western Hemisphere (Munkacsi et al., 2007). We can thus hypothesize that members of the *frugiperda* complex were primitively polyphagous (potentially on dicots) and that they only recently used cultivated grasses as host-plants, in relation to the extension of human agricultural activities, and particularly to spread of maize cultivation and the introduction of rice into the Western Hemisphere. Interestingly a similar pattern was found in the maize stalk borer (*Busseola fusca*) in Africa (Sezonlin et al., 2006). In this stemborer species three distinct populations have secondarily shifted to maize and sorghum long after their differentiation around 2.5 Myr (Sezonlin et al., 2006). Shifts in *S. frugiperda* may also have been facilitated by the fact that, in contrast to the Old World, there were few noctuid stemborers in the American fauna (Goldstein and Fibiger, 2005; Moyal, 2006). That means that the *frugiperda* maize strain whose last instars are acting as stemborers in maize likely filled an empty ecological niche.

During the diversification of the genus *Spodoptera*, the trend towards polyphagy on herbaceous dicots is relatively well conserved, with the preferred host-families being the species-rich Fabaceae, Malvaceae and Solanaceae. A similar tendency is found in several noctuid lineages, especially among the trifines (a group that encompasses the genus *Spodoptera*) that are predominantly herbaceous dicots feeders (Mitchell et al., 2006; Cho et al., 2008).

Our analyses suggest at least three independent specialization events on monocots have occurred in the evolutionary history of the genus. On two occurrences, these shifts towards monocots are highly specific: either on a single species (the water lettuce *Pistia stratiotes* in the case of *S. pectinicornis*) or on a single host-plant family (a few species of Liliaceae for *S. picta*). For a third lineage the shift is mostly directed toward grasses and sedges, which usually contain high contents of silica in their leaves. Species in this clade provide an example of the concept of key-innovation (sensu Simpson, 1953), as their mouthpart morphology has evolved in

response of the selective pressure of the abrasiveness of leaves, like other groups of herbivores (Massey et al., 2006). Simpson envisioned a key-innovation as an adaptation that permitted access to a new range of resources, resulting in a wave of subsequent speciation. In the case of *Spodoptera*, this key-innovation was indeed followed by a subsequent diversification of the corresponding lineage, but with only seven (or eight if considering *S. m. acronyctoides* as a distinct taxon) species, this radiation is not as spectacular as in other groups of grass feeding herbivores (Peña and Wahlberg, 2008; Micó et al., 2009; Toussaint et al., 2012). Species in this clade seem to have originated in the Afrotropical and Oriental region about 18 Myr ago. The divergence time estimates thus support the hypothesis that the radiation of *Spodoptera* grass feeders is connected to the transition to open grasslands in the late Miocene, in agreement with the views of Mitchell et al. (2006) and Toussaint et al. (2012) on the evolution of trifine noctuids. To summarize, our study suggests that the evolution of the genus was likely driven by a combination of biogeographic (range expansion, LDD, reverse colonization), and ecological (host-plant specialization) processes, which played an important role in the colonization of more open and arid habitats, in relation with late Cenozoic climate changes.

Acknowledgments

We thank S.L. Cameron, B. Riddle and two anonymous reviewers for very helpful comments on earlier drafts of the manuscript. We also thank the numerous persons who have helped providing samples for this study, and especially S. Moriya and C. Goto for the help with the Japanese specimens of *S. depravata*. We are indebted to B. Lalanne-Cassou for his help in identifying several individuals and for the *Spodoptera* pictures used in Fig. 2, and to G. Genson for technical assistance with the molecular work. Valuable feedbacks were also gained from insightful discussions with B. Lalanne-Cassou and M. Pagès. This study was financially supported by proper funds of INRA and IRD and a grant for G.J. Kergoat (project SPODOPTERA) from the Plant Health and Environment (SPE) division of the French National Institute for Agronomic Research (INRA). The contribution of A. Mitchell was supported by a NSW BioFirst Award. Part of the sequencing was supported by a Genoscope project @-Speed-Id (Accurate SPEciEs Delimitation and IDentification of eukaryotic biodiversity using DNA markers) proposed by F-BoL, the French Barcode of life initiative.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ympev.2012.08.006>.

References

- Ahmad, M., Sayyed, A.H., Crickmore, N., Saleem, M.A., 2007. Genetics and mechanism of resistance to deltamethrin in a field population of *Spodoptera litura* (Lepidoptera: Noctuidae). *Pest Manag. Sci.* 63, 1002–1010.
- Barros, E.M., Torres, J.B., Ruberson, J.R., Oliveira, M.D., 2010. Development of *Spodoptera frugiperda* on different hosts and damage to reproductive structures in cotton. *Entomol. Exp. Appl.* 137, 237–245.
- Bellemain, E., Ricklefs, R.E., 2008. Are islands the end of the colonization road? *Trends Ecol. Evol.* 23, 461–468.
- Bernays, E.A., 1991. Evolution of insect morphology in relation to plants. *Philos. Trans. Roy. Soc. Lond.* 333, 257–264.
- Braby, M.F., Pierce, N.E., Vila, R., 2007. Phylogeny and historical biogeography of the subtribe Aporiina (Lepidoptera: Pieridae): implications for the origin of Australian butterflies. *Biol. J. Linn. Soc.* 90, 413–440.
- Brown, E.S., Dewhurst, C.F., 1975. The genus *Spodoptera* (Lepidoptera: Noctuidae) in Africa and the Near East. *Bull. Entomol. Res.* 65, 221–262.
- Brown, J.M., Lemmon, A.R., 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Syst. Biol.* 56, 643–655.

- Canfield, M.R., Greene, E., Moreau, C.S., Chen, N., Pierce, N.E., 2008. Exploring phenotypic plasticity and biogeography in emerald moths: a phylogeny of the genus *Nemoria* (Lepidoptera: Geometridae). *Mol. Phylogenet. Evol.* 49, 477–487.
- Cho, S., Mitchell, A., Mitter, C., Regier, J., Matthews, M., Robertson, R., 2008. Molecular phylogenetics of heliothine moths (Lepidoptera: Noctuidae: Heliothinae), with comments on the evolution of host range and pest status. *Syst. Entomol.* 33, 581–594.
- Clayton, J.W., Soltis, P.S., Soltis, D.E., 2009. Recent long-distance dispersal overshadows ancient biogeographical patterns in a pantropical angiosperm family (Simaroubaceae, Sapindales). *Syst. Biol.* 58, 395–410.
- Condamine, F.L., Sperling, F.A.H., Wahlberg, N., Rasplus, J.-Y., Kergoat, G.J., 2012. What causes latitudinal gradients in species diversity? Evolutionary processes and ecological constraints on swallowtail biodiversity. *Ecol. Lett.* 15, 267–277.
- d'Alençon, E., Nègre, N., Stanojic, S., Allassoer, B., Gimenez, S., Léger, A., Abd-Alla, A., Juliant, S., Fournier, P., 2011. Characterization of a CENP-B homolog in the holocentric Lepidoptera *Spodoptera frugiperda*. *Gene* 485, 91–101.
- Day, R.K., Haggis, M.J., Odiyo, P.O., Mallya, G., Norton, G.A., Mumford, J.D., 1996. WormBase: a data management and information system for forecasting *Spodoptera exempta* (Lepidoptera: Noctuidae) in eastern Africa. *J. Econ. Entomol.* 89, 1–10.
- Djamine, A., Pathak, M.D., 1967. Role of silica in resistance to Asiatic borer (rice) *Chilo suppressalis* in rice varieties. *J. Econ. Entomol.* 60, 347–351.
- Douglas, S.D., 1991. Flower Form and Pollinator Diversity in the Middle Eocene of British Columbia and Washington. M. Sc. Thesis, University of Alberta, Edmonton.
- Douglas, S.D., Stockey, R.A., 1996. Insect fossils in middle Eocene deposits from British Columbia and Washington State: faunal diversity and geological range extensions. *Rev. Canad. Zool.* 74, 1140–1157.
- Dravé, E.-H., Laugé, G., 1978. Etude de l'action de la silice sur l'usure des mandibules de la pyrale du riz: *Chilo suppressalis* (F. Walker) (Lep. Pyralidae: Crambinae). *Bull. Soc. Entomol. Fr.* 83, 159–162.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214.
- Drummond, A.J., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Thierer, T., Wilson, A., 2007. Geneious v3.8. <<http://www.geneious.com/>>.
- Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? *Evolution* 63, 1–19.
- Erixon, P., Sennblad, B., Britton, T., Oxelman, B., 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst. Biol.* 52, 665–673.
- Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1994. Testing significance of incongruence. *Cladistics* 10, 315–319.
- Funk, D.J., Omland, K.E., 2003. Species-level paraphyly and polyphyly: frequency, causes, consequences, with insights from animal and mitochondrial DNA. *Annu. Rev. Ecol. Syst.* 34, 397–423.
- Gatehouse, A.G., 1997. Behavior and ecological genetics of wind-borne migration by insects. *Annu. Rev. Entomol.* 42, 475–502.
- Gillespie, R.G., Baldwin, B.G., Waters, J.M., Fraser, C.I., Nikula, R., Roderick, G., 2011. Long-distance dispersal: a framework for hypothesis testing. *Trends Ecol. Evol.* 27, 47–56.
- Godfrey, G.L., 1987. Noctuidae. In: Stehr, S.W. (Ed.), *Immature Insects*. Dubuque, Kendall/Hunt, pp. 549–578.
- Goldstein, P.Z., Fibiger, M.F., 2005. Biosystematics and evolution of the Apameini: a global synopsis. In: Zilli, A., Ronkay, L., Fibiger, M. (Eds.), *Noctuidae Europeae*, vol. 8. Apameini. Sora, Denmark, pp. 15–23.
- Goloboff, P.A., 1999. Analyzing large data sets in reasonable times: solutions for composite optima. *Cladistics* 15, 415–428.
- Goloboff, P.A., Farris, J.S., Nixon, K.C., 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24, 1–13.
- Groot, A.T., Marr, M., Heckel, D.G., Schöfl, G., 2010. The roles and interactions of reproductive isolation mechanisms in fall armyworm (Lepidoptera: Noctuidae) host strains. *Ecol. Entomol.* 35 (S1), 105–118.
- Gunn, A., Gatehouse, A.G., 1993. The migration syndrome in the African armyworm moth, *Spodoptera exempta*: allocation of resources to flight and reproduction. *Physiol. Entomol.* 14, 419–427.
- Hall, R., 2002. Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. *J. Asian Earth Sci.* 20, 353–431.
- Harvey, A.W., Mallya, G.A., 1995. Predicting the severity of *Spodoptera exempta* (Lepidoptera: Noctuidae) outbreak seasons in Tanzania. *Bull. Entomol. Res.* 85, 479–487.
- Herrero, S., Ansems, M., Van Oers, M.M., Vlak, J.M., Bakker, P.L., de Maagd, R.A., 2007. REPAT, a new family of proteins induced by bacterial toxins and baculovirus infection in *Spodoptera exigua*. *Ins. Biochem. Mol. Biol.* 37, 1109–1118.
- Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* 42, 182–192.
- Hines, H.M., 2008. Historical biogeography, divergence times, and diversification patterns of bumble bees (Hymenoptera: Apidae: *Bombus*). *Syst. Biol.* 57, 58–75.
- Ho, S.Y.W., Larson, G., 2006. Molecular clocks: when times are a-changin'. *Trends Genet.* 22, 79–83.
- Ho, S.Y.W., Phillips, M.J., Cooper, A., Drummond, A.J., 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol. Biol. Evol.* 22, 1561–1568.
- Holloway, J.D., 1989. The Moths of Borneo: Family Noctuidae, Trifine Subfamilies: Noctuinae, Heliothinae, Hadeninae, Acronictinae, Amphipyriinae, Agaristinae. Part 12. Southdene Sdn. Bhd., Kuala Lumpur.
- Huelsenbeck, J.P., Larget, B., Miller, R.E., Ronquist, F., 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.* 51, 673–688.
- Johnson, S.J., 1987. Migration and life history strategy of the fall armyworm, *Spodoptera frugiperda* in the Western Hemisphere. *Ins. Sci. Appl.* 8, 543–549.
- Jukes, T., Cantor, C., 1969. Evolution of protein molecules. In: Munro, H. (Ed.), *Mammalian Protein Metabolism*. Academic Press, New York, pp. 21–132.
- Kass, R.E., Raftery, A.E., 1995. Bayes factors. *J. Am. Stat. Assoc.* 90, 773–795.
- Kim, C.-G., Zhou, H.-Z., Imura, Y., Tominaga, O., Su, Z.-H., Osawa, S., 2000. Pattern of morphological diversification in the *Leptocarabus* ground beetles (Coleoptera: Carabidae) as deduced from mitochondrial ND5 gene and nuclear 28S rDNA sequences. *Mol. Biol. Evol.* 17, 137–145.
- Kimura, M., 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120.
- King, A.B.S., Saunders, J.L., 1984. *The Invertebrate Pests of Annual Food Crops in Central America*. ODA, London.
- Lalanne-Cassou, B., Silvain, J.-F., Monti, L., Malosse, C., 1994. Description d'une nouvelle espèce de *Spodoptera* de Guyane Française: *S. descoinsi* (Lepidoptera: Noctuidae: Amphirinae), découverte grâce à des attractifs sexuels. *Ann. Soc. Entomol. Fr. (NS)* 30, 25–32.
- Lewis, P.O., 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* 50, 913–925.
- Lopez-Vaamonde, C., Wikström, N., Kjer, K.M., Weiblen, G.D., Rasplus, J.-Y., Machado, C.A., Cook, J.M., 2009. Molecular dating and biogeography of fig-pollinating wasps. *Mol. Phylogenet. Evol.* 52, 715–726.
- Maddison, W.P., Maddison, D.R., 2011. Mesquite: A Modular System for Evolutionary Analysis. v2.75. <<http://mesquiteproject.org>>.
- Mansion, G., Rosenbaum, G., Schoenenberger, N., Bacchetta, G., Rosselló, J.A., Conti, E., 2008. Phylogenetic analysis informed by geological history supports multiple, sequential invasions of the Mediterranean basin by the angiosperm family Araceae. *Syst. Biol.* 57, 269–285.
- Massey, F.P., Ennos, A.R., Hartley, S.E., 2006. Silica in grasses as a defence against insect herbivores: contrasting effects on folivores and a phloem feeder. *J. Anim. Ecol.* 75, 595–603.
- Micó, E., Sanmartín, I., Galante, E., 2009. Mediterranean diversification of the grass-feeding *Anisopliina* beetles (Scarabaeidae, Rutelinae, Anomalini) as inferred by bootstrap-averaged dispersal-variance analysis. *J. Biogeogr.* 36, 546–560.
- Mitchell, A., Cho, S., Regier, J.C., Mitter, C., Poole, R.W., Matthews, M., 1997. Phylogenetic utility of elongation factor-1 α in Noctuoidea (Insecta: Lepidoptera): the limits of synonymous substitution. *Mol. Biol. Evol.* 14, 381–390.
- Mitchell, A., Mitter, C., Regier, J., 2006. Systematics and evolution of the cutworm moths (Lepidoptera: Noctuidae): evidence from two protein-coding nuclear genes. *Syst. Entomol.* 31, 21–46.
- Monti, L., Lalanne-Cassou, B., Lucas, P., Malosse, C., Silvain, J.-F., 1995. Differences in sex pheromone communication systems of closely related species: *Spodoptera latifascia* (Walker) and *S. descoinsi* Lalanne-Cassou and Silvain (Lepidoptera: Noctuidae). *J. Chem. Ecol.* 21, 641–660.
- Moyal, P., 2006. History of the systematics of the group of African Noctuid stem borers of monocotyledonous plants (Lepidoptera). *Ann. Soc. Entomol. Fr.* 42, 285–291.
- Munkacsí, A.B., Stoxen, S., May, G., 2007. Domestication of maize, sorghum and sugarcane did not drive the divergence of their smut pathogens. *Evolution* 61, 388–403.
- Mutanen, M., Wahlberg, N., Kaila, L., 2010. Comprehensive gene and taxon coverage elucidates radiation patterns in moths and butterflies. *Proc. Roy. Soc. B* 277, 2839–2848.
- Nagoshi, R.N., Meagher, R., 2003. Fall armyworm FR sequences map to sex chromosomes and their distribution in the wild indicate limitations in interstrain mating. *Ins. Mol. Biol.* 12, 453–458.
- Nagoshi, R.N., Brambila, J., Meagher, R., 2011. Use of DNA barcodes to identify invasive armyworm *Spodoptera* species in Florida. *J. Insect. Sci.* 11, 154.
- Nègre, V., Hôtelier, T., Volkoff, A.-N., Gimenez, S., Cousserans, F., Mita, K., Sabau, X., Rocher, J., López-Ferber, M., d'Alençon, E., Audant, P., Sabourault, C., Bidegainberry, V., Hilliou, F., Fournier, P., 2006. SPODBASE: an EST database for the lepidopteran crop pest *Spodoptera*. *BMC Bioinform.* 7, 322.
- Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. *Syst. Biol.* 53, 47–67.
- Nylander, J.A.A., Olsson, U., Alström, P., Sanmartín, I., 2008. Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to dispersal-variance analysis of the trushes (Aves: *Turdus*). *Syst. Biol.* 57, 257–268.
- Osborne, C.P., 2008. Atmosphere, ecology and evolution: what drove the Miocene expansion of *C₄* grasslands? *J. Ecol.* 96, 35–45.
- Pagel, M., 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst. Biol.* 48, 612–622.
- Pair, S.D., Raulston, J.R., Sparks, A.N., Westbrook, J.K., Dounce, G.K., 1986. Fall armyworm distribution and population dynamics in the southeastern states. *Florida Entomol.* 69, 468–487.
- Pashley, D.P., Martin, J.A., 1987. Reproductive incompatibility between host strains of fall armyworm (Lepidoptera: Noctuidae). *Ann. Entomol. Soc. Am.* 80, 731–733.
- Pashley, D.P., McMichael, M., Silvain, J.-F., 2004. Multilocus genetic analysis of host use, introgression, and speciation in host strains of fall armyworm (Lepidoptera: Noctuidae). *Ann. Entomol. Soc. Am.* 97, 1034–1044.
- Peña, C., Wahlberg, N., 2008. Prehistorical climate change increased diversification of a group of butterflies. *Biol. Lett.* 4, 274–278.

- Pogue, M.G., 2002. A world revision of the genus *Spodoptera* Guenée (Lepidoptera: Noctuidae). Mem. Am. Entomol. Soc. 43, 1–202.
- Posada, D., 2008. JModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.
- Rambaut, A., Drummond, A.J., 2011. Tracer v1.5. <<http://tree.bio.ed.ac.uk/software/tracer/>>.
- Ree, R.H., Smith, S.A., 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57, 4–14.
- Ree, R.H., Moore, B.R., Webb, C.O., Donoghue, M.J., 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Syst. Biol. 59, 2299–2311.
- Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
- Saito, O., 2001. Flight activity of three *Spodoptera* spp., *Spodoptera litura*, *S. exigua* and *S. depravata*, measured by a flight actograph. Physiol. Entomol. 2, 112–119.
- Sanderson, M.J., 2003. R8s; inferring absolute rates of evolution and divergence times in the absence of a molecular clock. Bioinformatics 19, 301–302.
- Sanderson, M.J., 2004. r8s, version 1.70. User's manual (December 2004). <<http://ginger.ucdavis.edu/r8s/>>.
- Sandhu, K.S., Naciri, M., Al-Rubeai, M., 2007. Prediction of recombinant protein production in an insect cell-baculovirus system using a flow cytometric technique. J. Immunol. Methods 325, 104–113.
- Sanmartín, I., Ronquist, F., 2004. Southern Hemisphere biogeography inferred by event-based models: plant versus animal patterns. Syst. Biol. 53, 216–243.
- Sanmartín, I., Enghoff, H., Ronquist, F., 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. Biol. J. Linn. Soc. 73, 345–390.
- Sauquet, H., Ho, S.Y.W., Gandolfo, M.A.A., Jordan, G.J., Wilf, P., Cantrill, D.J., Bayly, M.J., Bromham, L., Brown, G.K., Carpenter, R.J., Lee, D.M., Murphy, D.J., Sniderman, J.M., Udovicic, F., 2012. Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). Syst. Biol. 61, 289–313.
- Schluter, D., Price, T., Mooers, A.O., Ludwig, D., 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51, 1699–1711.
- Sezonlin, M., Dupas, S., Le Rü, B., Le Gall, P., Moyal, P., Calatayud, P.-A., Giffard, I., Faure, N., Silvain, J.-F., 2006. Phylogeography and population genetics of the maize stalk borer *Busseola fusca* (Lepidoptera, Noctuidae) in sub-Saharan Africa. Mol. Ecol. 15, 407–420.
- Silvain, J.-F., Lalanne-Cassou, B., 1997. Distinction entre *Spodoptera latifascia* (Walker) et *Spodoptera cosmiodes* (Walker), bona species (Lepidoptera, Noctuidae). Rev. Fr. Entomol. (NS) 19, 95–97.
- Silvain, J.-F., Thiberville, F., 1984. The noctuids (Lepidoptera: Noctuidae) dangerous to food and industrial crops in French Guiana. In: Proc. Caribbean Food Crops Soc. 19th. Annu. Meeting. Puerto Rico, pp. 217–239.
- Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87, 651–702.
- Simon, C., Buckley, T.R., Frati, F., Stewart, J.B., Beckenbach, A.T., 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annu. Rev. Ecol. Syst. 37, 545–579.
- Simpson, G.G., 1953. The Major Features of Evolution. Columbia University Press, New York, NY.
- Smith, C.M., 2005. Plant Resistance to Arthropods: Molecular and Conventional Approaches. Springer Science and Business.
- Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface. Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids Res. 25, 4876–4882.
- Todd, E.L., Poole, R.W., 1980. Keys and illustrations for the armyworm moths of the noctuid genus *Spodoptera* Guenée from the Western Hemisphere. Ann. Entomol. Soc. Am. 73, 722–738.
- Toussaint, E.F.A., Condamine, F.L., Kergoat, G.J., Silvain, J.-F., Capdevielle-Dulac, C., Barbut, J., Le Ru, B.P., 2012. Palaeoenvironmental shifts drove the adaptive radiation of a noctuid stemborer tribe (Lepidoptera, Noctuidae, Apameini) in the Miocene. PLoS One 7, e41377.
- Turner, A.H., Smith, N.D., Callery, J.A., 2009. Gauging the effect of sampling failure in biogeographical analysis. J. Biogeogr. 36, 612–625.
- Viette, P., 1967. Insectes. Lépidoptères Noctuidae. Faune de Madagascar, vol. 20, pp. 491–825.
- Wahlberg, N., 2006. The awkward age for butterflies: insights from the age of butterfly subfamily Nymphalinae (Lepidoptera: Nymphalidae). Syst. Biol. 55, 703–714.
- Wahlberg, N., Brower, A.V.Z., Nylin, S., 2005. Phylogenetic relationships and historical biogeography of tribes and genera in the subfamily Nymphalinae (Lepidoptera: Nymphalidae). Biol. J. Linn. Soc. 86, 227–251.
- Westbrook, J.K., 2008. Noctuid migration in Texas within the nocturnal aerocological boundary layer. Integr. Comp. Biol. 48, 99–106.
- Wilson, J.W., 1932. Notes on the biology of *Laphrygma exigua* Hübner. Florida Entomol. 16, 33–39.
- Wolfe, J.A., Wehr, W., 1987. Middle Eocene dicotyledonous plants from Republic, Northeastern Washington. US Geol. Surv. Bull. No. 1597.
- Yamamoto, S., Sota, T., 2007. Phylogeny of the Geometridae and the evolution of winter moths inferred from a simultaneous analysis of mitochondrial and nuclear genes. Mol. Phylogenet. Evol. 44, 711–723.
- Zachos, J.C., Pagani, M., Sloan, L., Thomas, E., Billups, K., 2001. Trends, rhythms and aberrations in global climate 65 Ma to Present. Science 292, 686–693.
- Zahiri, R., Kitching, I.J., Lafontaine, D., Mutanen, M., Kaila, L., Holloway, J.D., Wahlberg, N., 2011. A new molecular phylogeny offers hope for a stable family level classification of the Noctuoidea (Lepidoptera). Zool. Scripta. 40, 158–173.
- Zhang, B.C., 1994. Index of Economically Important Lepidoptera. CAB Cambridge University Press, UK.
- Zilli, A., Ronkay, L., Fibiger, M., 2005. Noctuidae Europaeae, vol. 8. Apameini. Entomological Press Sorø.

1.2. Analyse de délimitation d'espèce

Article 2 soumis à *BMC Evolutionary Biology: The effect of Bayesian branch length priors on species delimitation procedures; at tale from Spodoptera armyworms (Lepidoptera, Noctuidae).*

**The effect of Bayesian branch length priors on species delimitation
procedures: at tale from *Spodoptera* armyworms (Lepidoptera, Noctuidae)**

Pascaline Dumas^{1, 2, 3*}

*Corresponding author

Email: pascalinedumas@gmail.com

Jérôme Barbut⁴

Email: barbut@mnhn.fr

Bruno P Le Ru^{5, 6}

Email: bleru@icipe.org

Jean-François Silvain⁵

Email: Jean-Francois.Silvain@legs.cnrs-gif.fr

Anne-Laure Clamens³

Email: Anne-Laure.Clamens@supagro.inra.fr

Emmanuelle d'Alençon^{1, 2}

Email: alencon@supagro.inra.fr

Gael J Kergoat³

Email: kergoat@supagro.inra.fr

¹UM2 - UMR 1333 DGIMI, Université Montpellier 2, Place Eugène Bataillon, 34095, Montpellier, France

²INRA - UMR 1333 DGIMI, Université Montpellier 2, Place Eugène Bataillon, 34095, Montpellier, France

³INRA - UMR 1062 CBGP (INRA, IRD, CIRAD, Montpellier SupAgro), Campus de Baillarguet, CS 30016, 34988, Montferrier-sur-Lez, France

⁴Département Systématique et Evolution USM 602, Muséum National d'Histoire Naturelle, Paris, France

⁵Unité de Recherche IRD 072, Laboratoire Evolution, Génomes et Spéciation, UPR 9034, CNRS, 91198 - Gif-sur-Yvette, France and Université Paris-Sud 11, 91405, Orsay, France

⁶Unité de Recherche IRD 072, ICIPE, PO Box 30772, 00100, Nairobi, Kenya

Abstract

Background

One of the most used molecular species delimitation procedures, the General Mixed Yule-Coalescent (GMYC) model, relies on ultrametric trees to define potential species clusters. Similarly to other methods, this approach is known for being sensitive to sampling biases or confounding gene signals. However very little attention has been paid on the impact of tree building procedures on species delimitations. Interestingly, a recent study has highlighted that the choice and settings of ultrametrisation methods can drastically affect the number of putative species clusters in GMYC analyses. Here we propose to explore another potential source of error, namely the influence of Bayesian branch length priors and partitioning strategies on GMYC analyses. To do so we rely on a comprehensive dataset of noctuid moth belonging to the genus *Spodoptera*.

Results

In most cases, the number of resulting species clusters is stable, whatever partitioning strategies and priors are used for phylogenetic reconstructions. Most of the currently recognized *Spodoptera* species are also properly assigned. However, when species are not differentiated enough and/or correspond to species complexes with incomplete reproductive isolation - as it is the case for *Spodoptera frugiperda* strains - the number of GMYC species clusters is more likely to be affected by parameter choices.

Conclusions

Our study highlights that the GMYC approach is not overly sensitive to the partitioning strategies or branch length priors used in the tree building procedures. However, the problem comes from the fact that the GMYC procedure is more error prone for cases that matter most. Additional attention is thus required when using ultrametric trees based on the results of Bayesian analyses. Not only Bayesian tree reconstruction parameters have to be defined with caution, but it also seems necessary to combine a variety of approaches in order to detect potential biases in the results of species delimitation procedures.

Keywords

Branch lengths, Fall armyworm, General mixed Yule-coalescent model, Molecular species delimitation, *Spodoptera frugiperda*

Background

Though millions of species remain to be discovered and described [1-4], the science of describing species - Taxonomy - currently faces an important deficit in term of funding and manpower (the so-called “taxonomy crisis”; [5]). The resulting taxonomic impediment is a major issue because of the ongoing 6th extinction [6, 7], and we can rightfully fear that most of the yet-to-know species will vanish before even being found, described and potentially protected [8, 9]. The problem is also aggravated by the fact that traditional morphological studies are usually extremely time consuming, especially when dealing with hyperdiverse groups such as insects or plants, or when assessing species complexes with complex or cryptic morphological attributes [10-15]. The latter is often problematic, especially when facing species complexes that threaten public health [16, 17] or target important crops [18-21]. Some hope comes from the burst of molecular resources over the last decades, leading to the development of Molecular systematics as a whole (e.g. [22]). Molecular taxonomy, based on universal barcode sequences was then proposed as an elegant and effective solution to speed up species identifications and discoveries [23-31]. However, this initiative was quickly criticized on both practical [32-36] and theoretical [37-43] grounds, with numerous studies underlining the fact that molecular barcodes alone cannot solve the current taxonomic crisis [44-47]. The general consensus that emerged from these debates was that faster and more accurate species identifications or delimitations definitely require multiple complementary sources of information [20, 48, 49], hence underlining the need for a more “integrative” taxonomy [9, 50].

As defined by B. Dayrat [50], integrative taxonomy is « *the science that aims to delimit the units of life’s diversity from multiple and complementary perspectives (phylogeography, comparative morphology, population genetics, ecology, development, behaviour, etc.)*”. Because of their ever-increasing accessibility and decreasing cost, it is no wonder that molecular data are routinely incorporated into integrative taxonomy studies [49, 51-55]. This integration of molecular data also benefits from the fact that specific analytic methods were explicitly developed to assist species identifications and delimitations [56-59]. Among these methods one can generally distinguish between distance-based and tree-based methods [60-62]. Distance-based methods investigate the level of genetic divergence among individuals using standard gene fragments, such as the mitochondrial cytochrome c oxidase subunit I (COI) gene for animals [63-65]. Sequence divergences are supposed to be much lower among individuals of a same species than between individuals from different species [66]. The existence of this theoretical gap between intraspecific and interspecific divergence (“barcode gap”) is thus mandatory to guarantee successful species delimitation or identification by barcoding [67-69]. However this approach proved to be quite controversial [35, 40, 70, 71] when numerous empirical studies underlined that this barcoding gap is rarely found in empirical datasets [36, 41, 72]. In an alternative way, it was also proposed to identify or delimit species using tree-based methods [56, 59, 73, 74]. As implied by their name, these methods are more based on the ‘phylogenetic species concept’ (*sensu* [75]) and hence generally require trees with branch length proportional to time (ultrametric tree) as inputs. Most of these methods are also probabilistic and incorporate a coalescent framework [76-78], whose central aim is to identify independently evolving lineages with

no gene flow, in which selection and drift operate, each representing a putative species [79].

Among these tree-based methods, one approach - the general mixed Yule-coalescent (GMYC) model [70, 73] - has been particularly used in recent studies [80-88]. This approach simultaneously assesses the distinct branching patterns expected under a Yule model (modeling interspecific speciation events) and a coalescent model (modeling expected coalescent times of genes at an intraspecific level). It then determines the part of the tree for which there is a greater support for a Yule or a coalescent model, thereby drawing a virtual line between the interspecific and the intraspecific part of the tree. Potential species

clusters are then determined by identifying the clades (or single lineages) that originate after this putative threshold. Though the GMYC was later refined to account for rate heterogeneity [89] or topological uncertainty [90], it is surprisingly robust even when relying on single-locus information [91]. Similarly to other tree-based methods, the performance of the GMYC model is known to be extremely sensitive to several biases such as insufficient or unbalanced sampling [36, 92], heterogeneous rates of gene evolution [93], incomplete lineage sorting, or hybridization after introgression events [94]. Though rarely look upon, tree building procedures also matter, from the tree building procedures to the estimation of branch lengths and subsequent ultrametrization (Figure 1).

The GMYC model is a tree-based approach and it is the last step of a phylogenetic analysis pipeline. At each step of this analysis pipeline different factors or parameters could affect results and have final repercussions on

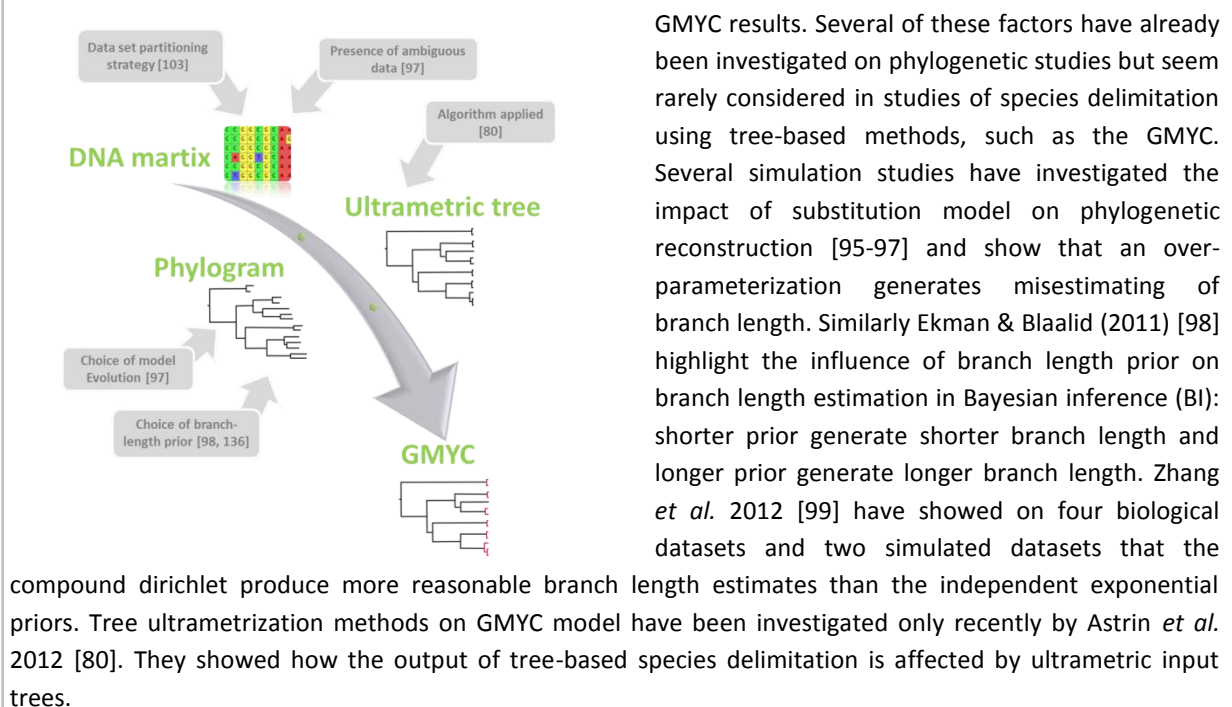


Figure1: analysis pipeline to use GMYC model

For example, a recent empirical study [80] has highlighted that results of GMYC procedures are sensitive to the choice of the ultrametrization method (nonparametric rate smoothing vs. penalized likelihood versus mean path length). In their study they found out that GMYC species delimitations relying on mean path length ultrametrization consistently outperformed analyses relying on penalized likelihood or nonparametric rate smoothing. As highlighted beforehand, another potential source of discrepancy is the estimation of branch lengths, prior to the ultrametrization procedure. Branch length estimates are known to be sensitive to the choice of substitution models [95, 96, 100] or partitioning strategies [101, 102]. It is especially the case in phylogenetic analyses that use multiple partitions and/or complex models, for which over-parameterization can potentially bias branch length estimates [101, 103]. In addition, it has been recently demonstrated [98] that branch length priors (within a Bayesian framework) can significantly impact branch length estimation, with shorter priors generating shorter branch lengths and longer priors generating longer branch lengths. To summarize, one can expect that any factor influencing the branch length estimation step can also impact the results of GMYC procedures, by generating either artefactually high or low number of putative species clusters.

To investigate the influence of partitioning or Bayesian branch length priors on species delimitation procedures, we propose to apply the GMYC method to an empirical dataset of moths belonging to the genus *Spodoptera* (Lepidoptera: Noctuidae). This genus encompasses about 30 species [104], and includes several species whose status has been questioned in the past based on biological [105-107], morphological [108, 109] and molecular evidence [110-113]. Clarifying

Spodoptera species boundaries is of prime importance, especially given the fact that half of the known species are attacking cultivated plants such as beans, cotton, maize, millet, rice, sorghum, soybean or tomatoes [104, 114, 115]. Here a focus will be made on a particular species, the fall armyworm (FAW; *Spodoptera frugiperda*), whose identity is particularly debated [116-118]. The fall armyworm is a widespread and well-known agricultural pest in the Western hemisphere [104, 114]. It is an extremely polyphagous species, which is known to develop on about 100 different plants species from 27 plant families [104]. Several features of the biology and the ecology of fall armyworm indicate that this species is composed of two morphologically undistinguishable ecological races (“corn strain” and “rice strain”), with distinct host-preferences [116, 119, 120]. Interestingly, the use of molecular markers [121-125] have indicated that both strains are generally found in sympatry [119, 126, 127], hence leading some authors that they could constitute a potential example of ongoing or sympatric speciation [106, 117]. In addition to the case of *S. frugiperda*, we will also examine more closely five other *Spodoptera* species, whose species status is still disputed. It is the case of *S. marima* and *S. ornithogalli*, which are morphologically related and possibly correspond to geographic races, as they have allopatric distributions and only differ by the absence of sexual dimorphism in *S. marima* males [104, 109]. These two species were found paraphyletic with a strong support (as indicated by Bayes factors) in a previous study [110], so we expect additional clarifications with the species delimitation procedures. Another species pair (*S. cosmiodes* and *S. descoinsi*) was also recovered paraphyletic with a strong support in the study of Kergoat *et al.* [110], but similarly to the case of *S. frugiperda*, additional evidences lend to think that these two entities possibly constitute

valid species [104, 105]. Finally, the last case that remains to be investigated is the status of the subspecies *S. mauritia acronyctoides*, which either constitutes a geographic race of *S. mauritia* or a valid species on its own [108]. For all these taxa, we plan to compare and discuss the results of GMYC analyses in the light of all other sources of information we have at our disposition.

In summary, the primary aims of this study are to: i) assess the sensitivity of branch length estimation on GMYC approach, using different conditions of phylogenetic analyses within a Bayesian framework, and ii) implement GMYC procedures to explore *Spodoptera* species boundaries in an integrative way.

Materiel and method

Sampling design

For this study we used 128 sequences from the study of [110] and 29 sequences for the study of [125, 128] to assemble a dataset of 157 individuals for 25 *Spodoptera* species. To complete this dataset, we sequenced specimens collected by J. Barbut by means of light trapping. Additional specimens were also obtained from museum collections (Table 1). For the latter, we preferentially picked specimens from distinct localities/countries in order to maximize the geographic coverage of each species. Out of the 30 currently recognized species [104], 232

Table 1: Taxon sampling.

Genus species	Voucher no.	Genbank accession no.	Origin
<i>Spodoptera</i>			
<i>albula</i> (Walker, 1857)	B20*	HQ177285	French Guiana
<i>albula</i> (Walker, 1857)	B53*	HQ177286	Honduras
<i>albula</i> (Walker, 1857)	LSU16*	HQ177288	French Guiana
<i>albula</i> (Walker, 1857)	LSU24*	HQ177289	USA - Louisiana
<i>albula</i> (Walker, 1857)	MNHN26**		Venezuela
<i>albula</i> (Walker, 1857)	MNHN27**		Jamaica
<i>albula</i> (Walker, 1857)	MNHN30**		Panama
<i>androgea</i> (Stoll, 1782)	B1*	HQ177290	Guadeloupe
<i>androgea</i> (Stoll, 1782)	MNHN20**		Dominican Republic
<i>androgea</i> (Stoll, 1782)	MNHN21**		Panama
<i>androgea</i> (Stoll, 1782)	MNHN22**		Ecuador
<i>androgea</i> (Stoll, 1782)	MNHN23**		Martinique
<i>androgea</i> (Stoll, 1782)	MNHN24**		Peru
<i>androgea</i> (Stoll, 1782)	MNHN66**		Dominican Republic
<i>cilium Guenée</i> , 1852	GK123*	HQ177292	Kenya
<i>cilium Guenée</i> , 1852	GK124*	HQ177293	Kenya
<i>cilium Guenée</i> , 1852	GK125*	HQ177294	Kenya
<i>cosmiodes</i> (Walker, 1858)	B100*	HQ177297	French Guiana
<i>cosmiodes</i> (Walker, 1858)	B150*	HQ177298	Brazil
<i>cosmiodes</i> (Walker, 1858)	B151*	HQ177299	Brazil
<i>cosmiodes</i> (Walker, 1858)	B166*	HQ177300	Brazil
<i>cosmiodes</i> (Walker, 1858)	B27*	HQ177296	French Guiana
<i>cosmiodes</i> (Walker, 1858)	LSU12*	HQ177301	French Guiana
<i>cosmiodes</i> (Walker, 1858)	B19*	HQ177295	French Guiana
<i>depravata</i> (Butler, 1879)	GK206*	HQ177302	Japan
<i>depravata</i> (Butler, 1879)	MNHN84**		Australia
<i>descoinsi</i> L.C.&Silvain, 1994	B102*	HQ177306	French Guiana
<i>descoinsi</i> L.C.&Silvain, 1994	B157*	HQ177307	French Guiana
<i>descoinsi</i> L.C.&Silvain, 1994	B17*	HQ177303	French Guiana
<i>descoinsi</i> L.C.&Silvain, 1994	B32*	HQ177304	French Guiana
<i>descoinsi</i> L.C.&Silvain, 1994	B38*	HQ177305	French Guiana
<i>descoinsi</i> L.C.&Silvain, 1994	LSU30*	HQ177308	French Guiana
<i>descoinsi</i> L.C.&Silvain, 1994	LSU31*	HQ177309	French Guiana
<i>descoinsi</i> L.C.&Silvain, 1994	MNHN43**		French Guiana
<i>descoinsi</i> L.C.&Silvain, 1994	MNHN44**		French Guiana

<i>descoinsi</i> L.C.&Silvain, 1994	MNHN45**		French Guiana
<i>dolichos</i> (Fabricius, 1794)	B104*	HQ177315	Guadeloupe
<i>dolichos</i> (Fabricius, 1794)	B122*	HQ177316	USA - Georgia
<i>dolichos</i> (Fabricius, 1794)	B35*	HQ177313	French Guiana
<i>dolichos</i> (Fabricius, 1794)	B37*	HQ177314	Guadeloupe
<i>dolichos</i> (Fabricius, 1794)	LSU14*	HQ177318	French Guiana
<i>dolichos</i> (Fabricius, 1794)	LSU27*	HQ177319	USA - Florida
<i>dolichos</i> (Fabricius, 1794)	LSU3*	HQ177317	Guadeloupe
<i>dolichos</i> (Fabricius, 1794)	MNHN34**		DominicanRepublic
<i>dolichos</i> (Fabricius, 1794)	MNHN35**		Venezuela
<i>dolichos</i> (Fabricius, 1794)	MNHN36**		French Guiana
<i>dolichos</i> (Fabricius, 1794)	MNHN37**		Jamaica
<i>dolichos</i> (Fabricius, 1794)	MNHN38**		Martinique
<i>dolichos</i> (Fabricius, 1794)	MNHN65**		Peru
<i>dolichos</i> (Fabricius, 1794)	MNHN7**		Martinique
<i>dolichos</i> (Fabricius, 1794)	MNHN84**		Australia
<i>eridania</i> (Stoll, 1782)	B108*	HQ177328	Peru
<i>eridania</i> (Stoll, 1782)	B109*	HQ177329	Peru
<i>eridania</i> (Stoll, 1782)	B16*	HQ177321	USA - Louisiana
<i>eridania</i> (Stoll, 1782)	B47*	HQ177322	Guadeloupe
<i>eridania</i> (Stoll, 1782)	B70*	HQ177323	USA - Louisiana
<i>eridania</i> (Stoll, 1782)	B71*	HQ177324	USA - Louisiana
<i>eridania</i> (Stoll, 1782)	B72*	HQ177325	USA - Louisiana
<i>eridania</i> (Stoll, 1782)	B73*	HQ177326	USA - Louisiana
<i>eridania</i> (Stoll, 1782)	B75*	HQ177327	Guadeloupe
<i>eridania</i> (Stoll, 1782)	HM756081***	HM756081	NA
<i>eridania</i> (Stoll, 1782)	HM756082***	HM756082	NA
<i>eridania</i> (Stoll, 1782)	HM756083***	HM756083	NA
<i>eridania</i> (Stoll, 1782)	HM756084***	HM756084	NA
<i>eridania</i> (Stoll, 1782)	HM756085***	HM756085	NA
<i>eridania</i> (Stoll, 1782)	LSU5*	HQ177330	USA
<i>evanida Schaus, 1914</i>	B155*	HQ177310	French Guiana
<i>evanida Schaus, 1914</i>	B156*	HQ177311	French Guiana
<i>evanida Schaus, 1914</i>	MNHN52**		French Guiana
<i>evanida Schaus, 1914</i>	MNHN53**		French Guiana
<i>evanida Schaus, 1914</i>	MNHN54**		French Guiana
<i>evanida Schaus, 1914</i>	MNHN55**		French Guiana
<i>evanida Schaus, 1914</i>	MNHN60**		French Guiana
<i>evanida Schaus, 1914</i>	MNHN61**		French Guiana
<i>evanida Schaus, 1914</i>	MNHN71**		Peru
<i>evanida Schaus, 1914</i>	MNHN72**		French Guiana
<i>exempta</i> (Walker, 1857)	B162*	HQ177334	Kenya
<i>exempta</i> (Walker, 1857)	B163*	HQ177335	Kenya
<i>exempta</i> (Walker, 1857)	B24*	HQ177331	Belgium (Laboratory)
<i>exempta</i> (Walker, 1857)	B25*	HQ177332	Belgium (Laboratory)
<i>exempta</i> (Walker, 1857)	B93*	HQ177333	Belgium (Laboratory)
<i>exigua</i> (Hübner, 1808)	B34*	HQ177337	Kenya
<i>exigua</i> (Hübner, 1808)	B67*	HQ177338	USA - California
<i>exigua</i> (Hübner, 1808)	B68*	HQ177339	France (Laboratory)
<i>exigua</i> (Hübner, 1808)	B69*	HQ177340	France (Laboratory)
<i>exigua</i> (Hübner, 1808)	G2178*	HQ177343	Egypt
<i>exigua</i> (Hübner, 1808)	G2182*	HQ177344	Egypt
<i>exigua</i> (Hübner, 1808)	G2185*	HQ177345	Egypt
<i>exigua</i> (Hübner, 1808)	HM756077***	HM756077	NA
<i>exigua</i> (Hübner, 1808)	HM756078***	HM756078	NA
<i>exigua</i> (Hübner, 1808)	HM756079***	HM756079	NA
<i>exigua</i> (Hübner, 1808)	HM756080***	HM756080	NA
<i>exigua</i> (Hübner, 1808)	LSU11*	HQ177341	USA
<i>exigua</i> (Hübner, 1808)	LSU15*	HQ177342	USA- California
<i>frugiperda</i> Corn (Smith, 1797)	B90C*	HQ177351	Peru (ex. maize)
<i>frugiperda</i> Corn (Smith, 1797)	FRUB1*		Brazil
<i>frugiperda</i> Corn (Smith, 1797)	FRUB11*		Brazil
<i>frugiperda</i> Corn (Smith, 1797)	FRUB12*		Brazil
<i>frugiperda</i> Corn (Smith, 1797)	FRUB13*		Brazil

<i>frugiperda</i> Corn (Smith, 1797)	FRUB14*		Brazil
<i>frugiperda</i> Corn (Smith, 1797)	FRUB15*		Brazil
<i>frugiperda</i> Corn (Smith, 1797)	FRUB16*		Brazil
<i>frugiperda</i> Corn (Smith, 1797)	FRUB17*		Brazil
<i>frugiperda</i> Corn (Smith, 1797)	FRUB3*		Brazil
<i>frugiperda</i> Corn (Smith, 1797)	FRUB4*		Brazil
<i>frugiperda</i> Corn (Smith, 1797)	FRUB7*		Brazil
<i>frugiperda</i> Corn (Smith, 1797)	FRUB8*		Brazil
<i>frugiperda</i> Corn (Smith, 1797)	G1 135*		Guadeloupe
<i>frugiperda</i> Corn (Smith, 1797)	G4 153*		Guadeloupe
<i>frugiperda</i> Corn (Smith, 1797)	HM136586***	HM136586	NA
<i>frugiperda</i> Corn (Smith, 1797)	HM136587***	HM136587	NA
<i>frugiperda</i> Corn (Smith, 1797)	HM136588***	HM136588	NA
<i>frugiperda</i> Corn (Smith, 1797)	HM136589***	HM136589	NA
<i>frugiperda</i> Corn (Smith, 1797)	HM136590***	HM136590	NA
<i>frugiperda</i> Corn (Smith, 1797)	HM136591***	HM136591	NA
<i>frugiperda</i> Corn (Smith, 1797)	HM136592***	HM136592	NA
<i>frugiperda</i> Corn (Smith, 1797)	LSU21*	HQ177353	Guadeloupe (ex.maize)
<i>frugiperda</i> Corn (Smith, 1797)	MNHN12**		Paraguay
<i>frugiperda</i> Corn (Smith, 1797)	MNHN14**		DominicanRepublic
<i>frugiperda</i> Corn (Smith, 1797)	MNHN16**		DominicanRepublic
<i>frugiperda</i> Corn (Smith, 1797)	MNHN56**		Cuba
<i>frugiperda</i> Corn (Smith, 1797)	MNHN81**		Peru
<i>frugiperda</i> Rice (Smith, 1797)	311 MF67*		French Guiana
<i>frugiperda</i> Rice (Smith, 1797)	337 MR7*		French Guiana
<i>frugiperda</i> Rice (Smith, 1797)	B10R*	HQ177348	French Guiana (ex.rice)
<i>frugiperda</i> Rice (Smith, 1797)	B12R*	HQ177349	French Guiana (ex.rice)
<i>frugiperda</i> Rice (Smith, 1797)	B7R*	HQ177346	French Guiana (ex.rice)
<i>frugiperda</i> Rice (Smith, 1797)	CE1*		French Guiana
<i>frugiperda</i> Rice (Smith, 1797)	F1 329*		French Guiana
<i>frugiperda</i> Rice (Smith, 1797)	F1 330*		French Guiana
<i>frugiperda</i> Rice (Smith, 1797)	F4 322*		French Guiana
<i>frugiperda</i> Rice (Smith, 1797)	FRUB10*		Brazil
<i>frugiperda</i> Rice (Smith, 1797)	FRUB9*		Brazil
<i>frugiperda</i> Rice (Smith, 1797)	G9 305*		French Guiana
<i>frugiperda</i> Rice (Smith, 1797)	G9 308*		French Guiana
<i>frugiperda</i> Rice (Smith, 1797)	GK424*		Brazil
<i>frugiperda</i> Rice (Smith, 1797)	GK427*		Brazil
<i>frugiperda</i> Rice (Smith, 1797)	GK429*		Brazil
<i>frugiperda</i> Rice (Smith, 1797)	HM136593***	HM136593	NA
<i>frugiperda</i> Rice (Smith, 1797)	HM136594R***	HM136594	NA
<i>frugiperda</i> Rice (Smith, 1797)	HM136595R***	HM136595	NA
<i>frugiperda</i> Rice (Smith, 1797)	HM136596R***	HM136596	NA
<i>frugiperda</i> Rice (Smith, 1797)	HM136597R***	HM136597	NA
<i>frugiperda</i> Rice (Smith, 1797)	HM136598R***	HM136598	NA
<i>frugiperda</i> Rice (Smith, 1797)	LSU22R*	HQ177350	Guadeloupe (ex.rice)
<i>frugiperda</i> Rice (Smith, 1797)	MFG 50*		French Guiana
<i>frugiperda</i> Rice (Smith, 1797)	MFG 56*		French Guiana
<i>frugiperda</i> Rice (Smith, 1797)	MNHN10**		DominicanRepublic
<i>frugiperda</i> Rice (Smith, 1797)	MNHN11**		Venezuela
<i>frugiperda</i> Rice (Smith, 1797)	MNHN13**		Mexico
<i>frugiperda</i> Rice (Smith, 1797)	MNHN15**		Guatemala
<i>frugiperda</i> Rice (Smith, 1797)	MNHN9**		DominicanRepublic
<i>latifascia</i> (Walker, 1856)	MNHN46**		Mexico
<i>latifascia</i> (Walker, 1856)	MNHN47**		Guatemala
<i>latifascia</i> (Walker, 1856)	MNHN48**		Costa Rica
<i>latifascia</i> (Walker, 1856)	MNHN51**		DominicanRepublic
<i>latifascia</i> (Walker, 1856)	B123*	HQ177361	USA - Georgia
<i>latifascia</i> (Walker, 1856)	B18*	HQ177355	Guadeloupe
<i>latifascia</i> (Walker, 1856)	B29*	HQ177356	Guadeloupe
<i>latifascia</i> (Walker, 1856)	B3*	HQ177354	USA - Louisiana
<i>latifascia</i> (Walker, 1856)	B31*	HQ177357	Barbados
<i>latifascia</i> (Walker, 1856)	B33*	HQ177358	USA - Louisiana
<i>latifascia</i> (Walker, 1856)	B36*	HQ177359	Barbados

<i>latifascia</i> (Walker, 1856)	B39*	HQ177360	USA - Louisiana
<i>latifascia</i> (Walker, 1856)	LSU17*	HQ177363	USA
<i>latifascia</i> (Walker, 1856)	LSU6*	HQ177362	Guadeloupe
<i>littoralis</i> (Boisduval, 1833)	ADN0*	HQ177371	Egypt
<i>littoralis</i> (Boisduval, 1833)	B14*	HQ177364	France (Laboratory)
<i>littoralis</i> (Boisduval, 1833)	B164*	HQ177367	Congo
<i>littoralis</i> (Boisduval, 1833)	B165*	HQ177368	Congo
<i>littoralis</i> (Boisduval, 1833)	B98*	HQ177365	Egypt
<i>littoralis</i> (Boisduval, 1833)	B99*	HQ177366	Egypt
<i>littoralis</i> (Boisduval, 1833)	GK118*	HQ177372	Mali
<i>littoralis</i> (Boisduval, 1833)	GK119*	HQ177373	Mali
<i>littoralis</i> (Boisduval, 1833)	GK120*	HQ177374	Mali
<i>littoralis</i> (Boisduval, 1833)	HM756074***	HM756074	NA
<i>littoralis</i> (Boisduval, 1833)	LSU35*	HQ177370	France (Laboratory)
<i>littoralis</i> (Boisduval, 1833)	LSU7*	HQ177369	France (Laboratory)
<i>litura</i> (Fabricius, 1775)	B113*	HQ177376	Australia
<i>litura</i> (Fabricius, 1775)	B114*	HQ177377	Australia
<i>litura</i> (Fabricius, 1775)	B15*	HQ177375	Indonesia
<i>litura</i> (Fabricius, 1775)	B153*	HQ177378	Indonesia
<i>litura</i> (Fabricius, 1775)	B154*	HQ177379	Philippines
<i>litura</i> (Fabricius, 1775)	B160*	HQ177380	Indonesia
<i>litura</i> (Fabricius, 1775)	GK425*		Indonesia
<i>litura</i> (Fabricius, 1775)	GK426*		Indonesia
<i>litura</i> (Fabricius, 1775)	HM756090***	HM756090	NA
<i>litura</i> (Fabricius, 1775)	HM756091***	HM756091	NA
<i>litura</i> (Fabricius, 1775)	HM756092***	HM756092	NA
<i>litura</i> (Fabricius, 1775)	HM756093***	HM756093	NA
<i>litura</i> (Fabricius, 1775)	JSTR00021 0103		China
<i>litura</i> (Fabricius, 1775)	JSTR00026 0201		China
<i>litura</i> (Fabricius, 1775)	JSTR00026 0202		China
<i>marima</i> (Schaus, 1904)	B145*	HQ177381	French Guiana
<i>marima</i> (Schaus, 1904)	MNHN39**		Venezuela
<i>marima</i> (Schaus, 1904)	MNHN40**		Venezuela
<i>marima</i> (Schaus, 1904)	MNHN58**		French Guiana
<i>mauritica</i> (Boisduval, 1833)	B65*	HQ177382	Reunion
<i>mauritica</i> (Boisduval, 1833)	B66*	HQ177383	Reunion
<i>mauritica</i> (Boisduval, 1833)	B94*	HQ177384	Reunion
<i>mauritica</i> (Boisduval, 1833)	B95*	HQ177385	Reunion
<i>m.acronyctoides</i>	B152*	HQ177386	Papua New Guinea
<i>ochrea</i> (Hampson, 1909)	B110*	HQ177390	Peru
<i>ochrea</i> (Hampson, 1909)	B84*	HQ177387	Peru
<i>ochrea</i> (Hampson, 1909)	B85*	HQ177388	Peru
<i>ochrea</i> (Hampson, 1909)	B86*	HQ177389	Peru
<i>ochrea</i> (Hampson, 1909)	MNHN75**		Peru
<i>ornithogalli</i> Guenée, 1852	B106*	HQ177394	USA
<i>ornithogalli</i> Guenée, 1853	B121*	HQ177395	USA - Georgia
<i>ornithogalli</i> Guenée, 1854	B146*	HQ177396	French Guiana
<i>ornithogalli</i> Guenée, 1855	B28*	HQ177392	USA - Louisiana
<i>ornithogalli</i> Guenée, 1856	B4*	HQ177391	USA - Louisiana
<i>ornithogalli</i> Guenée, 1857	B54*	HQ177393	USA - Louisiana
<i>ornithogalli</i> Guenée, 1858	LSU34*	HQ177398	USA - Louisiana
<i>ornithogalli</i> Guenée, 1859	LSU8*	HQ177397	USA
<i>ornithogalli</i> Guenée, 1860	MNHN41**		Guatemala
<i>ornithogalli</i> Guenée, 1861	MNHN69**		Dominican Republic
<i>pecten</i> Guenée, 1852	P1*	HQ177399	Indonesia
<i>pecten</i> Guenée, 1852	P2*	HQ177400	Indonesia
<i>pectinicornis</i> (Hampson, 1895)	B142*	HQ177401	Australia
<i>pectinicornis</i> (Hampson, 1895)	B143*	HQ177402	Australia
<i>pectinicornis</i> (Hampson, 1895)	B144*	HQ177403	Australia
<i>picta</i> (Guérin-Méneville, 1838)	B158*	HQ177404	Australia
<i>picta</i> (Guérin-Méneville, 1838)	B159*	HQ177405	Australia
<i>praefica</i> (Grote, 1875)	B105*	HQ177408	USA - California
<i>praefica</i> (Grote, 1875)	B30*	HQ177407	USA - California
<i>praefica</i> (Grote, 1875)	B5*	HQ177406	USA - California

<i>praefica</i> (Grote, 1875)	LSU32*	HQ177409	USA - California
<i>pulchella</i> (H.-Shäffer, 1868)	HM756075***	HM756075	NA
<i>pulchella</i> (H.-Shäffer, 1868)	HM756076***	HM756076	NA
<i>triturrata</i> (Walker, 1857)	GK121*	HQ177411	Kenya
<i>triturrata</i> (Walker, 1857)	Tri*	HQ177410	Kenya
<i>Psychomorpha epimenis</i> (Drury, 1782)			
<i>Agrotis ipsilon</i> (Hufnagel, 1766)			
<i>Helicoverpa armigera</i> (Hübner, 1809)			
<i>Heliothis virescens</i> (Fabricius, 1777)			
<i>Mythimna unipuncta</i> (Haworth, 1809)			
<i>Noctua atlantica</i> (Warren, 1905)			
<i>Sesamia inferens</i> (Walker, 1856)	JN039362		

*CBGP laboratory.

**Museum National d'Histoire Naturelle, Paris.

***Genbank Nagoshi & Meagher

individuals from 25 species were included in the dataset (mean of 9.28 individuals per species; minimum of two individuals and maximum of 58 individuals for *S. frugiperda*) (Table 1). Finally we used the results of the study of Mitchell *et al.* (2006) [129] to select seven outgroup taxa and root the trees with *Psychomorpha epimenis*.

DNA extraction and polymerase chain reaction

Total genomic DNA of new samples was extracted by grinding up hind legs using DNAeasy Blood and Tissue Kit (Qiagen). Partial fragments of the cytochrome c oxidase subunit I (COI) gene were amplified using standard primers used in a previous study [110]. The polymerase chain reaction (PCR) mix (25 μ L) consisted of 1xPCR buffer, 1.5mM MgCl₂, 0.1mM of each desoxynucleotide triphosphates, 1 μ M of each primer, 1 unit of Taq DNA polymerase (Qiagen) and 2 μ L (diluted to tenth) of extracted DNA and nuclease-free water to 25 μ L. PCR cycling conditions were the following: initial denaturation step at 95°C for 3 min followed by 35 cycles including a denaturation step at 92°C for 1 min, an annealing step at 48°C for 1 min, and an elongation step at 72°C for 1 min

and a final elongation step at 72°C for 10 min. The resulting PCR products were processed by the French sequencing centre Genoscope using a BigDye v3.1 sequencing kit and Applied 3730xl sequencers. Both strands were sequenced for all specimens to minimize PCR artifacts and ambiguities. Sequences of complementary strands were edited and reconciled using Geneious v3.8 software [130]. The alignment of all sequences was carried out using ClustalW version 2 [131] with default option settings. No gap was present in the corresponding alignment. We then used Mesquite v2.75 [132] to check the coding frame for possible errors or stop codons. The final dataset encompasses 239 individuals and 1,403 aligned characters. Among the 335,317 cells in the data matrix, 107,865 (32.16 %) correspond to missing data.

Bayesian inference analyses

Our dataset was analyzed under Bayesian inference (BI), using a diverse array of settings and priors that can be enforced with MrBayes v3.1.2 [133]. For more clarity, we provided an overview of the computational pipeline implemented for this study in Figure 2.

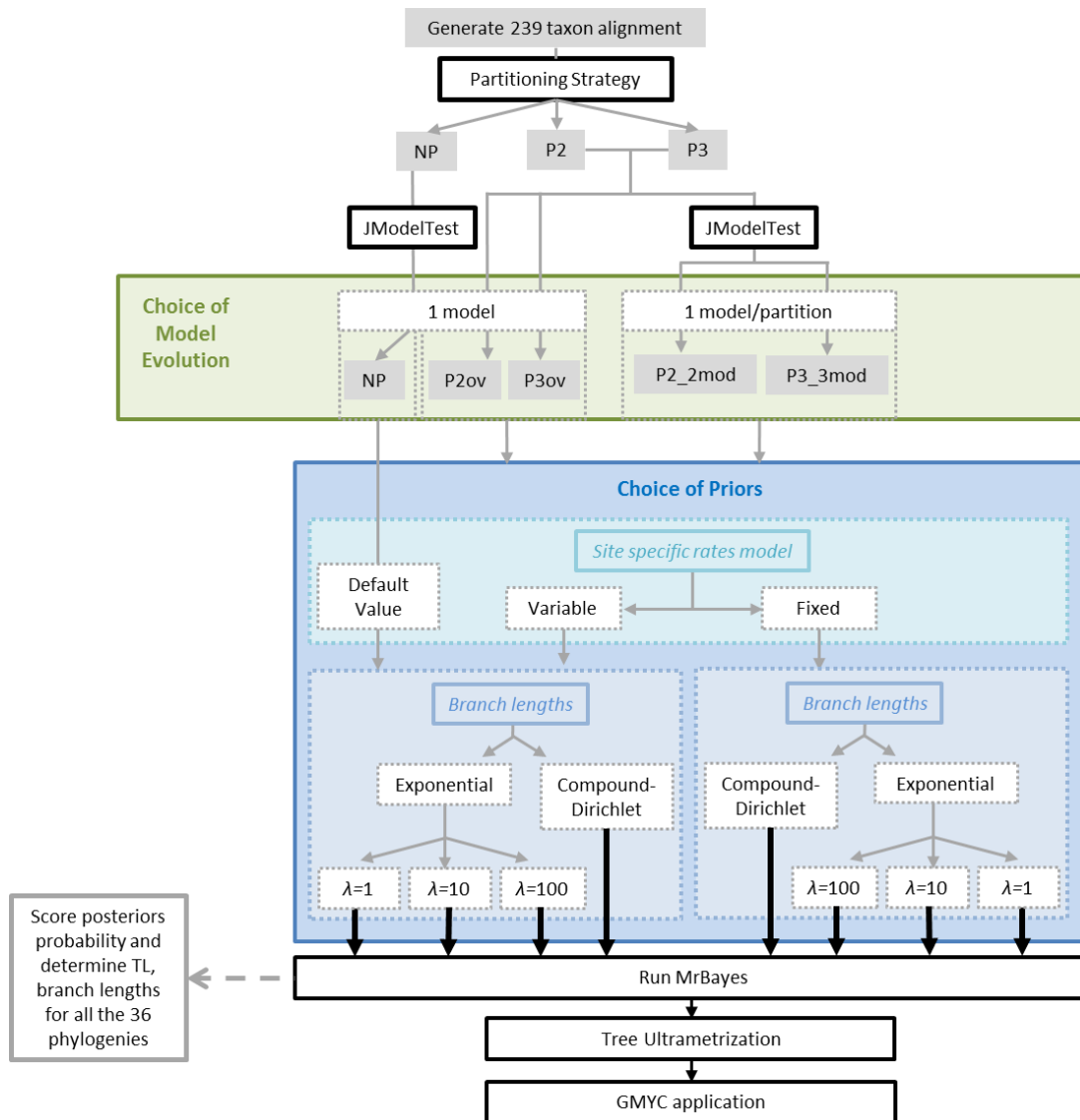


Figure 2: Computational pipeline implemented for species delimitation analysis. Grey boxes indicate a dataset, grey arrows indicate modifications of parameter analysis, and grey dotted boxes indicate that particular analysis parameter. Black arrows and boxes indicate phylogenetic analysis and other type of analysis performed on particular datasets. The grey dotted lines indicate the final analyses performed to obtain scores for each phylogenetic reconstruction.

i. Partitioning strategies

To assess whether the use of parameter-rich strategies under Bayesian inference may influence the outputs of GMYC procedures, we performed both non-partitioned (NP) and partitioned analyses ([101, 134]). For the partitioned analyses, we either used two partitions (one partition for codon positions 1 and 2 and one partition for codon position 3; P2ov) or three partitions (one partition for

each codon position; P3). The choice of substitution models was then performed using the corrected Akaike information criterion (AICc) as implemented in jModelTest2.0.1 [135] (Table 2). We also conducted additional partitioned analyses by enforcing the most complex substitution model (GTR+G +I) for each partition (corresponding analyses were named P2_2mod and P3_3mod).

Table 2: Partitioning strategies used in this study.

Partitioning strategy*	Definition
NP(1)	Unpartitioned data set
P2(2)	1 partition for the codon position 1 and 2 + 1 partition for the codon position 3
P3(3)	1 partition per codon position

*The number of partition is given into brackets

ii. Choice of priors

Two distinct site-specific rates models [133] were used for each batch of analyses (NP, P2ov, P3ov, P2_2mod and P3_3mod). We either specified a fixed rate (default value in MrBayes), which assumes that the average substitution rate (per site) is identical for each partition (obtained with the command *prsetratepr=fixed*), or a rate allowed to vary across partitions (obtained with the command *prsetratepr=variable*). Four distinct branch length priors were also implemented. First we tested three different exponentially distributed branch-length priors by changing the lambda argument (λ) of the exponential distribution to 1, 10 (default value in MrBayes) or 100 (obtained with the *prsetbrlenspr=unconstrained:exponential [λ] command*). Finally we conducted analyses using a multivariate prior for the branch lengths ("compound Dirichlet" priors [136]), which can be specified using the command *prsetbrlenspr=unconstrained:gammadir($\alpha_\tau, \beta_\tau, \alpha, c$)* implemented in a modified version of MrBayes 3.1.2 [99]. The compound Dirichlet prior specifies a fairly diffuse prior on the tree length (the sum of branch lengths) and uses a Dirichlet distribution to partition the tree length into branch lengths. We used the recommended value $\alpha_\tau = 1$ [136], so that $\beta_\tau = 1$ and parameters $\alpha = c = 1$ are fixed [99].

iii. Configuration of runs

Each analysis was run four times independently with eight Markov Chains

Monte Carlo (MCMC; one cold and seven incrementally heated), which ran for 10 million generations and were sampled every 1000 generations. Tracer [137] was used to examine the convergence across runs and the ESS values of sampled parameters. A burn-in of 25% was applied after checking for stability on the log-likelihood curves and the split-frequencies of the runs.

Processing of trees and GMYC procedure

We compared data from Bayesian posterior tree samples between all the analyses, for *Tree Length*, *Internal Branch Lengths* and *External Branch Lengths*. All corresponding values were estimated using the TreeStat package (http://beast.bio.ed.ac.uk/Main_Page). We then applied the mean path length (MPL) method as implemented in Pathd8 [138] to produce the GMYC input tree [80]. This approach was preferred because it does not require any specific calibration [139]. The GMYC procedure was then implemented using the *splits* package in R (available at <http://r-forge.r-project.org/projects/splits/>). We used here a single threshold value for our input tree [80, 84, 140].

Results

Bayesian trees

Figure 3 shows the topology of the consensus tree with the best mean likelihood value (lnL = -9033.66) resulting from the P3ov analysis

with a compound dirichlet branch-length prior and the same substitution rate for each partition. In all 45 analyses, the genus *Spodoptera* is recovered monophyletic with high clade posterior probabilities (CPP \geq 95 [141]). Within the genus *Spodoptera*, high support values are also recovered for the clades that correspond to putative species.

Spodoptera frugiperda consists of two well-supported clades (CPP of 1.0) whereas *S. m. acronyctoides* is found in a sister group position to the remaining members of *S. mauritia* (CPP of 1.0). Conversely, the two species, *S. descoinsi* and *S. ornithogalli* are found paraphyletic due to the position of individuals from *S. cosmiodes* and *S. marima*.

Variations in Tree Length

For all three partitioning strategies, tree lengths displayed unimodal responses to priors on branch lengths, regardless of the substitution rate priors used (Figure 4). Long branch length priors effectively cause tree lengths to become unrealistically high (more than 400 times longer than a default analysis).

Similarly, the shorter the prior is, the shorter the tree length is. Internal and external branches are similarly affected by the branch-length prior and substitution rate prior (Figure 5). Finally, tree lengths values obtained with short exponential branch-length prior ($\lambda = 100$) and compound dirichlet branch-length prior are the same.

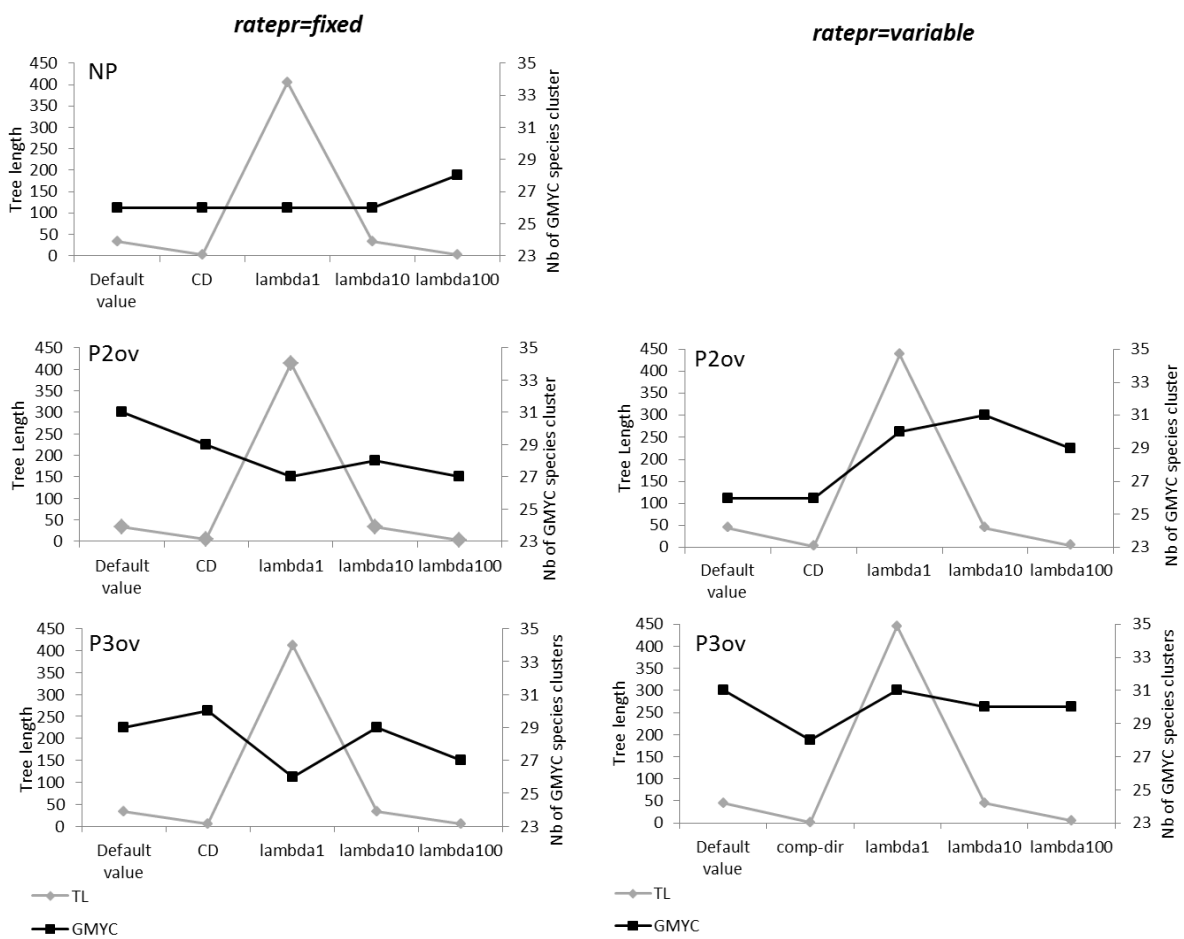


Figure 4: Tree length value and GMYC clusters number depending on branch-lengths prior for the three partitioning strategies under whether a fixed or a variable substitution rate priors.

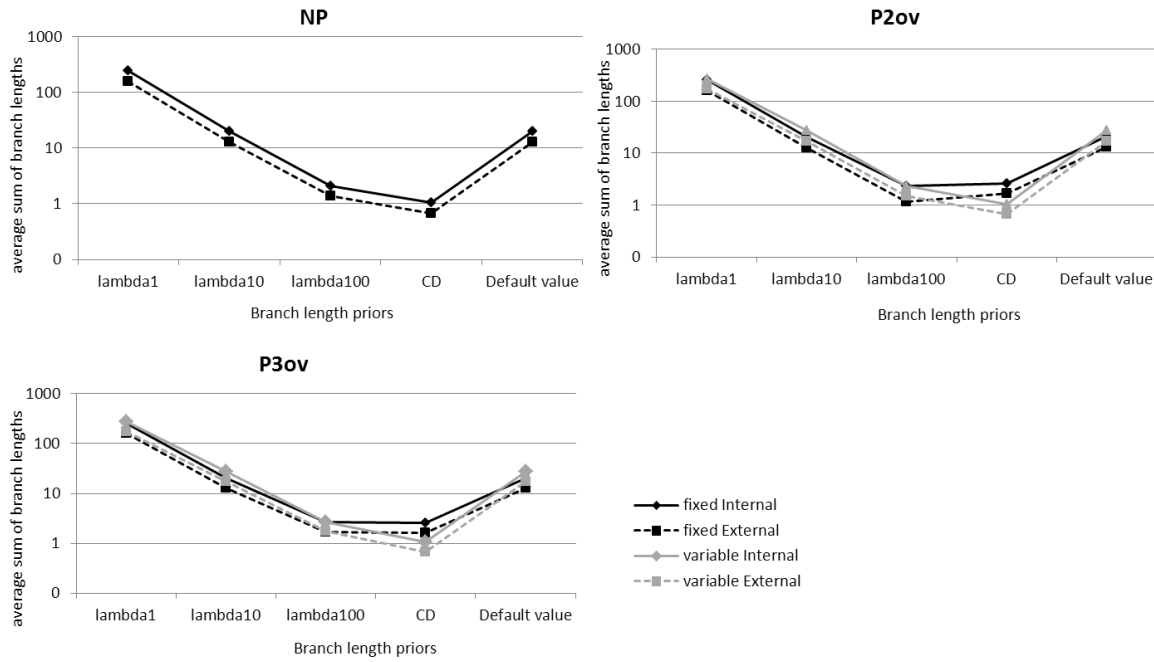


Figure 5: Average of summed internal and external branch lengths depending on the branch lengths prior for the three partitioning strategies.

Table 3: Synthetic table of likelihood values and means of tree length obtained for the different phylogenetic reconstructions. For each phylogeny the number of species clusters obtained with the GMYC analyses is indicated.

Substitution rate prior	Branch lengths prior	P2ov						P3ov						NP					
		LnL	TL	cluster GMYC			LnL	TL	cluster GMYC			LnL	TL	cluster GMYC					
				min	mean	max			min	mean	max			min	mean	max			
fixed	λ1	-9542,35	413,69	25	27	33	-9501,02	411,56	25	26	30	-10103,07	405,04	24	26	29			
	λ10 (default value)	-9323,93	33,11	28	28	31	-9265,96	33,7	28	29	32	-9870,33	33,2	24	26	29			
	λ100	-9143,25	3,48	25	27	32	-9088,63	4,38	24	27	31	-9609,25	3,47	24	28	31			
	Compound dirichlet (CD)	-9144,69	4,33	25	29	33	-9090,25	4,24	23	30	34	-9575,66	1,73	23	26	186			
variable	λ1	-9190,73	438,02	25	30	31	-9056,37	448,26	30	34	37								
	λ10	-9183,43	44,25	24	31	35	-9056,44	44,5	25	27	31								
	λ100	-9124,67	3,83	23	29	36	-9054,17	4,44	24	32	34								
	Compound dirichlet (CD)	-9079,45	1,72	25	26	28	-9033,66	1,72	26	28	32								

Substitution rate prior	Branch lengths prior	P2_2mod						P3_3mod					
		LnL	TL	cluster GMYC			LnL	TL	cluster GMYC				
				min	mean	max			min	mean	max		
fixed	λ1	-9513,95	419,06	25	27	36	-9488,71	412,4	25	28	32		
	λ10 (default value)	-9325,18	34,3	28	28	30	-9266,11	34,13	28	29	30		
	λ100	-9145,94	4,41	24	30	34	-9091,07	4,36	24	28	34		
	Compound dirichlet (CD)	-9147,94	4,27	25	27	34	-9092,91	4,16	24	27	193		
variable	λ1	-9204,54	438,64	25	27	30	-9065,33	445,26	25	31	35		
	λ10 (default value)	-9188,85	44,37	24	26	32	-9066,57	44,45	23	30	34		
	λ100	-9110,35	3,86	27	33	35	-9057,59	4,44	24	30	34		
	Compound dirichlet (CD)	-9084,06	1,7	23	31	194	-9034,14	1,7	23	28	191		

Effect on GMYC analyses

The number of species clusters found for each analysis is summarized in Table 3. The mean number of GMYC species clusters varies between 26 and 34. Figure 5-A displays the effect of partitioning strategies on the mean

number of GMYC species clusters found for the analyses using a fixed substitution rate prior. The mean number of GMYC species clusters is significantly higher for P2 and P3 partitioning strategies than NP strategy, as indicated by *p-value* of 0.026 and 0.007, respectively. However, the mean number of

GMYC species clusters is not significantly different between the P2 and P3 partitioning strategies, regardless of the prior on substitution rate used (Figure 6-B). Similarly, no correlation can be made between branch lengths priors and the number of resulting

species clusters (Figure 4). The mean number of species clusters obtained for P2 and P3 partitioning strategy (Figures 7-A and 7-B) is greater under a variable substitution rate prior than a fixed substitution rate prior.

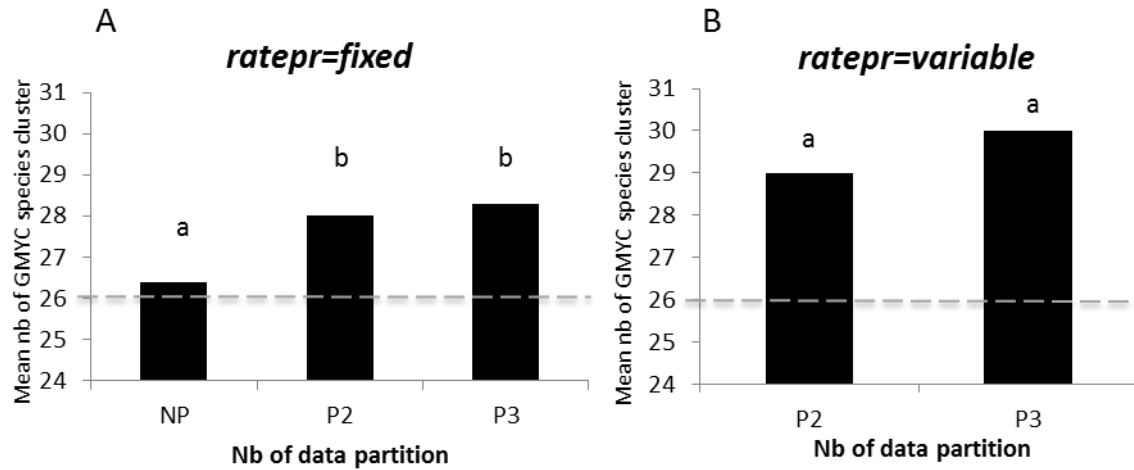


Figure 6: Mean number of GMYC species clusters depending of the partitioning strategies under whether a fixed (A) or a variable (B) substitution rate prior. The grey dotted line corresponds to the number of biological species expected.

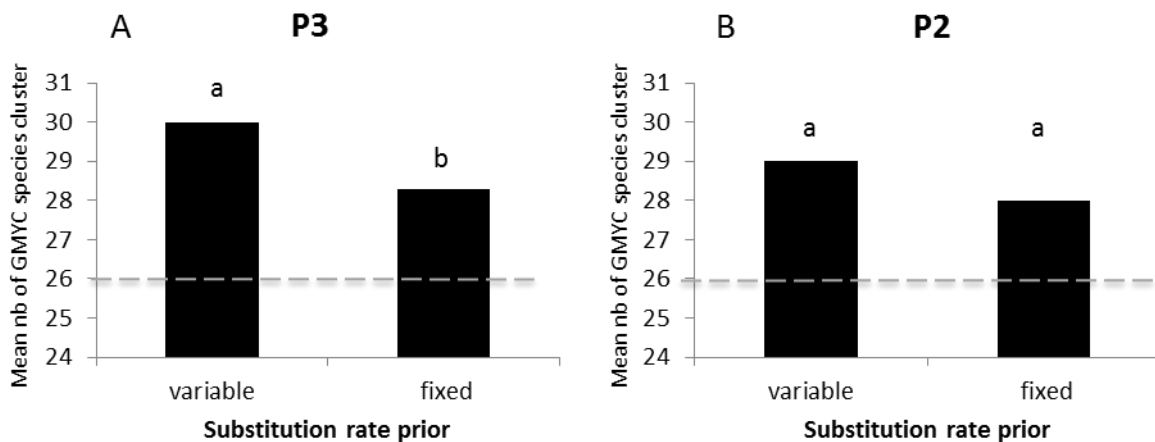


Figure 7: Mean number of GMYC species clusters depending on the substitution rate prior for the P3 partitioned strategy (A) and the P2 partitioned strategy (B). The grey dotted line corresponds to the number of biological species expected.

Nevertheless the increase is only significant for the P3 partitioning strategy, as indicate by a *p-value* of 0,048. Figure 8 displays the mean number of species clusters obtained with the

GMYC approach for the 36 analyses. Variation of the species cluster number between analyses is systematically situated within the *Spodoptera frugiperda* species complex.

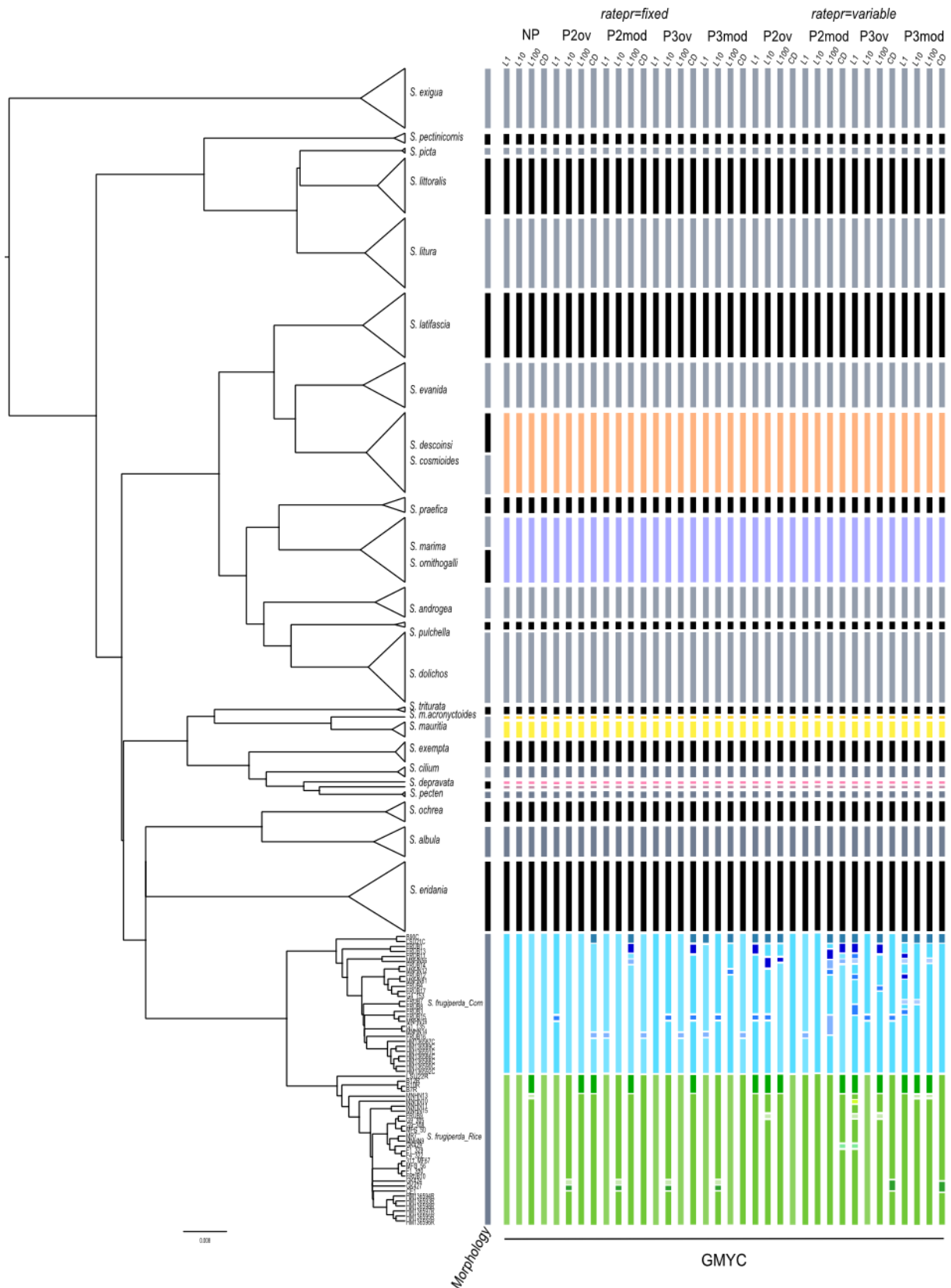


Figure 8: Species clusters as obtained according to the morphological delineation (column 1) and the GMYC method (columns 2-36). Each column (2-36) corresponds to GMYC results obtained for a particular phylogenetic analysis with specific parameters. Results are shown next to the tree obtained for the NP analysis.

Discussion

Our results allow to explore how the partitioning strategies or branch length priors used in the tree building procedures may impact the results of the GMYC approach. Our study highlights that the GMYC approach is not too sensitive to the partitioning strategies or branch length priors used in Bayesian inference. With the exception of *S. frugiperda*, our results show that the number of species clusters resulting from the application of the GMYC procedure is particularly stable (24 species cluster for our dataset, if excluding *S. frugiperda* specimens), whatever partitioning strategies and priors were used for phylogenetic reconstruction. Overall, this suggests that the GMYC method is a quite robust approach to delimitate species, especially when species are well differentiated. Thus, these results allow us to discuss the ambiguous species status for at least five taxa that already have been questioned [110] and which seem to be well differentiated, considering our species delimitation results.

First, for *S. mauritia* and *S. m. acronyctoides* the results of GMYC for each of the 36 analyses recovered the two taxa as separated species. However for these taxa we have not been able to increase sampling as was recommended by Kergoat *et al.* (2012) [110] that is why this result must be considered with caution. Second, regarding *S. ornithogalli* and *S. marima*, for which only few morphological (hindwing coloration [109]) and ecological (repartition area [104]) differences exist,

GMYC results are consistent between the analyses and cluster these two taxa into a single species cluster, confirming the synonymy hypothesis [110]. Third, the ambiguous species status of *S. cosmiodes* and *S. descoinsi*, for which morphological and biological marked differences exist [104, 105], seems paradoxically stable in view of our species delimitation results. Indeed, for each of the 36 analyses, the GMYC method clusters individuals of *S. cosmiodes* and *S. descoinsi* in only one species cluster. These results are in agreement with those of Kergoat *et al.* 2012 [110], at least concerning their results in parsimony analysis.

Interestingly, *S. frugiperda* is the only species for which GMYC results vary from one analysis to another. Indeed, for this taxon the GMYC approach seems to be sensitive to priors and partitioning strategy used for the phylogenetic reconstruction in Bayesian Inference. However, even though the mean number of GMYC clusters for *S. frugiperda* is variable (2 to 7), in most of case individuals are separated in two clusters corresponding to the two strains (rice and corn) of *S. frugiperda*. Our GMYC results could be related to the fact that the two strains of *S. frugiperda* are not fully differentiated and correspond to a species complex.

This result confirms the existence of a high level of genetic differentiation between the two strains, which is in agreement with biological characteristics of this moth. Both strains exhibit limitations in inter-strain mating in nature which has been the topic of several studies (Table 4).

Table 4: List of evidences arguing in favor of a high level of differentiation between the two FAW strains. Some evidences are more or less controversial.

	Suitability	References
Pre-zygotic barrier		
Habitat isolation	+ / -	[143]
Temporal isolation	+	[105] [142]
Compartmental isolation	++	[104] [118] [141]
Unidirectional mating	+/-	[144]
Post-zygotic barrier		
Reduction of hybrid fertility & reduction of hybrid female number	+	[147]
Genetic differentiation		
Phylogenetic analysis	++	[106]
GMYC analysis	++	Our results

Different prezygotic isolation mechanisms have been found, including differences in the composition of females' pheromones [106, 142], allochronism in mating activity between the two strains [107, 143] and oviposition preferences [120, 144]. However, success of mating between the two strains is highly controversial. Indeed Pashley & Martin, 1987 [145] have observed unidirectional mating between corn strain females and rice strain males while the reciprocal cross gives a viable offspring. On the other hand, other studies have shown that the two strains crossed successfully in both directions [144, 146]. Furthermore in backcrosses, the hybrid females mated with low success with their brothers but not at all with males of either parental strain [144, 145, 147]. Recently Velasquez-Velez *et al.*, 2011 [148] have found in both strains, postzygotic isolation for several life-history traits. They have likewise identified a reduction of the number of hybrid females and a reduction in hybrid fertility in *S. frugiperda*. Thus, results of our species delimitation study bring a new clue in favour of the assumption that the two strains of *S. frugiperda* may constitute separate species. However it is necessary to be cautious with these conclusions because some studies have found in nature a high level of hybrid, like 14 % in Brazil [149] and 16% in the USA and Central America [116].

In case of species complex, our results show that species delimitation with GMYC approach is very sensitive to analysis parameters used for BI phylogenetic reconstruction and exhibit some limitations. Indeed, the priors and the partitioning strategy must be carefully chosen because biases on branch lengths estimations affect the number of cluster found in GMYC. The number of species clusters obtained from non-partitioned analyses is stable, regardless of priors on branch length, whereas mean number of species clusters obtained from

partitioned dataset is variable and significantly higher than the expected number of biological species. In addition, when data is divided into numerous partitions, a variable substitution rate priors leads to a significant increase in the number of putative species clusters. These results suggest using more cautiously partitioning strategies for phylogenetic reconstruction in a Bayesian context, when species are not totally differentiated.

Conclusion

In this study we test the effect of Bayesian branch length priors on species delimitation analyses, through the study of the genus *Spodoptera*, for which several species status were disputed. When species status is clear and non-ambiguous, our study highlights robustness of GMYC approach. However when species are not well-differentiated, as is the case for *S. frugiperda* strains, this tree-based method present some limitations and the Bayesian tree reconstruction parameters, must be defined with caution. In this case, an excessive number of analysis parameters tends to increase in an unrealistically way the number of putative species cluster found in GMYC. The case of *frugiperda* species illustrates this problem and underlines the necessity to combine a variety of approaches in order to crosscheck information and better define species boundaries.

Acknowledgements

This work was partially supported by a grant from the SPE department of INRA (AAP SPE SPODOPTERA) and by a grant from the French National Research Agency grant (ANR-12-BSV7-0004-01; <http://www.agence-nationale-recherche.fr/>). Pascaline Dumas held doctoral fellowships from INRA and Montpellier 2 University. Part of the sequencing was

supported by a Genoscope project @-Speed-Id (Accurate SPEciEs Delimitation and Identification of eukaryotic biodiversity using DNA markers) proposed by F-BoL, the French Barcode of life initiative and by the program 'Bibliothèque du Vivant' (Project SPODOBAR) supported by a joint CNRS, INRA and MNHN consortium. Finally, we would also like to thank E. Joussein for her constructive comments on a preliminary version of the manuscript.

References

1. Joppa LN, Roberts DL, Myers N, Pimm SL: **Biodiversity hotspots house most undiscovered plant species.** *Proceedings of the National Academy of Sciences* 2011, **108**(32):13171-13176.
2. Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B: **How Many Species Are There on Earth and in the Ocean?** *PLoS Biology* 2011, **9**(8):e1001127.
3. Scheffers BR, Joppa LN, Pimm SL, Laurance WF: **What we know and don't know about Earth's missing biodiversity.** *Trends in Ecology & Evolution* 2012, **27**(9):501-510.
4. Wilson E: **Taxonomy as a fundamental discipline.** *Phil Trans Roy Soc Lond B* 2004, **359**:739.
5. Joppa LN, Roberts DL, Pimm SL: **The population ecology and social behaviour of taxonomists.** *Trends in Ecology & Evolution* 2011, **26**(11):551-553.
6. Fontaine B, Perrard A, Bouchet P: **21 years of shelf life between discovery and description of new species.** *Current Biology* 2012, **22**(22):R943-R944.
7. Pimm SL, Russell GJ, Gittleman JL, Brooks TM: **The Future of Biodiversity.** *Science* 1995, **269**(5222):347-350.
8. Mace GM: **The role of taxonomy in species conservation.** *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* 2004, **359**(1444):711-719.
9. Schlick-Steiner B, Steiner F, Seifert B, Stauffer C, Christian E, Crozier R: **Integrative taxonomy: a multisource approach to exploring Biodiversity.** *Annu Rev Entomol* 2010, **55**:421-438.
10. Alvarez N, Hossaert-McKey M, Rasplus J-Y, McKey D, Mercier L, Soldati L, Aebi A, Shani T, Benrey B: **Sibling species of bean bruchids : a morphological and phylogenetic study of *Acanthoscelides obtectus* Say and *Acanthoscelides obvelatus* Bridwell.** *Journal of Zoological Systematics and Evolutionary Research* 2004, **43** 29-37.
11. Bond JE, Sierwald P: **Cryptic speciation in the *Anadenobolus excisus* millipede species complex on the island of Jamaica.** *Evolution* 2002, **56**:1123-1135.
12. Collin R: **The utility of morphological characters in gastropod phylogenetics : an example from the Calyptraeidae.** *Biological Journal of the Linnean Society* 2003, **78**:541-593.
13. Huber BA, Rheims CA, Brescovit AD: **Speciation without changes in genital shape: a case study on Brazilian pholcid spiders (Araneae: Pholcidae).** *Zoologischer Anzeiger* 2005 **243**:273-279.
14. Kergoat GJ, Alvarez N: **Assessing the phylogenetic usefulness of a previously neglected morphological structure through elliptic Fourier analyses: a case study in *Bruchus* seed-beetles (Coleoptera: Chrysomelidae: Bruchinae: Bruchini).**

- Systematic Entomology* 2008, **33**:289-300.
15. Kodandaramaiah U, Weingartner E, Janz N, Leski M, Slove J, Warren A, Nylin S: **Investigating Concordance among Genetic Data, Subspecies Circumscriptions and Hostplant Use in the Nymphalid Butterfly *Polygonia faunus***. *Plos One* 2012, **7**(7):e41058.
 16. Gentile G, della Torre A, Maegga B, Powell JR, Caccone A: **Genetic Differentiation in the African Malaria Vector, *Anopheles gambiae* s.s., and the Problem of Taxonomic Status**. *Genetics* 2002, **161**(4):1561-1578.
 17. Pringle A, Baker DM, Platt JL, Wares JP, Latgé JP, Taylor JW: **Cryptic speciation in the cosmopolitan and clonal human pathogenic fungus *Aspergillus fumigatus***. *Evolution* 2005, **59**(9):1886-1899.
 18. Armstrong K, Ball S: **DNA barcodes for biosecurity: invasive species identification**. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* 2005, **360**(1462):1813 - 1823.
 19. Frey J, Guillen L, Frey B, Samietz J, Rull J, Aluja M: **Developing diagnostic SNP panels for the identification of true fruit flies (Diptera: Tephritidae) within the limits of COI-based species delimitation**. *BMC Evolutionary Biology* 2013, **13**(1):106.
 20. Roe A, Sperling F: **Population structure and species boundary delimitation of cryptic *Dioryctria* moths: an integrative approach**. *Molecular Ecology* 2007, **16**:3617 - 3633.
 21. Ross KG, Gotzek D, Ascunce MS, Shoemaker DD: **Species Delimitation: A Case Study in a Problematic Ant Taxon**. *Systematic Biology* 2010, **59**(2):162-184.
 22. Edwards SV: **Is a new and general theory of molecular systematic emerging?** *Evolution* 2009, **63**(1):1-19.
 23. Blaxter ML: **The promise of a DNA taxonomy**. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* 2004, **359**(1444):669-679.
 24. Hebert P, Cywinska A, Ball S, deWaard J: **Biological identifications through DNA barcodes**. *Proceedings of the Royal Society B: Biological Sciences* 2003, **270**(1512):313 - 321.
 25. Hebert P, Penton E, Burns J, Janzen D, Hallwachs W: **Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator***. *Proceedings of the National Academy of Sciences U S A* 2004, **101**(41):14812 - 14817.
 26. Hebert PDN, Gregory TR: **The Promise of DNA Barcoding for Taxonomy**. *Systematic Biology* 2005, **54**(5):852-859.
 27. Hebert PDN, Stoeckle MY, Zemplak TS, Francis CM: **Identification of Birds through DNA Barcodes**. *PLoS Biol* 2004, **2**(10):e312.
 28. Janzen DH: **Now is the time**. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* 2004, **359**(1444):731-732.
 29. Dincă V, Zakharov EV, Hebert PDN, Vila R: **Complete DNA barcode reference library for a country's butterfly fauna reveals high performance for temperate Europe**. *Proceedings of the Royal Society B: Biological Sciences* 2011, **278**(1704):347-355.
 30. Janzen D, Hallwachs W, Blandin P, Burns J, Cadiou J, Chacon I, Dapkey T, Deans A, Epstein M, Espinoza B *et al*: **Integration of DNA barcoding into an**

- ongoing inventory of complex tropical biodiversity. *Molecular Ecology Resources* 2009, **8**(Suppl 1):1 - 26.**
31. Prado B, Pozo C, Valdez-Moreno M, Hebert P: **Beyond the colours: discovering hidden diversity in the Nymphalidae of the Yucatan Peninsula in Mexico through DNA barcoding. *Plos One* 2011, **6**(11):e27776.**
32. Gompert Z, Nice CC, Fordyce JA, Forister ML, Shapiro AM: **Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. *Molecular Ecology* 2006, **15**(7):1759-1768.**
33. Meier R, Shiyang K, Vaidya G, Ng P: **DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology* 2006, **55**(5):715 - 728.**
34. Will K, Mishler B, Wheeler Q: **The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology* 2005, **54**(5):844 - 851.**
35. Whitworth TL, Dawson RD, Magalon H, Baudry E: **DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae).** *Proceedings of the Royal Society B: Biological Sciences* 2007, **274**(1619):1731-1739.
36. Wiemers M, Fiedler K: **Does the DNA barcoding gap exist? - a case study in blue butterflies (Lepidoptera: Lycaenidae).** *Frontiers in Zoology* 2007, **4**(1):8.
37. Galtier N, Nabholz B, GléMin S, Hurst GDD: **Mitochondrial DNA as a marker of molecular diversity: a reappraisal.** *Molecular Ecology* 2009, **18**(22):4541-4550.
38. Hendrich L, Pons J, Ribera I, Balke M: **Mitochondrial Cox1 Sequence Data Reliably Uncover Patterns of Insect Diversity But Suffer from High Lineage-Idiosyncratic Error Rates.** *Plos One* 2010, **5**(12):e14448.
39. Hickerson M, Meyer C, Moritz C: **DNA barcoding will often fail to discover new animal species over broad parameter space.** *Systematic Biology* 2006, **55**(5):729 - 739.
40. Lohse K: **Can mtDNA Barcodes Be Used to Delimit Species? A Response to Pons et al. (2006).** *Systematic Biology* 2009, **58**(4):439-442.
41. Meyer CP, Paulay G: **DNA Barcoding: Error Rates Based on Comprehensive Sampling.** *PLoS Biol* 2005, **3**(12):e422.
42. Moritz C, Cicero C: **DNA barcoding: promise and pitfalls.** *PLoS Biol* 2004, **2**(10):e354.
43. van Velzen R, Weitschek E, Felici G, Bakker FT: **DNA Barcoding of Recently Diverged Species: Relative Performance of Matching Methods.** *Plos One* 2012, **7**(1):e30490.
44. Ebach MCHC: **DNA barcoding is no substitute for taxonomy.** *Nature* 2005, **434**(7034):697-697.
45. Rubinoff D, Cameron S, Will K: **A Genomic Perspective on the Shortcomings of Mitochondrial DNA for "Barcoding" Identification.** *Journal of Heredity* 2006, **97**(6):581-594.
46. Will KW, Rubinoff D: **Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification.** *Cladistics* 2004, **20**(1):47-55.
47. Dupuis JR, Roe AD, Sperling FAH: **Multi-locus species delimitation in closely related animals and fungi: one marker is not enough.** *Molecular Ecology* 2012, **21**(18):4422-4436.

48. Padial J, Miralles A, De la Riva I, Vences M: **The integrative future of taxonomy.** *Frontiers in Zoology* 2010, **7**(1):16.
49. Gebiola M, GÓmez-Zurita J, Monti MM, Navone P, Bernardo U: **Integration of molecular, ecological, morphological and endosymbiont data for species delimitation within the *Pnigalio soemius* complex (Hymenoptera: Eulophidae).** *Molecular Ecology* 2012, **21**(5):1190-1208.
50. Dayrat B: **Toward integrative taxonomy.** *Biological Journal of the Linnean Society* 2005, **85**:407 - 415.
51. Rougerie R, Naumann S, Nässig WA: **Morphology and Molecules Reveal Unexpected Cryptic Diversity in the Enigmatic Genus *Sinobirma* Bryk, 1944 (Lepidoptera: Saturniidae).** *Plos One* 2012, **7**(9):e43920.
52. Padial JM, De La Riva I: **Integrative taxonomy reveals cryptic Amazonian species of *Pristimantis* (Anura: Strabomantidae).** *Zoological Journal of the Linnean Society* 2009, **155**(1):97-122.
53. Lumley L, Sperling F: **Integrating morphology and mitochondrial DNA for species delimitation within the spruce budworm (*Choristoneura fumiferana*) cryptic species complex (Lepidoptera: Tortricidae).** *Systematic Entomology* 2010, **35** (3):416-428.
54. Lumley LM, Sperling FAH: **Utility of microsatellites and mitochondrial DNA for species delimitation in the spruce budworm (*Choristoneura fumiferana*) species complex (Lepidoptera: Tortricidae).** *Molecular Phylogenetics and Evolution* 2011, **58**(2):232-243.
55. Roe AD, Sperling FAH: **Population structure and species boundary delimitation of cryptic *Dioryctria* moths: an integrative approach.** *Molecular Ecology* 2007, **16**(17):3617-3633.
56. Ence DD, Carstens BC: **SpedeSTEM: a rapid and accurate method for species delimitation.** *Molecular Ecology Resources* 2011, **11**(3):473-480.
57. O'Meara BC, Ané C, Sanderson MJ, Wainwright PC: **Testing for different rates of continuous trait evolution using likelihood.** *Evolution* 2006, **60**(5):922-933.
58. Pons J, Barraclough T, Gomez-Zurita J, Cardoso A, Duran D, Hazell S, Kamoun S, Sumlin W, Vogler A: **Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects.** *Systematic Biology* 2006, **55**(4):595-609.
59. Yang Z, Rannala B: **Bayesian species delimitation using multilocus sequence data.** *Proceedings of the National Academy of Sciences U S A* 2010, **107**(20):9264 - 9269.
60. Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C: **Coalescent-based species delimitation in an integrative taxonomy.** *Trends in Ecology & Evolution (Personal edition)* 2012, **27**(9):480-488.
61. Lim GS, Balke M, Meier R: **Determining Species Boundaries in a World Full of Rarity: Singletons, Species Delimitation Methods.** *Systematic Biology* 2012, **61**(1):165-169.
62. Sauer J, Hausdorf B: **A comparison of DNA-based methods for delimiting species in a Cretan land snail radiation reveals shortcomings of exclusively molecular taxonomy.** *Cladistics* 2012, **28**(3):300-316.

63. Huang D, Meier R, Todd P, Chou L: **Slow Mitochondrial COI Sequence Evolution at the Base of the Metazoan Tree and Its Implications for DNA Barcoding.** *Journal of Molecular Evolution* 2008, **66**(2):167-174.
64. Li Y, Zhou XIN, Feng GUI, Hu H, Niu L, Hebert PDN, Huang D: **COI and ITS2 sequences delimit species, reveal cryptic taxa and host specificity of fig-associated Sycophila (Hymenoptera, Eurytomidae).** *Molecular Ecology Resources* 2010, **10**(1):31-40.
65. Nagy ZT, Sonet G, Glaw F, Vences M: **First Large-Scale DNA Barcoding Assessment of Reptiles in the Biodiversity Hotspot of Madagascar, Based on Newly Designed COI Primers.** *Plos One* 2012, **7**(3):e34506.
66. Hebert P, Penton E, Burns J, Janzen D, Hallwachs W: **Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*.** *Proceedings of the National Academy of Sciences U S A* 2004, **101**(41):14812 - 14817.
67. Hebert P, Ratnasingham S, deWaard J: **Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species.** *Proceedings of the Royal Society B: Biological Sciences* 2003, **270** Suppl 1:96-99.
68. Smith M, Woodley N, Janzen D, Hallwachs W, Hebert P: **DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae).** *Proceedings of the National Academy of Sciences U S A* 2006, **103**(10):3657 - 3662.
69. Hajibabaei M, Janzen D, Burns J, Hallwachs W, Hebert P: **DNA barcodes distinguish species of tropical Lepidoptera.** *Proceedings of the National Academy of Sciences U S A* 2006, **103**(4):968 - 971.
70. Fontaneto D, Herniou EA, Boschetti C, Caprioli M, Melone G, Ricci C, Barraclough TG: **Independently Evolving Species in Asexual Bdelloid Rotifers.** *PLoS Biology* 2007, **5**(4):e87.
71. Hurst GDD, Jiggins FM: **Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts.** *Proceedings of the Royal Society B: Biological Sciences* 2005, **272**(1572):1525-1534.
72. Meier R, Zhang G, Ali F: **The Use of Mean Instead of Smallest Interspecific Distances Exaggerates the Size of the "Barcoding Gap" and Leads to Misidentification.** *Systematic Biology* 2008, **57**(5):809-813.
73. Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP: **Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects.** *Systematic Biology* 2006, **55**(4):595-609.
74. Hambäck P, Weingartner E, Ericson L, Fors L, Cassel-Lundhagen A, Stenberg J, Bergsten J: **Bayesian species delimitation reveals generalist and specialist parasitic wasps on *Galerucella* beetles (Chrysomelidae): sorting by herbivore or plant host.** *BMC Evolutionary Biology* 2013, **13**(1):92.
75. de Queiroz K: **Species concepts and species delimitation.** *Syst Biol* 2007, **56**:879-886.
76. Rannala B, Yang Z: **Bayes Estimation of Species Divergence Times and Ancestral Population Sizes Using DNA Sequences From Multiple Loci.** *Genetics* 2003, **164**(4):1645-1656.

77. Knowles L, Carstens B: **Delimiting species without monophyletic gene trees.** *Systematic Biology* 2007, **56**:887-895.
78. Yang Z, Rannala B: **Bayesian species delimitation using multilocus sequence data.** *Proceedings of the National Academy of Sciences* 2010, **107**(20):9264-9269.
79. Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C: **Coalescent-based species delimitation in an integrative taxonomy.** *Trends in Ecology & Evolution* 2012, **27**(9):480-488.
80. Astrin JJ, Stüben PE, Misof B, Wägele JW, Gimnich F, Raupach MJ, Ahrens D: **Exploring diversity in cryptorhynchine weevils (Coleoptera) using distance-, character- and tree-based species delineation.** *Molecular Phylogenetics and Evolution* 2012, **63**(1):1-14.
81. Bryson RW, Savary WE, Prendini L: **Biogeography of scorpions in the *Pseudouroctonus minimus* complex (Vaejovidae) from south-western North America: implications of ecological specialization for pre-Quaternary diversification.** *Journal of Biogeography* 2013.
82. Fernández-Flores S, Fernández-Triana JL, Martínez JJ, Zaldívar-Riverón A: **DNA barcoding species inventory of *Microgastrinae* wasps (Hymenoptera, Braconidae) from a Mexican tropical dry forest.** *Molecular Ecology Resources* 2013.
83. Fontaneto D, Iakovenko N, Eyres I, Kaya M, Wyman M, Barraclough T: **Cryptic diversity in the genus *Adineta* Hudson & Gosse, 1886 (Rotifera: Bdelloidea: Adinetidae): a DNA taxonomy approach.** *Hydrobiologia* 2011, **662**(1):27-33.
84. Kergoat GJ, Le Ru BP, Genson G, Cruaud C, Couloux A, Delobel A: **Phylogenetics, species boundaries and timing of resource tracking in a highly specialized group of seed beetles (Coleoptera: Chrysomelidae: Bruchinae).** *Molecular Phylogenetics and Evolution* 2011, **59**(3):746-760.
85. McBride CS, Van Velzen R, Larsen TB: **Allopatric origin of cryptic butterfly species that were discovered feeding on distinct host plants in sympatry.** *Molecular Ecology* 2009, **18**:3639-3651.
86. Nuñez JJ, Vejar-Pardo A, Guzmán BE, Barriga EH, Gallardo CS: **Phylogenetic and mixed Yule-coalescent analyses reveal cryptic lineages within two South American marine snails of the genus *Crepidatella* (Gastropoda: Calyptraeidae).** *Invertebrate Biology* 2012, **131**(4):301-311.
87. Pereira-da-Conceicao L, Price B, Barber-James H, Barker N, de Moor F, Villet M: **Cryptic variation in an ecological indicator organism: mitochondrial and nuclear DNA sequence data confirm distinct lineages of *Baetis harrisoni* Barnard (Ephemeroptera: Baetidae) in southern Africa.** *BMC Evolutionary Biology* 2012, **12**(1):1-14.
88. Vuataz L, Sartori M, Wagner A, Monaghan MT: **Toward a DNA Taxonomy of Alpine *Rhithrogena* (Ephemeroptera: Heptageniidae) Using a Mixed Yule-Coalescent Analysis of Mitochondrial and Nuclear DNA.** *Plos One* 2011, **6**(5):e19728.
89. Monaghan MT, Wild R, Elliot M, Fujisawa T, Balke M, Inward DJG, Lees DC, Ranaivosolo R, Eggleton P, Barraclough TG *et al*: **Accelerated Species Inventory on Madagascar**

- Using Coalescent-Based Models of Species Delineation. *Systematic Biology* 2009, **58**(3):298-311.**
90. Reid N, Carstens B: **Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent model.** *BMC Evolutionary Biology* 2012, **12**(1):196.
91. Fujisawa T, Barraclough TG: **Delimiting Species Using Single-locus Data and the Generalized Mixed Yule Coalescent (GMYC) Approach: A Revised Method and Evaluation on Simulated Datasets.** *Systematic Biology* 2013.
92. Meyer C, Paulay G: **DNA barcoding: error rates based on comprehensive sampling.** *PLoS Biology* 2005, **3**(12):e422.
93. Williams ST, Knowlton N: **Mitochondrial Pseudogenes Are Pervasive and Often Insidious in the Snapping Shrimp Genus *Alpheus*.** *Molecular Biology and Evolution* 2001, **18**(8):1484-1493.
94. Petit R, Excoffier L: **Gene flow and species delimitation.** *Trends in Ecology & Evolution* 2009, **24**:386 - 393.
95. Lemmon AR, Moriarty EC: **The Importance of Proper Model Assumption in Bayesian Phylogenetics.** *Systematic Biology* 2004, **53**(2):265-277.
96. Phillips MJ: **Branch-length estimation bias misleads molecular dating for a vertebrate mitochondrial phylogeny.** *Gene* 2009, **441**(1-2):132-140.
97. Lemmon AR, Brown JM, Stanger-Hall K, Lemmon EM: **The Effect of Ambiguous Data on Phylogenetic Estimates Obtained by Maximum Likelihood and Bayesian Inference.** *Systematic Biology* 2009, **58**(1):130-145.
98. Ekman S, Blaalid R: **The Devil in the Details: Interactions between the Branch-Length Prior and Likelihood Model Affect Node Support and Branch Lengths in the Phylogeny of the Psoraceae.** *Systematic Biology* 2011, **60**(4):541-561.
99. Zhang C, Rannala B, Yang Z: **Robustness of Compound Dirichlet Priors for Bayesian Inference of Branch Lengths.** *Systematic Biology* 2012, **61**(5):779-784.
100. Brown JM, Hedtke SM, Lemmon AR, Lemmon EM: **When Trees Grow Too Long: Investigating the Causes of Highly Inaccurate Bayesian Branch-Length Estimates.** *Systematic Biology* 2010, **59**(2):145-161.
101. Brown JM, Lemmon AR: **The Importance of Data Partitioning and the Utility of Bayes Factors in Bayesian Phylogenetics.** *Systematic Biology* 2007, **56**(4):643-655.
102. Brandley MC, Schmitz A, Reeder TW: **Partitioned Bayesian Analyses, Partition Choice, and the Phylogenetic Relationships of Scincid Lizards.** *Systematic Biology* 2005, **54**(3):373-390.
103. Marshall DC, Simon C, Buckley TR: **Accurate Branch Length Estimation in Partitioned Bayesian Analyses Requires Accommodation of Among-Partition Rate Variation and Attention to Branch Length Priors.** *Systematic Biology* 2006, **55**(6):993-1003.
104. Pogue MG: **World revision of the genus *Spodoptera* Guenée (Lepidoptera: Noctuidae).** *Memoirs of the American Entomological Society* 2002, **43**:1-202.

105. Monti L, Générumont J, Malosse C, Lalanne-Cassou B: **A genetic analysis of some components of reproductive isolation between two closely related species, *Spodoptera latifascia* (Walker) and *S. descoinsi* (Lalanne-Cassou and Silvain) (Lepidoptera: Noctuidae).** *Journal of Evolutionary Biology* 1997, **10**(1):121-134.
106. Groot AT, Marr M, Schöfl G, Lorenz S, Svatos A, Heckel DG: **Host strain specific sex pheromone variation in *Spodoptera frugiperda*.** *Frontiers in Zoology* 2008, **5**(1):20.
107. Schöfl G, Heckel DG, Groot AT: **Time-shifted reproductive behaviours among fall armyworm (Noctuidae: *Spodoptera frugiperda*) host strains: evidence for differing modes of inheritance.** *Journal of Evolutionary Biology* 2009, **22**:1447–1459.
108. Brown ES, Dewhurst CF: **The genus *Spodoptera* (Lepidoptera: Noctuidae) in Africa and the Near East.** *Bulletin of Entomological Research* 1975, **65**:221–262.
109. Todd EL, Poole RW: **Keys and illustrations for the armyworm moths of the noctuid genus *Spodoptera* Guenée from the Western Hemisphere.** *Annals of the Entomological Society of America* 1980, **73**:722–738.
110. Kergoat GJ, Prowell DP, Le Ru BP, Mitchell A, Dumas P, Clamens A-L, Condamine FL, Silvain J-F: **Disentangling dispersal, vicariance and adaptive radiation patterns: A case study using armyworms in the pest genus *Spodoptera* (Lepidoptera: Noctuidae).** *Molecular Phylogenetics and Evolution* 2012, **65**(3):855-870.
111. Martinelli S, Barata RM, Zucchi MI, Silva-Filho MDC, Omoto C: **Molecular variability of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) populations associated to maize and cotton crops in Brazil.** *Journal of Economic Entomology* 2006, **99**(2):519-526.
112. Salinas-Hernandez H, Saldamando-Benjumea C: **Haplotype identification within *Spodoptera frugiperda* (J.E. Smith) (Lepidoptera: Noctuidae) corn and rice strains from Colombia.** *Neotropical Entomology* 2011, **40**:421-430.
113. Silvain J-F, Lalanne-Cassou B: **Distinction entre *Spodoptera latifascia* (Walker) et *Spodoptera cosmioides* (Walker), bona species [Lepidoptera, Noctuidae].** *Revue française d'entomologie* 1997, **19**:95-97.
114. Barros EM, Torres JB, Ruberson JR, Oliveira MD: **Development of *Spodoptera frugiperda* on different hosts and damage to reproductive structures in cotton.** *Entomologia Experimentalis et Applicata* 2010, **137**(3):237-245.
115. Greenberg SM, Sappington TW, Legaspi BC, Liu TX, Sétamou M: **Feeding and Life History of *Spodoptera exigua* (Lepidoptera: Noctuidae) on Different Host Plants.** *Annals of the Entomological Society of America* 2001, **94**(4):566-575.
116. Prowell DP, McMichael M, Silvain J-F: **Multilocus genetic analysis of host use, introgression, and speciation in host strains of fall armyworm (Lepidoptera: Noctuidae).** *Annals of the Entomological Society of America* 2004, **97**(5):1034-1044
117. Drès M, Mallet J: **Host races in plant-feeding insects and their importance in sympatric speciation.** *The Royal Society* 2002, **357**:471-492.
118. Bolnick DI, Fitzpatrick BM: **Sympatric Speciation: Models and Empirical**

- Evidence.** *Annual Review of Ecology, Evolution, and Systematics* 2007, **38**(1):459-487.
119. Pashley DP: **Host-associated genetic differentiation in fall armyworm (Lepidoptera: Noctuidae): a sibling species complex?** *Annals of the Entomological Society of America* 1986, **79**:898-904.
120. Meagher R, Nagoshi R, Stuhl C: **Oviposition Choice of Two Fall Armyworm (Lepidoptera: Noctuidae) Host Strains.** *Journal of Insect Behavior* 2011:1-11.
121. Levy HC, Garcia-Maruniak A, Maruniak JE: **Strain identification of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) insects and cell line: PCR-RFLP of cytochrome oxidase C subunit I gene.** *The Florida Entomologist* 2002, **85**(1):186-190.
122. Lu YJ, Adang MJ: **Distinguishing fall armyworm (Lepidoptera: Noctuidae) strain using a diagnostic mitochondrial DNA marker.** *Florida Entomologist* 1996, **79**(1):48-55.
123. McMichael M, Pashley Prowell DP: **Differences in amplified fragment-length polymorphisms in fall armyworm (Lepidoptera: Noctuidae) host strain.** *Annals of the Entomological Society of America* 1999, **92**(2):175-181.
124. Meagher RL, Gallo-Meagher M: **Identifying host strains of fall armyworm (Lepidoptera : Noctuidae) in Florida using mitochondrial markers.** *Florida Entomologist* 2003, **86**(4):450-455.
125. Nagoshi RN, Armstrong JS, Silvie P, Meagher RL: **Structure and distribution of a strain-biased tandem repeat element in Fall armyworm (Lepidoptera: Noctuidae) populations in Florida, Texas, and Brazil.** *Annals of the Entomological Society of America* 2008, **101**(6):1112-1120
126. Machado V, Wunder M, Baldissera VD, Oliveira JV, Fiúza LM, Nagoshi RN: **Molecular Characterization of Host Strains of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in Southern Brazil.** *Annals of the Entomological Society of America* 2008, **101**(3):619-626.
127. Pair SD, Raulston JR, Sparks AN, Westbrook JK, Douce GK: **Fall Armyworm Distribution and Population Dynamics in the Southeastern States.** *The Florida Entomologist* 1986, **69**(3):468-487.
128. Nagoshi RN, Meagher RL, Adamczyk JJ, Braman SK, Brandenburg RL, Nuessly G: **New restriction fragment length polymorphisms in the cytochrome oxidase I gene facilitate host strain identification of Fall armyworm (Lepidoptera: Noctuidae) populations in the southeastern United States.** *Journal of Economic Entomology* 2006, **99**(3):671-677.
129. Mitchell A, Mitter C, Regier JC: **Systematics and evolution of the cutworm moths (Lepidoptera: Noctuidae): evidence from two protein-coding nuclear genes.** *Systematic Entomology* 2006, **31**(1):21-46.
130. Drummond AJ, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson A: **Geneious v3.8.** . 2007.
131. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al*: **ClustalW and ClustalX version 2.** *Bioinformatics* 2007 doi:10.1093/bioinformatics/btm404 **23**(21):2947-2948.

132. Maddison WPaDRM: **Mesquite: a modular system for evolutionary analysis. Version 2.0** <http://mesquiteproject.org>. 2007.
133. Ronquist F, Huelsenbeck, J.P.: **MRBAYES 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003., **19**:1572–1574.
134. Nylander JAA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey J: **Bayesian Phylogenetic Analysis of Combined Data.** *Systematic Biology* 2004, **53**(1):47-67.
135. Darriba D TG, Doallo R, Posada D: **jModelTest 2: more models, new heuristics and parallel computing.** *Nature Methods* 2012, **9**(8):772.
136. Rannala B, Zhu T, Yang Z: **Tail Paradox, Partial Identifiability, and Influential Priors in Bayesian Branch Length Inference.** *Molecular Biology and Evolution* 2012, **29**(1):325-335.
137. Rambaut A, Drummond AJ: **Tracer v1.4, Available from** <http://beast.bio.ed.ac.uk/Tracer> 2007.
138. Britton T, Anderson CL, Jacquet D, Lundqvist S, Bremer K: **Estimating Divergence Times in Large Phylogenetic Trees.** *Systematic Biology* 2007, **56**(5):741-752.
139. Ho SYW, Saarma U, Barnett R, Haile J, Shapiro B: **The Effect of Inappropriate Calibration: Three Case Studies in Molecular Ecology.** *Plos One* 2008, **3**(2):e1615.
140. Monaghan M, Balke M, Gregory T, Vogler A: **DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers.** *Philosophical Transactions of the Royal Society B: Biological Sciences* 2005, **360**(1462):1925 - 1933.
141. Erixon P, Svennblad B, Britton T, Oxelman B: **Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics.** *Systematic Biology* 2003, **52**:665–673.
142. Lima ER, McNeil JN: **Female sex pheromones in the host races and hybrids of the fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae).** *Chemoecology* 2009, **19**:29-36.
143. Schöfl G, Dill A, Heckel DG, Groot AT: **Allochronic Separation versus Mate Choice: Nonrandom Patterns of Mating between Fall Armyworm Host Strains.** *The American Naturalist* 2011, **177**(4):470-485.
144. Whitford F, Quisenberry SS, Riley TJ, JW L: **Oviposition preference, mating compatibility, and development of two fall armyworm strains.** *The Florida Entomologist* 1988, **71**(3):234-243.
145. Pashley Prowell DP, Martin JA: **Reproductive incompatibility between host strains of the Fall armyworm (Lepidoptera: Noctuidae).** *Annals of the Entomological Society of America* 1987:731-733.
146. Quisenberry SS: **Fall armyworm (Lepidoptera: Noctuidae) host strain reproductive compatibility.** *Florida Entomologist* 1991, **74**(2):194-199.
147. Groot AT, M. M, Heckel DG, Schöfl G: **The roles and interactions of reproductive isolation mechanisms in fall armyworm (Lepidoptera: Noctuidae) host strains.** *Ecological Entomology* 2010, **35**(1):105-118.
148. Velásquez-Vélez MI, Saldamando-Benjumea CI, Ríos-Diez JD: **Reproductive Isolation between Two Populations of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) Collected in Corn and Rice Fields from Central Colombia.** *Annals of the*

- Entomological Society of America*
2011, **104**(4):826-833.
149. Nagoshi RN, Fleischer S, Meagher RL:
**Texas Is the Overwintering Source of
Fall Armyworm in Central
Pennsylvania: Implications for**

**Migration Into the Northeastern
United States.** *Environmental
Entomology* 2009, **38**(6):1546-1554

1.3. Différenciation génétique et structuration des populations des deux variants de *S. frugiperda*

Article 3 en préparation (version préliminaire) : Génétique et structuration des populations des deux variants de *Spodoptera frugiperda* (Noctuidae, Lepidoptera). Il est envisagé, par la suite, de compléter l'échantillonnage existant pour avoir une couverture plus dense pour le variant maïs, de façon à tester si il y a une structuration des populations pour ce variant.

Génétique et structuration des populations des deux variants de *Spodoptera frugiperda* (Noctuidae, Lepidoptera)

P. Dumas, A.-L. Clamens, L. Soldati, E. d'Alençon et G. J. Kergoat

Introduction

Le genre *Spodoptera* regroupe une trentaine d'espèces mondialement réparties et dont la moitié a un statut de bio agresseur (Pogue 2002). *Spodoptera frugiperda* (J.E. Smith) (Lepidoptera : Noctuidae), encore appelée *Fall armyworm* en anglais (FAW), représente l'une des plus ravageuses de l'hémisphère Ouest. *Spodoptera frugiperda* est une espèce hautement polyphage pouvant s'attaquer à plus de 98 espèces au sein de 27 familles botaniques différentes (Capinera 1999; Pogue 2002). La présence de cette espèce a été documentée dans toutes les Amériques, de l'Argentine au sud du Canada en passant par l'Amérique Centrale ainsi que dans les Caraïbes (Juárez *et al.* 2012; Nagoshi *et al.* 2009; Nagoshi *et al.* 2012b; Nagoshi *et al.* 2010; Prowell *et al.* 2004; Saldamando & Vélez-Arango 2010). L'étendue géographique de son aire de répartition est directement liée à sa capacité de dispersion. En effet le cycle biologique de *S. frugiperda* n'ayant pas d'étape de diapause, les individus ne sont pas capables de résister à des températures négatives et s'adaptent donc en migrant vers des régions plus clémentes et vers de nouvelles ressources, grâce à leur capacité de vol à longue distance (Mitchell *et al.* 1991; Pair *et al.* 1986; Sparks 1979).

Deux variants sont décrits pour *S. frugiperda* : le variant maïs et le variant riz (Pashley 1986; Pashley *et al.* 1985). Les larves appartenant au variant maïs sont préférentiellement retrouvées sur maïs (*Zea mays*), sorgho (*Sorghum bicolor*) et coton (*Gossypium hirsutum*) alors que celles du variants riz sont connues pour infester préférentiellement le riz (*Oryza sativa*), le soja (*Glycine max*) et différentes plantes herbacées comme le chiendent (*Cynodon dactylon*). Les deux variants sont morphologiquement identiques mais génétiquement différenciés. En effet, plusieurs marqueurs moléculaires polymorphes que ce soit sur l'ADN mitochondrial (Levy *et al.* 2002; Lu & Adang 1996; Meagher & Gallo-Meagher 2003; Nagoshi *et al.* 2006; Pashley 1989; Pashley & Ke 1992) ou nucléaire (McMichael & Pashley Prowell 1999; Nagoshi & Meagher 2003; Prowell *et al.* 2004) ont été caractérisés entre les deux

variants. La présence en sympatrie des deux variants a été mise en évidence grâce à ces différents marqueurs, dans plusieurs régions en Amérique du Sud (Juárez *et al.* 2012; Machado *et al.* 2008; Martinelli *et al.* 2006; Nagoshi *et al.* 2012a; Saldamando & Vélez-Arango 2010) et Amérique du Nord (Nagoshi *et al.* 2009; Nagoshi *et al.* 2008; Nagoshi *et al.* 2012b; Prowell *et al.* 2004) ainsi que dans les Caraïbes (Nagoshi *et al.* 2010; Prowell *et al.* 2004). De plus, la présence de mécanismes d'isolement reproducteur a été identifiée entre les deux variants impliquant des barrières de différentes natures. Plusieurs barrières pré-zygotiques ont été décrites, incluant la préférence de plante hôte lors de l'oviposition (Schöfl *et al.* 2011; Whitford F *et al.* 1988), un allochronisme durant la nuit des périodes d'accouplement (Schöfl *et al.* 2009) ainsi que des différences en proportions des composés présents dans les phéromones sexuelles émises par les femelles (Groot *et al.* 2008). Les études sur la présence de barrières post-zygotiques entre les deux variants sont plus controversées. Certaines mettent en évidence un sens d'accouplement unidirectionnel (Pashley Prowell & Martin 1987; Velásquez-Vélez *et al.* 2011) alors que d'autres montrent l'inverse (Nagoshi & Meagher 2004; Quisenberry 1991; Whitford F *et al.* 1988). De plus en *back cross* les femelles hybrides ne s'accouplent que difficilement avec leur frère et pas du tout avec tout autre mâle (Groot *et al.* 2010; Pashley Prowell & Martin 1987; Whitford *et al.* 1988). Enfin des études en champs ont montré un fort taux d'hybridation dans la nature comme au Brésil (14%) ou aux Etats-Unis et Amérique Centrale (16%) (Nagoshi *et al.* 2009; Prowell *et al.* 2004) ou encore dans les Caraïbes (Nagoshi *et al.* 2010), grâce à la combinaison de marqueurs mitochondriaux et nucléaires.

Plusieurs études menées sur des populations naturelles de FAW mettent en évidence une différenciation génétique importante en fonction de la plante hôte (Busato *et al.* 2004; Clark *et al.* 2006; Martinelli *et al.* 2006). Cependant jusqu'à présent très peu d'études (Saldamando & Vélez-Arango 2010) ont été réalisées dans le but de regarder le niveau de différenciation et de structuration des populations des deux variants de *S. frugiperda* et aucune sur l'ensemble de son aire de répartition. En effet l'étude de Saldamando & Vélez-Arango (2010) a montré une différenciation génétique élevée entre les populations des deux variants échantillonnées en Colombie. Dans cette étude nous proposons de regarder, grâce à l'étude d'un fragment du gène mitochondrial COI (cytochrome oxydas I), le niveau de

structuration et de différenciation de plusieurs populations de *S. frugiperda* variant riz et variant maïs prises sur l'ensemble de leur aire de répartition.

Méthode

Constitution du jeu de données

Un total de 182 séquences de *Spodoptera frugiperda* a été utilisé dans cette étude. Le jeu de données rassemble 62 séquences précédemment utilisées dans les études récentes (Kergoat *et al.* 2012); Dumas *et al.* soumis), 60 séquences récupérées sur la base de données de séquences internationales GenBank et de 60 nouveaux spécimens récoltés sur le terrain (Tableau S1). Toutes ces séquences ont été sélectionnées afin d'avoir la meilleure couverture géographique possible pour les deux variants de *S. frugiperda* (Figure 1).

Les 60 nouveaux échantillons - récoltés lors de deux missions de terrain en Guadeloupe en 2013- ont été capturés dans quatre localités distinctes par piégeage phéromonal ou directement prélevés sur plante hôte et mis dans de l'éthanol à 96 % pour être conservés à -20°C. Leurs ADN génomiques ont été obtenus en suivant le protocole du kit d'extraction DNeasy® Blood and Tissue Kit (Qiagen). Deux fragments du gène mitochondrial codant pour la sous unité I du Cytochrome Oxydase (COI) ont été amplifiés grâce à des amorces utilisées dans d'autres études et suivant un protocole de *polymerase chain reaction* « PCR » précédemment utilisé (Dumas *et al.* en préparation). Les produits PCR ont été ensuite séquencés par le centre Génoscope qui utilise le kit de séquençage BigDye v3.1 et le séquenceur Applied 3730xl. Les séquences ont enfin été vérifiées et nettoyées grâce au programme Geneious v3.8 (Drummond *et al.* 2007).

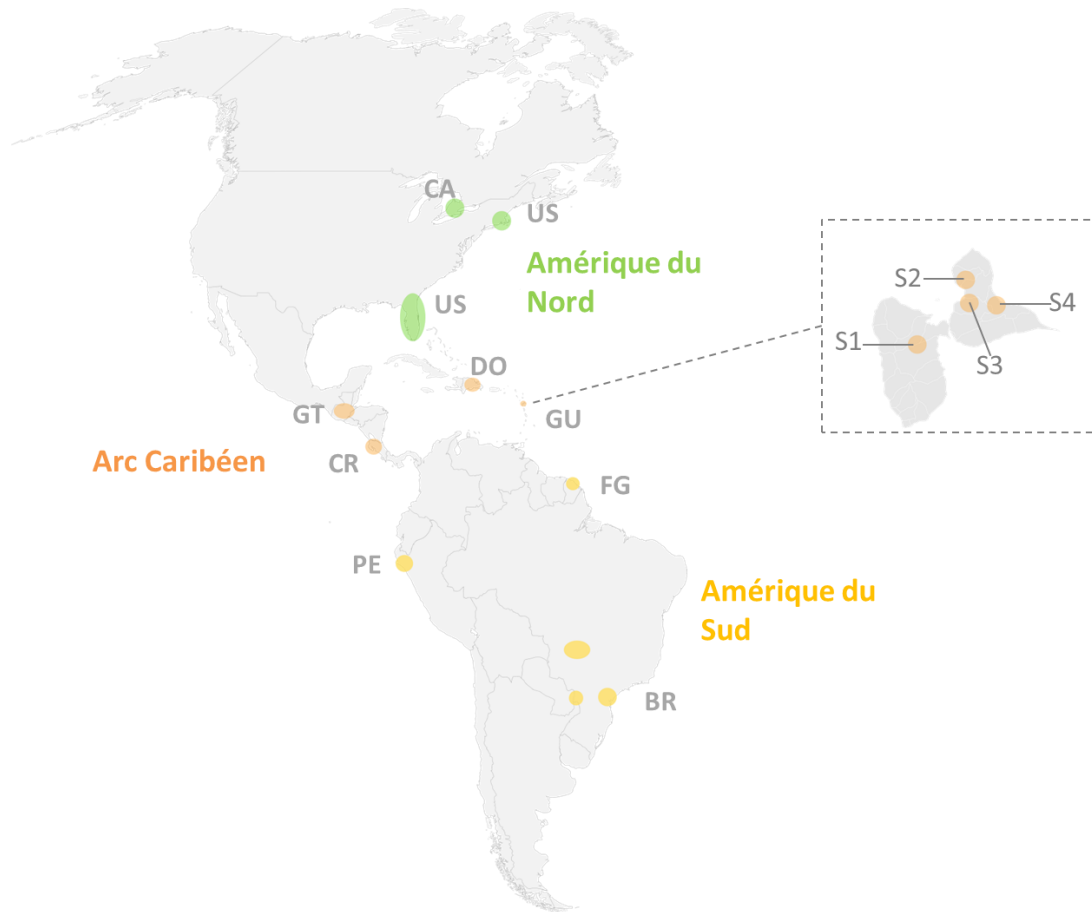


Figure 1 : Carte du continent américain avec la localisation des échantillons utilisés pour les analyses de différenciation des populations de *S. frugiperda*. En Guadeloupe quatre localités différentes ont été échantillonnées (S1-S4).

L'ensemble des 182 séquences nucléotidiques ont été alignées grâce au programme ClustalW version 2 (Larkin *et al.* 2007 doi:10.1093/bioinformatics/btm404) et le programme Mesquite (Maddison& Maddison 2009) a servi à tronquer les séquences pour garder la section du fragment COI pour lequel on a 100% de recouvrement pour tous les différents échantillons (433pb).

Identification des variants riz et maïs au sein des nouvelles séquences

Les nouveaux échantillons ainsi que les séquences prises sur GenBank ont été analysés pour savoir à quel variant de *S. frugiperda* ils appartenaient. Une analyse phylogénétique a donc été réalisée en utilisant la phylogénie du genre *Spodoptera* (Kergoat *et al.* 2012). Les nouvelles séquences ont été ajoutées au jeu de données pour être ré-analysées en inférence Bayésienne à l'aide du logiciel MRBAYES version 3.2.1.(Ronquist

&Huelsenbeck 2013) en utilisant le modèle de substitution (GTR +G) sélectionné par le programme jModelTest 2.0.1 4 (Darriba 2012). Dix millions de générations avec huit chaîne de Markov indépendantes ont été utilisées. Toutes les 100 générations, un échantillonnage des arbres issus de chaque chaîne de Markov est réalisé. Tous les arbres ainsi échantillonnés serviront à estimer l'arbre consensus qui donnera simultanément une estimation de la topologie ainsi que les valeurs de soutien des nœuds. Afin de limiter les biais liés à une mauvaise estimation des paramètres en début d'analyse, ce consensus est effectué à partir des topologies issues de la phase dite stationnaire. Pour cela on ne considère pas les arbres générés par les toutes premières 2.5 millions de générations. Cette phase dite de *burn-in* est aussi appliquée après avoir testé la stabilité et la bonne convergence des différentes analyses grâce au programme Tracer (Rambaut & Drummond 2007).

Analyses statistiques de la structuration des populations

Les séquences ont été analysées en divisant le jeu de données en différents grands groupes potentiels. Dans un premier temps, la structuration potentielle entre l'ensemble des individus assignés au variant maïs et l'ensemble des individus assignés au variant riz a été testée. Puis pour chacun des deux variants trois grands groupes potentiels ont été définis : l'ensemble des individus échantillonnés en Amérique du Nord (Canada et USA), l'ensemble des individus échantillonnés dans l'arc Caraïbéen (Guadeloupe, République Dominicaine et Costa Rica) et l'ensemble des individus échantillonnés en Amérique du Sud (Brésil, Pérou et Guyane Française). Enfin, quatre groupes ont été définis pour les individus échantillonnés dans quatre localités distinctes en Guadeloupe. La différenciation génétique entre ces différents groupes a été estimée en calculant avec le programme DNASP version 4.10.8 (Rozas *et al.* 2003), les indices de F_{st} , de K_{st}^* (Hudson *et al.* 1992) et de S_{nn} (Hudson 2000). La significativité des différents indices a été évaluée en réalisant un test de permutation (1000 répliques) présent dans ce programme. L'estimation de ces trois indices de façon complémentaire est souhaitable car ils permettent de prendre en compte des biais distincts : l'indice de F_{st} est plus adéquat lorsque le niveau de diversité est faible (Nagylaki 1998) ; le K_{st}^* est défini comme un test puissant en cas de taille d'échantillons faible et le S_{nn} est requis en cas d'hétérogénéité des tailles d'échantillons et des niveaux de variation (Hudson 2000). La diversité nucléotidique (π) et haplotypique (H_d) ainsi que le nombre d'haplotypes et de sites polymorphes ont également été estimés grâce au programme DNASP version

4.10.8. Enfin, le logiciel Network 4.6.1.1 a été utilisé pour estimer en parcimonie le réseau de relations entre les différents haplotypes.

Résultats

Structure et diversité génétique

Le jeu de données final est constitué de 143 séquences d'une longueur de 433 paires de bases (pb) pour lequel il y a 100% de recouvrement. Le jeu de données ne présente aucune donnée ambiguë ni d'*indels* dans les séquences. Un total de 34 sites polymorphes a été identifié pour ce fragment pour l'ensemble des 143 séquences, divisées en 12 haplotypes différents (Figure 2). Le nombre d'haplotypes, de sites polymorphes, ainsi que l'estimation de la diversité nucléotidique (P_i) et haplotypique (H_d) sont présentés dans le tableau 1.

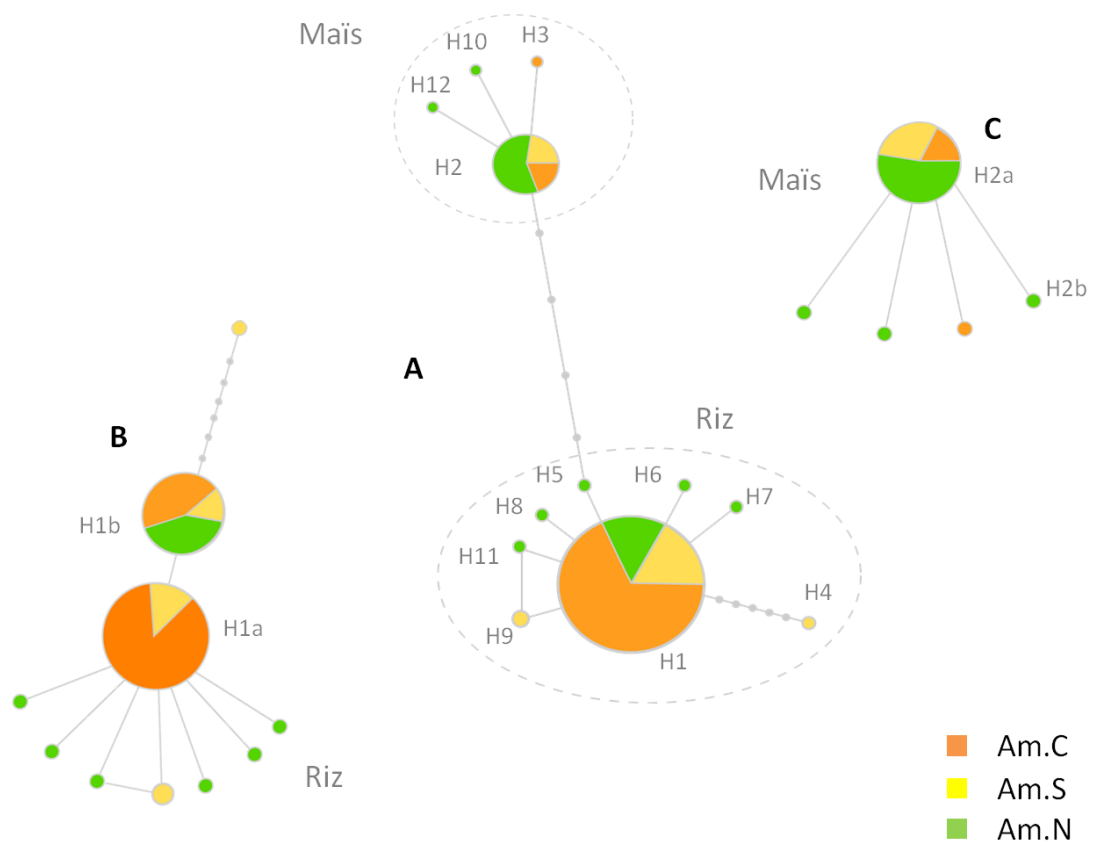


Figure 2 : Réseaux d'haplotypes pour *Spodoptera frugiperda* obtenus à partir du marqueur mitochondrial COI. Les cercles sont proportionnels à la fréquence des haplotypes retrouvés. Les points gris sur les branches représentent le nombre d'étapes de mutation entre haplotypes. Les spécimens ont été échantillonnés sur l'aire de distribution de *S. frugiperda* : Amérique Centrale (orange), Amérique du Sud (jaune) et Amérique du Nord (vert).

Tableau 1 : Statistiques de diversité pour les populations étudiées à partir des séquences COI (ns, non significatif. *, 0.01<P<0.05; **, 0.001<P<0.01; ***, P<0.001).

Population	Nombre d'individus	Nombre de Sites polymorphes	Nombre d'haplotypes	Hd	Pi(%)
Riz	117	13	10	0,527	0,287
Mais	25	4	5	0,300	0,088
Riz-Guadeloupe	60	1	2	0,183	0,044
Riz-Am.C	15	0	1	0,000	0,000
Riz-Am.S	22	10	4	0,597	0,894
Riz-Am.N	20	6	7	0,521	0,149
Riz-Arc.C	74	1	2	0,311	0,079
Riz-Am.C+Am.S	37	8	4	0,599	0,578
Mais-Am.C+Am.S	10	1	2	0,200	0,047
Mais-Am.N	15	3	4	0,371	0,107
Riz-autres	58	19	11	0,605	0,449
Riz-S1	11	4	4	0,533	0,081
Riz-S2	16	4	5	0,608	0,072
Riz-S3	19	4	4	0,380	0,052
Riz-S4	14	7	5	0,593	0,115

Différenciation génétique

Les individus des différents groupes sont scindés en deux grands groupes d'haplotypes séparés par quatre pas mutationnels (Figure 2-A). L'ensemble des individus riz sont compris dans le premier groupe, qui rassemble l'haplotype majoritaire (H1), retrouvé dans tous les groupes et dans la plupart des localités (tableau 2). S'ajoute à cela, cinq haplotypes uniques (H5-H8 et H11) chacun séparé par un pas mutationnel de H1 (retrouvés dans les populations du groupe Amérique du Nord), un haplotype H9 rassemblant deux individus du groupe Amérique du Sud (séparé d'un pas mutationnel de l'haplotype H1 et de l'haplotype H11) et un haplotype unique H4 (séparé de l'haplotype H1 par cinq pas mutationnels). Les individus du variant maïs quant à eux sont retrouvés dans le deuxième groupe, qui rassemble le deuxième plus gros haplotype (H2), lui aussi retrouvé dans toutes les populations, et trois haplotypes uniques H12 et H10 correspondant à deux individus d'Amérique du Nord et H3 correspondant à un individu d'Amérique centrale. Les trois haplotypes uniques sont chacun séparés de l'haplotype H2 par un pas mutationnel. L'arbre phylogénétique consensus obtenu en Inférence Bayésienne (Figure S1) met en évidence (probabilité postérieures $pp = 1$) la monophylie des clades correspondant aux deux variants

de *S. frugiperda* et permet d'identifier les individus du variant riz et du variant maïs au sein des différentes populations de *S. frugiperda* (Figure S1).

Tableau 2 : Localités échantillonnées et nombre d'individus collectés par populations (N) pour *Spodoptera frugiperda*.

Population	N	Localité	Occurrence (code haplotype)
Am. du Nord	35	USA	12(H1) 7(H2) 1(H6) 1(H7) 1(H8) 1(H11)
		Canada	6(H2) 3(H1) 1(H12) 1(H13)
Arc. Caraïbéen	80	Guadeloupe	60(H1) 1(H2) 1(H3)
		Rép. Dominicaine	2(H2)
		Costa Rica (Am. Centrale)	15(H1)
Am. du Sud	28	Brésil	3(H2) 6(H1) 2(H9) 1(H4) 1(H10)
		Pérou	2(H2)
		Guyane Fr.	13(H1)

La Figure 2-b et 2-C représente les réseaux d'haplotypes obtenus au sein de chacun des deux variants de *S. frugiperda*. Pour le variant riz (Figure 2-B), dix haplotypes sont identifiés avec la séparation de l'haplotype majoritaire en deux (H1a et h1b). L'haplotype H1a rassemble toute la population Guadeloupe et une partie de la population du groupe Amérique du sud (Brésil et Guyane Française) alors que l'haplotype H1b rassemble des individus de toutes les localités. Concernant le variant maïs (Figure 2-c), H2 est scindé en deux créant un cinquième haplotype H2b unique séparé de l'haplotype H2a par un pas mutationnel.

Enfin au sein de la population Guadeloupe variant riz il y a identification de deux haplotypes séparé par un pas mutationnel (tableau 1).

Les indices de différenciation F_{st} , K_{st} et S_{nn} entre les différents groupes inter et intra variants sont présentés dans le tableau 3 et mettent en évidence un niveau de différenciation moléculaire élevé entre les individus appartenant au variant riz et les individus appartenant au variant maïs. Au niveau intra-variant, contrairement au variant maïs, les trois valeurs des indices de différenciation génétique révèlent que les populations du variant riz sont génétiquement différentes en fonction de l'origine géographique des

individus (p -value < 0.001) même si cette différenciation est un peu moins soutenue statistiquement entre le groupe échantillonné dans l'Arc Caraïbéen et le groupe échantillonné en Amérique du Sud (p -value entre 0.01 et 0.05). Enfin au sein des individus du variant riz, échantillonnés en Guadeloupe, les indices de différenciation montrent une homogénéité génétique entre les individus des différentes localités S1, S2, S3 et S4.

Tableau 3 : Statistiques de différenciation des populations pour la séquence mitochondriale de la sous-unité I du cytochrome oxydase COI (alignement de 433 sites, sans gap). **, 0.001 < P < 0.01; ***, P < 0.001

Population	Population	Fst	Kst*	Snn
Riz	Maïs	0,9049	0,7261***	1,0000***
Riz-Guadeloupe	Riz-Am.C+Am.S	0,2409	0,3289***	0,7569***
Riz-Guadeloupe	Riz-Am.N	0,7692	0,7204***	1,0000***
Riz-Guadeloupe	Riz-Am.C	1,0000	1,0000***	1,0000***
Riz-Guadeloupe	Riz-Am.S	0,0568	0,1021***	0,7161***
Riz-Am.S	Riz-Am.N	0,1672	0,1712***	0,7575***
Riz-arc.C	Riz-Am.N	0,5899	0,3517***	0,8136***
Riz-arc.C	Riz-Am.S	0,0083	0,03071*	0,6687*
Maïs-Am.N	Maïs-Am.C+Am.S	0,0000	-0,01530	0,4937
Riz-Guadeloupe	Riz-autres	0,3879	0,3820***	0,7981***
Riz-S1	Riz-S2	-0,0833	-0,04194	0,48252
Riz-S1	Riz-S3	0,0000	0,072	0,53982
Riz-S1	Riz-S4	-0,08205	-0,04373	0,561
Riz-S2	Riz-S3	0,066667	0,05002	0,525
Riz-S2	Riz-S4	-0,07061	-0,03531	0,46311
Riz-S3	Riz-S4	0,076	0,07062	0,53939

Discussion

Cette étude présente les résultats obtenus à partir d'un fragment de séquence COI pour des individus représentatifs de l'aire de répartition de *S. frugiperda*. Le réseau d'haplotype mitochondrial ainsi que les estimations des différents indices statistiques de différenciation des populations montrent une structuration forte des individus du variant maïs par rapport aux individus des variants riz, sans corrélation avec la localité géographique. Ces résultats confirment le fort niveau de différenciation génétique entre les deux variants de *S. frugiperda* mis en évidence par différentes études récentes (Kergoat *et al.* 2012); Dumas *et al.* soumis) et apporte un argument de plus en faveur de l'hypothèse selon laquelle les variants de *S. frugiperda* correspondent à des espèces différentes. En effet les variants de *S. frugiperda* sont considérés comme pouvant être un exemple de spéciation en cours (Berlocher & Feder 2002; Groot *et al.* 2010).

Au sein du variant maïs, l'absence de structuration géographique pourrait s'expliquer par une plus grande capacité de migration du variant maïs par rapport au variant riz. Cependant le nombre d'individus échantillonnés pour le variant maïs étant faible il ne faut pas exclure la possibilité d'une sous-estimation du niveau de structuration pour ce variant et que cette absence de signal soit artéfactuelle. A l'inverse, en ce qui concerne les populations du variant riz, les indices de différenciation génétique suggèrent une structuration géographique des populations assez forte et ce même avec des échantillons d'effectifs relativement faible. Les résultats obtenus pour le variant riz échantillonné en Guadeloupe montrent une très forte différenciation génétique avec les autres populations du variant riz, ce qui suggère que l'augmentation du nombre d'échantillons pour les autres localités ne ferait que renforcer les résultats que nous avons obtenus. Cependant au sein des spécimens de Guadeloupe, échantillonnés sur quatre sites différents la même année, les résultats montrent qu'il y a très peu de variation génétique et que les individus forment un groupe génétiquement homogène.

L'ensemble de ces résultats sont en accord avec les études menées sur le niveau de différenciation génétiques réalisées sur des aires géographiques plus restreintes (Saldamando & Vélez-Arango 2010). Ces résultats préliminaires vont donc dans le sens de l'hypothèse proposée par Groot et ses collaborateurs (2010) au sujet d'un effet « population » relatif aux différences de niveau d'isolement reproducteur retrouvés dans la littérature entre les deux variants de *S. frugiperda*.

Références

- Berlocher SH, Feder JL (2002) Sympatric speciation in phytophagous insects: moving beyond controversy. *Annual Review in Entomology***47**, 773-815.
- Busato GR, Grutzmacher AD, de Oliveira AC, *et al.* (2004) Analysis of the molecular structure and diversity of *Spodoptera frugiperda* (JE smith) (Lepidoptera : Noctuidae) populations associated to the corn and rice crops in Rio Grande do Sul State, Brazil. *Neotropical Entomology***33**, 709-716.
- Capinera JL (1999) Fall Armyworm, *Spodoptera frugiperda* (J.E. Smith) (Insecta: Lepidoptera: Noctuidae). *Featured Creatures from the Entomology and Nematology Department, Florida Cooperative Extension Service.*

- Clark PL, Molina-Ochoa J, Martinelli S, *et al.* (2006) Population variation of the fall armyworm, *Spodoptera frugiperda*, in the Western Hemisphere. *Journal of Insect Science* **7**.
- Darriba D TG, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**, 772.
- Drummond AJ, Ashton B, Cheung M, *et al.* (2007) Geneious v3.8. .
- Groot A, Marr M, Schofl G, *et al.* (2008) Host strain specific sex pheromone variation in *Spodoptera frugiperda*. *Frontiers in Zoology* **5**, 20.
- Groot AT, M. M, Heckel DG, *et al.* (2010) The roles and interactions of reproductive isolation mechanisms in fall armyworm (Lepidoptera: Noctuidae) host strains. *Ecological Entomology* **35**, 105-118.
- Hudson R (2000) A new statistic for detecting genetic differentiation. *Genetics* **155**, 2011–2014.
- Hudson R, Boos D, Kaplan N (1992) A statistical test for detecting geographic subdivision *Molecular Biology and Evolution* **9**, 138–151.
- Juárez ML, Murúa MG, García MG, *et al.* (2012) Host Association of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) Corn and Rice Strains in Argentina, Brazil, and Paraguay. *Journal of Economic Entomology* **105**, 573-582.
- Kergoat GJ, Prowell DP, Le Ru BP, *et al.* (2012) Disentangling dispersal, vicariance and adaptive radiation patterns: A case study using armyworms in the pest genus *Spodoptera* (Lepidoptera: Noctuidae). *Molecular Phylogenetics and Evolution* **65**, 855-870.
- Larkin MA, Blackshields G, Brown NP, *et al.* (2007 doi:10.1093/bioinformatics/btm404) ClustalW and ClustalX version 2. *Bioinformatics* **23**, 2947-2948.
- Levy HC, Garcia-Maruniak A, Maruniak JE (2002) Strain identification of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) insects and cell line: PCR-RFLP of cytochrome oxidase C subunit I gene. *The Florida Entomologist* **85**, 186-190.
- Lu YJ, Adang MJ (1996) Distinguishing fall armyworm (Lepidoptera: Noctuidae) strain using a diagnostic mitochondrial DNA marker. *Florida Entomologist* **79**, 48-55.
- Machado V, Wunder M, Baldissera VD, *et al.* (2008) Molecular Characterization of Host Strains of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in Southern Brazil. *Annals of the Entomological Society of America* **101**, 619-626.
- Maddison W, Maddison D (2009) Mesquite: a modular system for evolutionary analysis Version 2.6.
- Martinelli S, Barata RM, Zucchi MI, *et al.* (2006) Molecular variability of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) populations associated to maize and cotton crops in Brazil. *Journal of Economic Entomology* **99**, 519-526.
- McMichael M, Pashley Prowell DP (1999) Differences in amplified fragment-length polymorphisms in fall armyworm (Lepidoptera: Noctuidae) host strain. *Annals of the Entomological Society of America* **92**, 175-181.
- Meagher RL, Gallo-Meagher M (2003) Identifying host strains of fall armyworm (Lepidoptera : Noctuidae) in Florida using mitochondrial markers. *Florida Entomologist* **86**, 450-455.

- Mitchell ER, McNeil JN, Westbrook JK, *et al.* (1991) Seasonal periodicity of Fall armyworm, (Lepidoptera: Noctuidae) in the Caribbean basin and northward to Canada. *Journal of Entomological Science***26**, 39-50.
- Nagoshi RN, Fleischer S, Meagher RL (2009) Texas Is the Overwintering Source of Fall Armyworm in Central Pennsylvania: Implications for Migration Into the Northeastern United States. *Environmental Entomology* **38**, 1546-1554
- Nagoshi RN, Gabriela Murúa M, Hay-Roe M, *et al.* (2012a) Genetic Characterization of Fall Armyworm (Lepidoptera: Noctuidae) Host Strains in Argentina. *Journal of Economic Entomology***105**, 418-428.
- Nagoshi RN, Meagher RL (2003) FR Tandem-Repeat Sequence in Fall Armyworm (Lepidoptera: Noctuidae) Host Strains. *Annals of the Entomological Society of America***96**, 329-335.
- Nagoshi RN, Meagher RL (2004) Behaviour and distribution of the two Fall armyworm host strains in Florida. *Florida Entomologist***87**, 440-449.
- Nagoshi RN, Meagher RL, Adamczyk JJ, *et al.* (2006) New restriction fragment length polymorphisms in the cytochrome oxidase I gene facilitate host strain identification of Fall armyworm (Lepidoptera: Noctuidae) populations in the southeastern United States. *Journal of Economic Entomology***99**, 671-677.
- Nagoshi RN, Meagher RL, Flanders K, *et al.* (2008) Using haplotypes to monitor the migration of Fall Armyworm (Lepidoptera: Noctuidae) corn-strain populations from Texas and Florida. *Journal of Economic Entomology***101**, 742-749.
- Nagoshi RN, Meagher RL, Hay-Roe M (2012b) Inferring the annual migration patterns of fall armyworm (Lepidoptera: Noctuidae) in the United States from mitochondrial haplotypes. *Ecology and Evolution***2**, 1458-1467.
- Nagoshi RN, Meagher RL, Jenkins DA (2010) Puerto Rico Fall Armyworm has only limited interactions with those from Brazil or Texas but could have substantial exchanges with Florida populations. *Journal of Economic Entomology***103**, 360-367.
- Nagylaki T (1998) Fixation indices in subdivided populations. *Genetics***148**, 1325-1332.
- Pair SD, Raulston JR, Sparks AN, *et al.* (1986) Fall Armyworm Distribution and Population Dynamics in the Southeastern States. *The Florida Entomologist***69**, 468-487.
- Pashley DP (1986) Host-associated genetic differentiation in fall armyworm (Lepidoptera: Noctuidae) a sibling species complex? *Ann Entomol Soc Am***79**, 898-904.
- Pashley DP (1989) Host-associated differentiation in armyworms (Lepidoptera: Noctuidae): An allozymic and mitochondrial DNA perspective. In: *Electrophoretic Studies on Agricultural Pests*, pp. 103-114, Oxford.
- Pashley DP, Johnson SJ, Sparks AN (1985) Genetic population structure of migratory moths: the fall armyworm (Lepidoptera: Noctuidae). *Annals of the Entomological Society of America***78**, 756-762.
- Pashley DP, Ke LD (1992) Sequence evolution in mitochondrial ribosomal and ND-1 genes in lepidoptera: implications for phylogenetic analyses. *Molecular Biology and Evolution***9**, 1061-1075.

- Pashley Prowell DP, Martin JA (1987) Reproductive incompatibility between host strains of the Fall armyworm (Lepidoptera: Noctuidae). *Annals of the Entomological Society of America*, 731-733.
- Pogue MG (2002) World revision of the genus *Spodoptera* Guenée (Lepidoptera: Noctuidae). *Memoirs of the American Entomological Society* **43**, 1-202.
- Prowell DP, McMichael M, Silvain J-F (2004) Multilocus genetic analysis of host use, introgression, and speciation in host strains of fall armyworm (Lepidoptera: Noctuidae). *Annals of the Entomological Society of America* **97**, 1034-1044
- Quisenberry SS (1991) Fall armyworm (Lepidoptera: Noctuidae) host strain reproductive compatibility. *Florida Entomologist* **74**, 194-199.
- Rambaut A, Drummond AJ (2007) Tracer v1.4, Available from <http://beast.bio.ed.ac.uk/Tracer>
- Ronquist F, Huelsenbeck J (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572 - 1574.
- Rozas J, Sánchez-DelBarrio JC, Messeguer X, *et al.* (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496-2497.
- Saldamando CI, Vélez-Arango AM (2010) Host plant association and genetic differentiation of corn and rice strains of *Spodoptera frugiperda* Smith (Lepidoptera: Noctuidae) in Colombia. *Neotropical Entomology* **39**, 921-929.
- Schöfl G, Dill A, Heckel DG, *et al.* (2011) Allochronic Separation versus Mate Choice: Nonrandom Patterns of Mating between Fall Armyworm Host Strains. *The American Naturalist* **177**, 470-485.
- Schöfl G, Heckel DG, Groot AT (2009) Time-shifted reproductive behaviours among fall armyworm (Noctuidae: *Spodoptera frugiperda*) host strains: evidence for differing modes of inheritance. *Journal of Evolutionary Biology* **22**, 1447–1459.
- Sparks AN (1979) A review of the biology of the Fall armyworm. *The Florida Entomologist* **62**, 82-87.
- Velásquez-Vélez MI, Saldamando-Benjumea CI, Ríos-Diez JD (2011) Reproductive Isolation between Two Populations of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) Collected in Corn and Rice Fields from Central Colombia. *Annals of the Entomological Society of America* **104**, 826-833.
- Whitford F, Quisenberry SS, Riley TJ, *et al.* (1988) Oviposition preference, mating compatibility, and development of two fall armyworm strains. *The Florida Entomologist* **71**, 234-243.

Tableau S1 : Echantillons de *Spodoptera frugiperda*

Genus species	Genbank accession no	Origine	Code Network	Population
<i>Spodoptera</i>				
<i>frugiperdamais</i> (Smith, 1797)		Peru	1CPE	AmSud
		Brazil	2CBR	AmSud
		Brazil	3CBR	AmSud
		Brazil	4CBR	AmSud
		Brazil	5CBR	AmSud
		Brazil	6CBR	AmSud
		Brazil	7CBR	AmSud
		Brazil	8CBR	AmSud
		Brazil	9CBR	AmSud
		Brazil	10CBR	AmSud
		Brazil	11CBR	AmSud
		Brazil	12CBR	AmSud
		Brazil	13CBR	AmSud
		Guadeloupe	14CGU	Caraïbes
		Guadeloupe	15CGU	Caraïbes
		USA (floride)	16CUS	AmNord
		USA (floride)	17CUS	AmNord
		USA (floride)	18CUS	AmNord
		USA (floride)	19CUS	AmNord
		USA (floride)	20CUS	AmNord
		USA (floride)	21CUS	AmNord
		USA (floride)	22CUS	AmNord
		Guadeloupe	23CGU	AmCentrale
		Paraguay	24CPY	AmSud
		DominicanRepublic	25CDO	AmCentrale
		DominicanRepublic	26CDO	AmCentrale
		Cuba	27CCU	AmCentrale
		Peru	28CPE	AmSud
		USA (massachusetts)	149US	AmNord
		Canada	157CA	AmNord
		Canada	158CA	AmNord
		Canada	159CA	AmNord
	Canada	160CA	AmNord	
	Canada	168CA	AmNord	
	Canada	169CA	AmNord	
	Canada	172CA	AmNord	
<i>frugiperda</i> riz (Smith, 1797)		Guadeloupe S1	59RGU	Caraïbes
		Guadeloupe S1	60RGU	Caraïbes
		Guadeloupe S1	61RGU	Caraïbes
		Guadeloupe S1	62RGU	Caraïbes
		Guadeloupe S1	63RGU	Caraïbes
		Guadeloupe S1	64RGU	Caraïbes
		Guadeloupe S1	65RGU	Caraïbes
		Guadeloupe S1	66RGU	Caraïbes
		Guadeloupe S1	67RGU	Caraïbes
		Guadeloupe S1	68RGU	Caraïbes
		Guadeloupe S1	69RGU	Caraïbes
		Guadeloupe S1	70RGU	Caraïbes
		Guadeloupe S1	71RGU	Caraïbes
		Guadeloupe S1	72RGU	Caraïbes
		Guadeloupe S1	73RGU	Caraïbes
		Guadeloupe S1	74RGU	Caraïbes
	Guadeloupe S2	75RGU	Caraïbes	
	Guadeloupe S2	76RGU	Caraïbes	

Guadeloupe S2	77RGU	Caraïbes
Guadeloupe S2	78RGU	Caraïbes
Guadeloupe S2	79RGU	Caraïbes
Guadeloupe S2	80RGU	Caraïbes
Guadeloupe S2	81RGU	Caraïbes
Guadeloupe S2	82RGU	Caraïbes
Guadeloupe S2	83RGU	Caraïbes
Guadeloupe S2	84RGU	Caraïbes
Guadeloupe S2	85RGU	Caraïbes
Guadeloupe S2	86RGU	Caraïbes
Guadeloupe S2	87RGU	Caraïbes
Guadeloupe S2	88RGU	Caraïbes
Guadeloupe S2	89RGU	Caraïbes
Guadeloupe S2	90RGU	Caraïbes
Guadeloupe S2	91RGU	Caraïbes
Guadeloupe S2	92RGU	Caraïbes
Guadeloupe S2	93RGU	Caraïbes
Guadeloupe S3	94RGU	Caraïbes
Guadeloupe S3	95RGU	Caraïbes
Guadeloupe S3	96RGU	Caraïbes
Guadeloupe S3	97RGU	Caraïbes
Guadeloupe S3	98RGU	Caraïbes
Guadeloupe S3	99RGU	Caraïbes
Guadeloupe S3	100RGU	Caraïbes
Guadeloupe S3	101RGU	Caraïbes
Guadeloupe S3	102RGU	Caraïbes
Guadeloupe S3	103RGU	Caraïbes
Guadeloupe S3	104RGU	Caraïbes
Guadeloupe S3	105RGU	Caraïbes
Guadeloupe S3	106RGU	Caraïbes
Guadeloupe S3	107RGU	Caraïbes
Guadeloupe S3	108RGU	Caraïbes
Guadeloupe S3	109RGU	Caraïbes
Guadeloupe S3	110RGU	Caraïbes
Guadeloupe S3	111RGU	Caraïbes
Guadeloupe S3	112RGU	Caraïbes
Guadeloupe S3	113RGU	Caraïbes
Guadeloupe S3	114RGU	Caraïbes
Guadeloupe S3	115RGU	Caraïbes
Guadeloupe S3	116RGU	Caraïbes
Guadeloupe S3	117RGU	Caraïbes
Guadeloupe S3	118RGU	Caraïbes
Guadeloupe S3	119RGU	Caraïbes
Guadeloupe S4	120RGU	Caraïbes
Guadeloupe S4	121RGU	Caraïbes
Guadeloupe S4	122RGU	Caraïbes
Guadeloupe S4	123RGU	Caraïbes
Guadeloupe S4	124RGU	Caraïbes
Guadeloupe S4	125RGU	Caraïbes
Guadeloupe S4	126RGU	Caraïbes
Guadeloupe S4	127RGU	Caraïbes
Guadeloupe S4	128RGU	Caraïbes
Guadeloupe S4	129RGU	Caraïbes
Guadeloupe S4	130RGU	Caraïbes
Guadeloupe S4	131RGU	Caraïbes

	Guadeloupe S4	132RGU	Caraïbes
	Guadeloupe S4	133RGU	Caraïbes
	Guadeloupe S4	134RGU	Caraïbes
	Guadeloupe S4	135RGU	Caraïbes
311 MF67	French Guiana	29RFG	AmSud
337 MR7	French Guiana	30RFG	AmSud
B10R	French Guiana	31RFG	AmSud
B12R	French Guiana	32RFG	AmSud
B7R	French Guiana	33RFG	AmSud
CE1	French Guiana	34RFG	AmSud
F1 329	French Guiana	35RFG	AmSud
F1 330	French Guiana	36RFG	AmSud
F4 322	French Guiana	37RFG	AmSud
FRUB10	Brazil	38RBR	AmSud
FRUB9	Brazil	39RBR	AmSud
G9 305	French Guiana	40RFG	AmSud
G9 308	French Guiana	41RFG	AmSud
GK424	Brazil	42RBR	AmSud
GK427	Brazil	43RBR	AmSud
GK429	Brazil	44RBR	AmSud
HM136593	USA (floride)	45RUS	AmNord
HM136594	USA (floride)	46RUS	AmNord
HM136595	USA (floride)	47RUS	AmNord
HM136596	USA (floride)	48RUS	AmNord
HM136597	USA (floride)	49RUS	AmNord
HM136598	USA (floride)	50RUS	AmNord
LSU22R	Guadeloupe	51RGU	AmCentrale
MFG 50	French Guiana	52RFG	AmSud
MFG 56	French Guiana	53RFG	AmSud
MNHN10	DominicanRepublic	54RDO	Caraïbes
MNHN11	Venezuela	55RVE	AmSud
MNHN13	Mexico	56RMX	AmCentrale/Caraïbes
MNHN15	Guatemala	57RGT	AmCentrale/Caraïbes
MNHN9	DominicanRepublic	58RDO	AmCentrale/Caraïbes
JQ559528	Costa Rica	136CR	AmCentrale/Caraïbes
JQ554012	Costa Rica	137CR	AmCentrale/Caraïbes
JQ572603	Costa Rica	138CR	AmCentrale/Caraïbes
JQ571459	Costa Rica	139CR	AmCentrale/Caraïbes
JQ547900	Costa Rica	140CR	AmCentrale/Caraïbes
JQ577923	Costa Rica	141CR	AmCentrale/Caraïbes
JF854746	Brazil	142BR	AmSud
JF854745	Brazil	143BR	AmSud
JF854744	Brazil	144BR	AmSud
JF854743	Brazil	145BR	AmSud
JF854742	Brazil	146BR	AmSud
JF854741	Brazil	147BR	AmSud
JF854740	Brazil	148BR	AmSud
HQ964487	USA (massachusetts)	150US	AmNord
HQ964486	USA (massachusetts)	151US	AmNord
HQ964485	USA (massachusetts)	152US	AmNord
HQ964443	USA (massachusetts)	153US	AmNord
HQ964442	USA (massachusetts)	154US	AmNord
HQ964441	USA (massachusetts)	155US	AmNord
GU094756	Canada	156CA	AmNord
HQ964440	USA (massachusetts)	161US	AmNord

HQ964439	USA (massachusetts)	162US	AmNord
HQ964394	USA (massachusetts)	163US	AmNord
HQ964393	USA (massachusetts)	164US	AmNord
HQ964353	USA (massachusetts)	165US	AmNord
HQ964352	USA (massachusetts)	166US	AmNord
HQ964351	USA (massachusetts)	167US	AmNord
GU439149	Canada	170CA	AmNord
GU439148	Canada	171CA	AmNord
GU095403	Canada	173CA	AmNord
GU159435	Costa Rica	174CR	AmCentrale/Caraïbes
GU159434	Costa Rica	175CR	AmCentrale/Caraïbes
GU159433	Costa Rica	176CR	AmCentrale/Caraïbes
GU159431	Costa Rica	177CR	AmCentrale/Caraïbes
GU159430	Costa Rica	178CR	AmCentrale/Caraïbes
GU159429	Costa Rica	179CR	AmCentrale/Caraïbes
GU159428	Costa Rica	180CR	AmCentrale/Caraïbes
GU159427	Costa Rica	181CR	AmCentrale/Caraïbes
GU159426	Costa Rica	182CR	AmCentrale/Caraïbes

* échantillons utilisés dans des études antérieures (Kergoat et al. 2012 and Dumas et al. soumis)

** nouvelles séquences GenBank

2. Deuxième niveau d'analyse : analyse génétique et génomique sur population de laboratoire

Dans ce second chapitre, l'échelle d'analyse devient plus fine et correspond à l'étude du degré de différenciation entre un petit nombre d'individus de chaque variant dans des conditions expérimentales contrôlées. En effet ce chapitre concerne des individus de populations de laboratoire et met en évidence l'existence d'un isolement post-zygotique entre les deux variants, grâce à des outils de génétique et de génomique.

Ce chapitre comprend un premier article technique sur la caractérisation d'un lot de 17 marqueurs microsatellites au sein des populations de laboratoire et montre que, malgré un temps relativement long d'élevage, il existe un niveau de polymorphisme relativement élevé au sein de chacun des variants de *S. frugiperda*.

Le deuxième article présenté dans ce chapitre est dédié à la caractérisation d'un isolement post-zygotique entre les deux variants, qui se base tout d'abord sur la mise en évidence d'un biais de taux de fertilité des couples interspécifiques, dépendant du sens d'accouplement. De plus cet article étudie la ségrégation des marqueurs microsatellites dans la descendance F2 hybride dans les deux sens de croisement interspécifiques et lors de croisements intraspécifiques. Les résultats obtenus montrent que la moitié des marqueurs microsatellites ne ségrégent pas correctement dans la descendance. Les régions génomiques contenant les séquences des microsatellites présentant ou non une distorsion de ségrégation sont ensuite identifiées. Dans un dernier temps, le niveau de différenciation entre les deux variants sera évalué dans des gènes portés par ces régions génomiques. Cette démarche nous permettra peut-être d'identifier des gènes candidats à une incompatibilité hybride.

2.1. Caractérisation des marqueurs microsatellites

Article 4 publié dans *Molecular Ecology Resources* : *Isolation and characterization of microsatellite loci in the fall armyworm, Spodoptera frugiperda (Lepidoptera, Noctuidae)*.

PERMANENT GENETIC RESOURCES NOTE

Permanent Genetic Resources added to Molecular Ecology Resources Database 1 December 2011 – 31 January 2012

MOLECULAR ECOLOGY RESOURCES PRIMER DEVELOPMENT CONSORTIUM,¹ M. C. ARIAS,² E. ARNOUX,³ JAMES J. BELL,⁴ ABEL BERNADOU,⁵ GIORGIA BINO,⁶ R. BLATRIX,⁷ DENIS BOURGUET,⁸ CECILIA CARREA,⁹ ANNE-LAURE CLAMENS,⁸ HAYDÉE A. CUNHA,^{10, 11} E. D'ALENÇON,¹² YI DING,¹³ C. DJIETO-LORDON,¹⁴ M. P. DUBOIS,⁷ P. DUMAS,^{8, 12} C. ERAUD,¹⁵ B. FAIVRE,³ F. O. FRANCISCO,² E. FRANÇOZO,² M. GARCIA,⁷ JONATHAN P. A. GARDNER,⁴ S. GARNIER,³ S. GIMENEZ,¹² JOHN R. GOLD,¹⁶ D. J. HARRIS,¹⁷ GUANGCUN HE,¹³ B. HELLEMANS,¹⁸ CHRISTOPHER M. HOLLENBECK,¹⁶ SHENGLI JING,¹³ G. J. KERGOAT,⁸ BINGFANG LIU,¹³ JAN R. McDOWELL,¹⁹ D. McKEY,⁷ TERRENCE L. MILLER,¹⁹ ERICA NEWTON,^{9, 20} KATRINA M. PAGENKOPP LOHAN,¹⁹ CHIARA PAPETTI,⁶ IAN PATERSON,⁹ J. PECCOUD,⁷ XINXIN PENG,¹³ F. PIATSCHHECK,⁷ SERGINE PONSARD,⁵ KIMBERLY S. REECE,¹⁹ CÉLINE M. O. REISSER,^{4, 21} MARK A. RENSHAW,¹⁶ DANIEL E. RUZZANTE,⁹ M. SAUVE,⁷ JEFFREY D. SHIELDS,¹⁹ ANTONIO SOLÉ-CAVA,¹⁰ E.L. SOUCHE,^{18, 22} J.K.J. VAN HOUDT,^{18, 23} ANDERSON VASCONCELLOS,¹⁰ F. A. M. VOLCKAERT,¹⁸ SHUZHEN WANG,¹³ JIE XIAO,¹⁹ HANGJIN YU,¹³ LORENZO ZANE,⁶ BARBARA ZANNATO,⁶ TYLER S. ZEMLAK,^{9, 24} CHUNXIAO ZHANG,¹³ YAN ZHAO,¹³ XI ZHOU¹³ and LILI ZHU¹³

¹Molecular Ecology Resources Editorial Office, 6270 University Blvd, Vancouver, BC, V6T 1Z4, Canada; ²Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, rua do Matão, 277, 05508-090, São Paulo, SP, Brazil;

³Biogéosciences UMR 5561, Université de Bourgogne, 6 Boulevard Gabriel, 21000 Dijon, France; ⁴Centre for Marine Environmental and Economic Research, School of Biological Sciences, Victoria University of Wellington, Wellington 6140, New Zealand; ⁵CNRS – UMR 5174, Laboratoire Evolution & Diversité Biologique, 118 route de Narbonne, 31062 Toulouse Cedex 09, France. Université de Toulouse – UMR 5174, Laboratoire Evolution & Diversité Biologique, 118 route de Narbonne, 31062 Toulouse Cedex 09, France;

⁶Department of Biology, University of Padova, via U. Bassi 58/B, I-35121 Padova, Italy; ⁷Centre d'Ecologie Fonctionnelle et Evolutive (CEFE), UMR 5175 CNRS, 1919 route de Mende, F-34293 Montpellier Cedex 5, France; ⁸Centre de Biologie pour la Gestion des Populations (CBGP), UMR INRA-IRD-CIRAD Montpellier SupAgro, Campus International de Baillarguet, CS 30016, 34988 Montferrier sur-Lez Cedex, France; ⁹Department of Biology, Dalhousie University, Halifax, NS, Canada B3H4R2;

¹⁰Laboratório de Biodiversidade Molecular, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil;

¹¹Laboratório de Mamíferos Aquáticos e Bioindicadores (MAQUA), Faculdade de Oceanografia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil; ¹²UMR 1333, INRA, Université Montpellier 2, Diversité, Génomes & Interactions Microorganismes-Insectes, Place Eugène Bataillon, 34095 Montpellier, France; ¹³State Key Laboratory of Hybrid Rice, College of Life Science, Wuhan University, Wuhan 430072, People's Republic of China; ¹⁴Laboratory of Zoology, Yaoundé University, Faculty of Science, PO Box 812, Yaoundé, Cameroun; ¹⁵Office National de la Chasse et de la Faune Sauvage, Carrefour de la Canauderie, Station biologique de Chizé, 79360 Villiers en Bois, France; ¹⁶Center for Biosystematics and Biodiversity, Department of Wildlife and Fisheries Sciences, Texas A and M University, College Station, TX 77843-2258, USA; ¹⁷Royal Botanic Garden Edinburgh, Edinburgh EH3 5LR, Scotland, UK; ¹⁸Laboratory of Biodiversity and Evolutionary Genomics, University of Leuven, Charles Deberiotstraat 32, B-3000 Leuven, Belgium; ¹⁹The Virginia Institute of Marine Science, The College of William & Mary, PO Box 1346, Gloucester Point, VA 23062, USA; ²⁰Environmental and Life Sciences Department, Trent University, Peterborough, ON, Canada K9J 7B8; ²¹Department of Biology, Unit of Ecology and Evolution, University of Fribourg, Fribourg CH-1700, Switzerland; ²²Bioinformatics for Genomic Analysis, Institut Pasteur, 28 rue du Docteur Roux, F-75724 Paris Cedex, France; ²³Laboratory for Cytogenetics and Genome Research, University of Leuven, O&N, Herestraat 49, B-3000 Leuven, Belgium; ²⁴Department of Microbiology & Immunology, Dalhousie University, Halifax, NS, Canada B3H4R2

Abstract

This article documents the addition of 473 microsatellite marker loci and 71 pairs of single-nucleotide polymorphism (SNP) sequencing primers to the Molecular Ecology Resources Database. Loci were developed for the following species: *Barteria fistulosa*, *Bombus morio*, *Galaxias platei*, *Hematodinium perezii*, *Macrocentrus cingulum* Brischke (a.k.a. *M. abdominalis*)

Correspondence: Molecular Ecology Resources Primer Development Consortium, E-mail: editorial.office@molecol.com

Fab., *M. grandii* Goidanich or *M. gifuensis* Ashmead), *Micropogonias furnieri*, *Nerita melanotragus*, *Nilaparvata lugens* Stål, *Sciaenops ocellatus*, *Scomber scombrus*, *Spodoptera frugiperda* and *Turdus lherminieri*. These loci were cross-tested on the following species: *Barteria dewevrei*, *Barteria nigritana*, *Barteria solida*, *Cynoscion acoupa*, *Cynoscion jamaicensis*, *Cynoscion leiarchus*, *Cynoscion nebulosus*, *Cynoscion striatus*, *Cynoscion virescens*, *Macrodon ancylodon*, *Menticirrhus americanus*, *Nilaparvata muiri* and *Umbrina canosai*. This article also documents the addition of 116 sequencing primer pairs for *Dicentrarchus labrax*.

This article documents the addition of 473 microsatellite marker loci and 71 pairs of single-nucleotide polymorphism (SNP) genotyping primers to the Molecular Ecology Resources Database. Table 1 contains information on

the focal species, the number of loci developed, any other species the loci were tested in and the accession numbers for the loci in both the Molecular Ecology Resources Database and GenBank. The authors responsible for each

Table 1 Information on the focal species, the number of loci developed, any other species the loci were tested in and the accession numbers for the loci in both the Molecular Ecology Resources Database and GenBank. The authors responsible for each set of loci are listed in the final column

Species	No. primers developed	Other species tested	MER database no.	GenBank accession no.	Authors
<i>Barteria fistulosa</i>	17	<i>B. dewevrei</i> , <i>B. nigritana</i> , <i>B. solida</i>	48615–48631	JQ289126–JQ289128, JQ289130–JQ289143	Sauve, M.; Garcia, M.; Djipto-Lordon, C.; Peccoud, J.; Piatscheck, F.; Dubois, M. P.; McKey, D.; Harris, D. J.; Blatrix, R.
<i>Bombus morio</i>	13	n/a	47715–47722, 47724–47728	JN701023–JN701035	Françoso, E.; Francisco, F. O.; Arias, M. C.
<i>Galaxias platei</i>	9	n/a	47692–47700	JQ043430–JQ043438	Zemlak, Tyler S.; Carrea, Cecilia; Newton, Erica; Paterson, Ian;
<i>Hematodinium perezii</i>	11	n/a	47681–47691	JN403077–JN403087	Ruzzante, Daniel E. Pagenkopp Lohan, Katrina M.; McDowell, Jan R.; Shields, Jeffrey D.; Xiao, Jie; Miller, Terrence L.; Reece, Kimberly S.
<i>Macrocentrus cingulum</i> Brischke (a.k.a. <i>M. abdominalis</i> Fab., <i>M. grandii</i> Goidanich or <i>M. gifuensis</i> Ashmead)	13	n/a	47202–47222	JN703391–JN703395, JN703399, JN703401, JN703403–JN703405, JN703409–JN703411	Clamens, Anne-Laure; Bernadou, Abel; Ponsard, Sergine; Bourguet, Denis
<i>Micropogonias furnieri</i>	16	<i>C. acoupa</i> , <i>C. jamaicensis</i> , <i>C. leiarchus</i> , <i>C. striatus</i> , <i>C. virescens</i> , <i>M. ancylodon</i> , <i>M. americanus</i> , <i>U. canosai</i>	48472–48487	JF798645–JF798650, JF798652–JF798656, JF798660–JF798664	Vasconcellos, Anderson; Cunha, Haydée A.; Solé-Cava, Antonio
<i>Nerita melanotragus</i>	10	n/a	48579–48588	JQ364728–JQ364737	Reisser, Céline M. O.; Bell, James J.; Gardner, Jonathan P. A.

Table 1 (Continued)

Species	No. primers developed	Other species tested	MER database no.	GenBank accession no.	Authors
<i>Nilaparvata lugens</i> Stål	17	<i>N. muiri</i>	47663–47677, 47679, 47680	JN834077–JN834091	Jing, Shengli; Zhang, Chunxiao; Wang, Shuzhen; Zhou, Xi; Yu, Hangjin; Peng, Xinxin; Liu, Bingfang; Zhao, Yan; Zhu, Lili; Ding, Yi; He, Guangcun
<i>Sciaenops ocellatus</i>	323	<i>Cynoscion nebulosus</i>	47770–47772, 47774, 47775, 47777, 47779, 47781, 47782, 47784–47788, 47813, 47815–47890, 48016–48043, 48082–48108, 48295–48471	EU726996, JF509110–JF509136, JQ235209–JQ235503	Renshaw, Mark A.; Hollenbeck, Christopher M.; Gold, John R.
<i>Scomber scombrus</i>	15	n/a	48512, 48514, 48515, 48517, 48520, 48522, 48525, 48527–48534	JQ219865–JQ219879	Papetti, Chiara; Bino, Giorgia; Zannato, Barbara; Zane, Lorenzo
<i>Spodoptera frugiperda</i>	17	n/a	48552–48568	JN790256–JN790271, JN796474	Dumas, P.; Clamens, A.-L.; Kergoat, G. J.; Gimenez, S.; d'Alençon, E.
<i>Turdus lherminieri</i>	12	n/a	47729–47740	JQ001753–JQ001762, JQ001764, JQ001765	Arnoux, E.; Faivre, B.; Eraud, C.; Garnier, S.

Table 2 Information on the focal species, the sequencing primer pairs developed, the number of single-nucleotide polymorphisms observed and any other species the loci were tested in. The next columns contain the number of allele specific primers and probes developed, and the Molecular Ecology Resources Database and GenBank accession numbers, respectively. The authors responsible for each set of loci are listed in the final column

Species	No. primer pairs	No. SNPs in sequence	Other species tested	No. Allele specific primers /probe	Target gene(s)	MER database numbers	Genbank Accession no	Authors
<i>Dicentrarchus labrax</i>	116	71	n/a	See MS for details	See MS for details	48237–48294	JM497134– JM497249	Souche, E. L.; Hellemans, B.; Van Houdt, J. K. J.; Volckaert, F. A. M.

set of loci are listed in the final column. Table 2 presents information on SNP genotyping resources added to the MER database and presents data on the focal species, the number of sequencing primer pairs, the observed number of SNPs, other species the loci were tested in and the number of allele specific primers or probes. The MER

database and GenBank accession numbers and the authors responsible are also listed. A full description of the development protocol for the loci presented here can be found on the Molecular Ecology Resources Database (<http://tomato.biol.trinity.edu/>).

1 **Isolation and characterization of microsatellite loci in the fall**
2 **armyworm, *Spodoptera frugiperda* (Lepidoptera, Noctuidae)**

3 P. Dumas^{1,2*}, A.-L. Clamens², G. J. Kergoat², S. Gimenez¹, E. d'Alençon¹,

4

5

6 Affiliations

7 ¹ UMR 1333, INRA, Université Montpellier 2, Diversité, Génomes & Interactions
8 Microorganismes-Insectes, Place Eugène Bataillon, 34095 Montpellier, France.

9 ² INRA, UMR Centre de Biologie pour la Gestion des Populations, CBGP, (INRA / IRD /
10 CIRAD / Montpellier SupAgro), Campus International de Baillarguet, CS30016, 34988
11 Montferrier-sur-Lez, France.

12

13 Keywords: fall armyworm, microsatellites, Lepidoptera, Noctuidae, pest, *Spodoptera*
14 *frugiperda*

15

16

17

18

19

20 * Corresponding author: Pascaline Dumas

21 Tel. +33 467 144 720

22 Fax. +33 467 144 299

23

24 E-mail: pascaline.dumas@univ-montp2.fr

25

26 **Abstract**

27 The fall armyworm, *Spodoptera frugiperda*, is a widespread and well-known agricultural pest
28 in the Western hemisphere. This moth is known to attack several major crops such as cotton,
29 maize and rice. It consists of two morphologically undistinguishable strains with different
30 host plant preferences (generally referred as the corn or rice strain). The development of
31 microsatellite markers is a prerequisite for more thorough studies on population and taxa
32 structure within *S. frugiperda*. In this study, a total of 17 polymorphic microsatellites were
33 isolated and characterized. We tested our loci on 46 individuals, one population of 23
34 individuals from each strain. Among them, 5 were unique alleles distinguishing the two
35 strains. Characterization of microsatellite loci was performed using laboratory strains of the
36 two taxa. No significant deviations from Hardy-Weinberg equilibrium were observed except
37 at one locus in the corn strain and one in the rice strain. These deviations are likely due to the
38 presence of null alleles at these loci. We will use this microsatellite markers set for future
39 study of phylogeographic structure and population genetics of the two *S. frugiperda* strains.

40

41

42

43

44

45

46

47

48

49

50

51 The fall armyworm *Spodoptera frugiperda* (J.E. Smith) is a noctuid moth (Lepidoptera,
52 Noctuidae). This moth is a well-known polyphagous pest species, which is considered as the
53 primary pest of corn in the Western hemisphere (Barros *et al.* 2010). It is also a species
54 responsible for economic damage on several important cultivated species of Poaceae such as
55 rice, sorghum, sugarcane and wheat (Pogue 2002). In addition, the fall armyworm is
56 increasingly becoming a pest of cotton and soybean (Martinelli *et al.* 2006). Populations of *S.*
57 *frugiperda* are subdivided into at least two strains which strongly differ in their host plant
58 preferences (Prowell *et al.* 2004). The corn strain is found on corn, sorghum, and cotton,
59 whereas the rice strain is mainly observed on rice and bermuda grass (Pashley 1986).
60 Individuals belonging to the two strains are morphologically similar but are differentiated by
61 both genomic (Nagoshi & Meagher 2003) and mitochondrial (Lu & Adang 1996) DNA.
62 Among all genetic markers, microsatellites are numerous, distributed over the whole genome,
63 and allow comprehensive studies of population genetic structure (Ellegren 2004).
64 In order to analyze the population structure of the two host strains of *S. frugiperda*, we
65 developed and characterized 21 microsatellite loci. In the course of our study, the group of
66 Arias *et al.*, (2011) published another set of 192 microsatellites of *S. frugiperda*. However,
67 they did not include information on differences between the two strains (Arias *et al.* 2011).
68 Microsatellite markers were isolated using the new generation 454 FLX Titanium
69 pyrosequencing technology. As starting material, total genomic DNA was extracted from four
70 adults of *S. frugiperda* (corn laboratory strain, originating from Guadeloupe), using Wizard
71 Genomic DNA purification Kit (PROMEGA) and following manufacturer's instructions.
72 Enrichment in microsatellites sequences was carried out at Genoscreen (Lille, France)
73 according to the procedure described by Malausa *et al.* (2011).
74 A total of 2,281 concatenated sequences, longer than 80 bp and containing microsatellites
75 (with more than five repeats) were obtained, of which 377 allowed design of PCR

76 amplification primer pairs using QDD software (Megléczy *et al.* 2010). Based on the expected
77 sizes (100 to 400 pb) of amplification products and annealing temperatures of primers, we
78 tested the amplification of 60 loci. Simplex PCR amplifications were carried out in a 10 μ L
79 final volume containing 2x Multiplex Master Mix (Qiagen), 0.2 μ M of each primer and 30 ng
80 of genomic DNA. The thermal profile started with an initial denaturation at 95 °C (15 min)
81 followed by 37 cycles of denaturation at 94°C (30 s), annealing at 60°C (1.5 min) and
82 extension at 72 °C (1 min), and final extension at 60°C for 30 min. From these trials 21
83 primer pairs which amplified DNA bands corresponding to the predicted size of the PCR
84 product, were selected for multiplex tests. One of the two primers of each locus was 5'end
85 labelled with a fluorescent dye (FAM, VIC, NED or PET). Dye colours were chosen, with
86 Multiplex Manager software, in order to analyse simultaneously loci of similar allelic size and
87 avoid overlap between loci with the same dye. The 21 loci were then amplified in four
88 multiplex PCRs and run on a capillary electrophoresis genotyping/sequencing apparatus
89 (ABI13130XL, 16 capillaries). The PCRs were performed on a Mastercycler ep gradient S
90 (Eppendorf) using the Multiplex PCR Kit (Qiagen). They were carried out in 10 μ L reaction
91 volume containing the QIAGEN Multiplex PCR Master Mix (1x) (including *Taq*, dNTPs and
92 3 mM of MgCl₂ as final concentration), 0.2 μ M of each primer, and 50 ng of genomic DNA.
93 PCRs started with an initial activation step at 95°C for 15 min, followed by 35 cycles with
94 denaturation at 94°C for 30 s, annealing at 60°C for 90 s, extension at 72°C for 60 s and final
95 extension of 60°C for 30 min. The PCR products were diluted 100 fold for all microsatellite
96 loci and then 2 μ L of each PCR was mixed with 0.1 μ L of a fluorescent size ladder (GeneScan
97 500 LIZ) and 15 μ L of Hidi Formamide (Applied Biosystems). The profiles obtained after
98 electrophoresis on the ABI 3130 automated sequencer were analysed using GeneMapper v.3.7
99 software (Applied Biosystems, USA).

100 We evaluated the polymorphism at these 21 loci of individuals from both strains. Those
101 strains have been seeded with 30 to 50 pupae ten and four years ago for the corn and rice
102 strain, respectively. Since then, they were reared in laboratory conditions (on an artificial
103 diet, at 24°C with a 16:8 photoperiod and hygrometry of 40 %). The individuals that seeded
104 the corn strain came from Guadeloupe whereas those that seeded the rice strain came from
105 Florida. We genotyped 23 individuals of both corn and rice strain. Genomic DNA was
106 extracted using DNAeasy Blood and Tissue Kit (Qiagen). Tests for linkage disequilibrium
107 and deviations from Hardy-Weinberg proportions were performed using GENEPOP v.4.0
108 (Rousset 2008). Among the 21 microsatellite loci, 17 were polymorphic and four were
109 monomorphic. Five of these 17 markers had unique alleles that distinguished the two strains.
110 Most loci showed no significant deviation from Hardy-Weinberg expectations in both taxa.
111 After Bonferroni correction, two loci, *Sfrugi33* and *Sfrugi 24* show a significant deficit in
112 heterozygotes in the corn and rice strain, respectively (Table 1). A large difference between
113 observed and expected heterozygosity is found at four loci (*Sfrugi2*, *Sfrugi29*, *Sfrugi38* and
114 *Sfrugi30*), that are not marked as being out of Hardy-Weinberg equilibrium. This large
115 difference may not be significant and due to the small number of alleles found. This would be
116 confirmed by genotyping more individuals from several populations. Analysis of the data with
117 the software MICRO-CHECKER (Van Oosterhout *et al.* 2004) indicated that this deficit was
118 likely due to the presence of null alleles at *Sfrugi33* and *Sfrugi24*. Only one pair of loci
119 displayed a significant linkage ($P = 0.014$): *Sfrugi2* and *Sfrugi38* in the corn strain. This may
120 be artifactual. First, the same pair of loci did not show any linkage in the rice strain and the P
121 value did not remain significant if we performed a correction to take into account the
122 multiplicity of tests.

123 Despite the long time that the strains used in this study were captively reared, most (17 out of
124 21) microsatellite loci showed polymorphism. The genetic diversity, the absence of linkage

125 and the deviation from Hardy-Weinberg equilibrium at only two loci make this new set of
126 microsatellite loci particularly suitable for studying the genetic structure of natural
127 populations of *S. frugiperda*. In both strains, the low proportion of monomorphic locus after
128 four to ten years of laboratory rearing might be explained by the relative high number of
129 individuals (30 to 50) that founded those strains and by the sizes of the strains which were
130 maintained each generation around 50-100 individuals. In natural populations we can expect a
131 higher level of genetic diversity at the 17 polymorphic loci revealed in the present study.
132 Further testing will be needed to confirm whether the loci continue to distinguish the two
133 strains from multiple populations.

134

135 **Acknowledgements**

136 This work was supported by a grant from the SPE department of INRA. P. Dumas is the
137 recipient of a PhD grant from Montpellier 2 University and the SPE department of INRA.
138 Data provided in this work were partly produced through technical facilities of the SFR
139 “Montpellier Environnement Biodiversité”. We thank Dr D. Bourguet for his constructive
140 comments on the manuscript, Dr R. Meagher for the kind gift of pupae from the *S. frugiperda*
141 “rice” strain and C. Gibard and G. Clabots for maintaining rearing facilities.

142

143

144 **References**

- 145 Arias RS, Blanco CA, Portilla M, *et al.* (2011) First microsatellites from *Spodoptera frugiperda*
146 (Lepidoptera: Noctuidae) and their potential use for population genetics. *Annals of the*
147 *Entomological Society of America* **104**, 576-587.
- 148 Barros EM, Torres JB, Ruberson JR, *et al.* (2010) Development of *Spodoptera frugiperda* on different
149 hosts and damage to reproductive structures in cotton. *Entomologia Experimentalis et*
150 *Applicata* **137**, 237-245.

- 151 Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* **5**, 435-
152 445.
- 153 Lu YJ, Adang MJ (1996) Distinguishing fall armyworm (Lepidoptera: Noctuidae) strain using a
154 diagnostic mitochondrial DNA marker. *Florida Entomologist* **79**, 48-55.
- 155 Malausa T, Gilles A, MeglÉCz E, *et al.* (2011) High-throughput microsatellite isolation through 454
156 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources*
157 **11**, 638-644.
- 158 Martinelli S, Barata RM, Zucchi MI, *et al.* (2006) Molecular variability of *Spodoptera frugiperda*
159 (Lepidoptera: Noctuidae) populations associated to maize and cotton crops in Brazil. *Journal*
160 *of Economic Entomology* **99**, 519-526.
- 161 MeglÉCz E, Costedoat C, Dubut V, *et al.* (2010) QDD: a user-friendly program to select microsatellite
162 markers and design primers from large sequencing projects. *Bioinformatics* **26**, 403-404.
- 163 Nagoshi RN, Meagher RL (2003) Fall armyworm FR sequences map to sex chromosomes and their
164 distribution in the wild indicate limitations in interstrain mating. *Insect Molecular Biology* **12**,
165 453-458.
- 166 Pashley DP (1986) Host-associated genetic differentiation in fall armyworm (Lepidoptera: Noctuidae):
167 a sibling species complex? *Annals of the Entomological Society of America* **79**, 898-904.
- 168 Pogue MG (2002) World revision of the genus *Spodoptera* Guenée (Lepidoptera: Noctuidae).
169 *Memoirs of the American Entomological Society* **43**, 1-202.
- 170 Prowell DP, McMichael M, Silvain J-F (2004) Multilocus genetic analysis of host use, introgression,
171 and speciation in host strains of fall armyworm (Lepidoptera: Noctuidae). *Annals of the*
172 *Entomological Society of America* **97**, 1034-1044
- 173 Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows
174 and Linux. *Molecular Biology and Evolution Resources* **8**, 103-106.
- 175 Van Oosterhout C, Hutchinson F, Wills DPM, *et al.* (2004) MICRO-CHECKER: software for
176 identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*
177 **4**, 535-538.

178

179

Table 1 Characteristics of loci isolated in *Spodoptera frugiperda*, with primer sequences, size of allele, number of alleles (No), size range of PCR products, expected (*He*) and observed (*Ho*) heterozygosities with deviation from Hardy-Weinberg proportions after Bonferroni correction (* $P < 0.03$, ** $P < 0.0006$).

locus	Primer sequence (5'-3') Include fluo-label dye	Repeat motif	Size (bp)	No Total	Corn (n=16)			Rice (n=15)			GenBank accession no.
					No	Size range	<i>Ho/He</i>	No	Size range	<i>Ho/He</i>	
Sfrugi34	F: VIC-TCATCATTTGCATGACCAACT R: AGGGAGGCAAGCAAAGAAAT	(TG) ₉	165	2	1	167	-	1	158	-	JN790264
Sfrugi37	F: FAM-AGTAAATCCTGGCAACACGC R: TGTCTGGAGCAGTTATGCATTT	(AG) ₈	115	2	1	112	-	2	110 - 112	0.61/0.50	JN790265
Sfrugi5	F: FAM-GCTAGGAGTTTTCGGGAAAGTA R: CGTCCATATTACCCCGAGTG	(AC) ₇	300	2	1	330	-	2	292 - 294	0.52/0.49	JN790256
Sfrugi2	F: NED-CGGCTAGGGGTGAAATGTAA R: GCGTTGAATATAATGCGGCT	(GT) ₆	369	3	2	365 - 367	0.70/0.48	2	367 - 371	0.43/0.39	JN796474
Sfrugi25	F: FAM-TGTCATTGATGCTGCAAACA R: CGGCAACTACGTACTGTATGGTTA	(TG) ₇	153	2	1	153	-	1	155	-	JN790260
Sfrugi29	F: PET-GCAAACGGCGATCAAAGTAT R: CGAAATGTGTTTAAAGTCACTACCT	(AC) ₇	291	4	2	290 - 307	0.25/0.49	2	308 - 311	0.50/0.38	JN790261
Sfrugi30	F: VIC-GGTGTAGGTTGTCTTGCGGT R: ACTCATTGACCCAGCGTACC	(GT) ₆	241	3	2	244 - 269	0.13/0.12	2	244 - 257	0.70/0.46	JN790262
Sfrugi39	F: FAM-GAGCTCGTTCTCCCTCACTG R: GCCCATTCGTAAGCAT	(AC) ₈	299	2	1	299	-	1	344	-	JN790267
Sfrugi50	F: VIC-ACCACCATGCGGTAGCTATT R: CACCTACCTCGTATCAGCGT	(AC) ₇	135		1	132		3	133 - 172	0.55/0.59	JN790269
Sfrugi21	F: FAM-CCTTGGGTAAGGTGAGAGGG R: CAAATCCAATGATGACCACG	(TC) ₁₀	201	3	1	202	-	2	163 - 186	0.64/0.45	JN790258
Sfrugi33	F: NED-ATTCCAAAGAGGCACCATAG R: GCAGCATGTTCAACCCACTA	(CT) ₉	139	3	2	130 - 152	0.24/0.49 **	1	162	-	JN790263
Sfrugi38	F: VIC-AAATCATTTCCCAGAAAACA R: GACAGACCATTTGACGGTGA	(AC) ₆	118	3	2	93 - 115	0.18/0.48	1	116	-	JN790266
Sfrugi43	F: NED-ACTTTGAAAAAGGCGTCGAA R: GCCCTCAACCAGACCAATAA	(TG) ₉	231	4	3	208 - 233	0.63/0.60	2	221 - 233	0.47/0.37	JN790268
Sfrugi24	F: VIC-TTACACATTCGTTTGTACACGC R: GAATGGAACAGACAGTCGTTGT	(CA) ₈	113	4	2	105 - 113	0.43/0.49	2	108-111	0.00/0.25 *	JN790259
Sfrugi58	F: NED-ATCGATGTGACACCACTCA R: TTAGTTGTGTGTGTAGAGTAAGC	(CA) ₇	132	2	1	129	-	2	108 - 129	0.43/0.36	JN790270
Sfrugi6	F: VIC-TGGAATATGTGCAGCGAAAA R: TCGTTATGCATACACCATTCAGA	(AC) ₅	324	2	1	326	-	1	322	-	JN790257
Sfrugi60	F: FAM-TGTTGCGCACTTCTACTGCT R: AATTGTTGAGGTTTGCCTCA	(CA) ₆	128	2	1	125	-	1	127	-	JN790271

2.2. Isolement reproducteur entre les deux variants de *S. frugiperda*

Article 5 en préparation pour *BMC Genetics: Spodoptera frugiperda host plant variants : two strains or two distinct species?*

***Spodoptera frugiperda* host plant variants : two strains or two distinct species?**

P. Dumas^{1,2,3*}, F. Legeai^{4,5}, S. Gimenez^{1,2}, A.-L. Clamens³, G.J. Kergoat³, and E. d'Alençon^{1,2}

Affiliations

¹UM2 - UMR 1333 DGIMI, Université Montpellier 2, Place Eugène Bataillon, 34095, Montpellier, France

²INRA - UMR 1333 DGIMI, Université Montpellier 2, Place Eugène Bataillon, 34095, Montpellier, France

³INRA - UMR 1062 CBGP (INRA, IRD, CIRAD, Montpellier SupAgro), Campus de Baillarguet, CS 30016, 34988, Montferrier-sur-Lez, France.

⁴INRA - UMR 1349 Institute of Genetics, Environment and Plant Protection, Domaine de la Motte, BP 35327, 35653, Le Rheu Cedex, France.

⁵INRIA Centre Rennes – Bretagne Atlantique, GenOuest, Campus de Beaulieu, 35042, Rennes, France.

Keywords: host-races, microsatellites, post-zygotic isolation, segregation distortion, *Spodoptera frugiperda*

* Corresponding author: Pascaline Dumas

Tel. +33 467 144 720

Fax. +33467 144 299

E-mail: pascaline.dumas@univ-montp2.fr

Abstract

The moth *Spodoptera frugiperda* (Lepidoptera: Noctuidae) is a well-known pest of crops throughout the Western hemisphere. This species consists of two strains adapted to different larval host-plants: the first feeds preferentially on corn, cotton and sorghum whereas the second is more associated with rice and several pasture grasses. Though morphologically indistinguishable, the two strains exhibit differences in their mating behavior, pheromone compositions, and present development variability according to the host plant. Though the later suggest that both strains are different species, this issue is still highly controversial because hybrids naturally occur in the wild, not to mention the discrepancies among published results concerning mating success between the two strains. In order to clarify the status of the two host plant strains of *S. frugiperda*, we analyze features that possibly reflect the level of post-zygotic isolation: (i) F1 hybrid lethality and sterility; (ii) patterns of meiotic segregation of hybrids in reciprocal second generation (F2), as compared to the meiosis of the two parental strains. We found a significant reduction of mating success in F1 in CR (a Corn female with a Rice male) cross as compared to reciprocal crosses (RC), RR, CC crosses (Rice female with Rice male and Corn female with Corn male) and a high level of microsatellite markers showing transmission ratio distortion (TRD) in the F2 progeny obtained by F1 hybrid intercrosses. Our results support the existence of post-zygotic reproductive isolation between the two laboratory strains and are in accordance with the marked level of genetic differentiation that was recovered between individuals of the two strains taken in the wild [1, 2]. Finally, by identifying the nature of the genomic regions the surrounding distorted markers we provided some new insights to discuss the basis of hybrid incompatibilities. Put altogether, these evidences are compatible with a sibling species status for the two host plant strains.

Introduction

Speciation, the process by which an ancestral lineage splits into two or more reproductively isolated lineages, is a central process in evolution. Within the biological species context ([3]), a fundamental component of this process is reproductive isolation, which may result from pre- and/or postzygotic barriers. Prezygotic barriers occur before fecundation and usually consist in differences between populations in habitat, biology or behavior. Postzygotic barriers exist between related species when fitness of hybrid genotypes is lower than those of parental genotypes. Three different kinds of postzygotic isolation between individuals from distinct species have been described [4]: (i) second-generation hybrid (F2) degeneracy, (ii) first generation hybrid (F1) lethality, and (iii) F1 hybrid sterility. Although in the former case, postzygotic isolation is not total because of the existence of gene flow, it is nonetheless considered as a standard feature of the evolution of complete sterility between species [5, 6]. Regarding the inviability or sterility of hybrids, Dobzhansky and Muller proposed that they may result from the accumulation of genes that function normally in a pure-species genome but produce epistatic interactions in hybrids [7, 8]. When hybrids between these lineages are obtained, these negative interactions can cause inviability and/or sterility in particular recombinant genotypes, such as they are removed by natural selection from hybrid populations. This nonrandom elimination of specific allelic combinations leads to segregation distortion (or transmission ratio distortion) *ie* significant deviation of allele or genotype frequencies from simple Mendelian expectations. One corollary of this explanation is that loci causing hybrid incompatibility are expected to be located at or near regions of transmission distortion in hybrid populations.

Interestingly, this corollary was formerly demonstrated by analyzing crosses between two plant species in the genus *Solanum* [9]. Even more remarkably, nonrandom elimination of specific allelic combinations has been also used to infer the genetic basis of hybrid incompatibility among species [10-13].

In this study we propose to investigate and clarify the status of a “so-called” species of noctuid moth: the fall armyworm (*Spodoptera frugiperda*). This moth is a widespread and important agricultural pest in the Western hemisphere [14, 15], which has been defined so far as one species with two plant-related strains [16-18]. One strain was originally identified from populations feeding preferentially on corn, sorghum and cotton (corn strain), while the other was identified from populations feeding preferentially on rice and on various pasture grasses (rice strain) [14, 15]. Corn and rice strains are morphologically identical but genetically distinguishable using strain-specific molecular markers [2, 19-24]. Because of the later it is possible to highlight the fact that both variants occur in sympatry [17, 25, 26]. Other molecular evidence comes from the results of molecular-based species delimitation analyses, which consistently split sequenced individuals into two putative species clusters corresponding to the two rice and corn strains (Dumas *et al.* submitted). Other analyses also indicate that the two strains have likely diverged more than 2 Myr ago [1]. Finally, several prezygotic and postzygotic incompatibilities are known for the two strains (for a review, see [27]), among which some are highly controversial: among the known prezygotic barriers is the difference in host-related performances of larvae from each strain [28, 29]. Interestingly this point was also recently disputed because of the results of some studies that have found that the rice strain larvae developed better on corn and

sorghum than corn strain larvae [27, 30]. Other known prezygotic barriers consist of behavioral isolation due to mating allochronism between the two strains [31] or to pheromone differences in females [32]. Concerning post-zygotic barriers to genes flow, Pashley & Martin (1987) [33] observed that in mating experiments between the C strain females and R strain males, no spermatophores were transferred while the reciprocal cross (RC) gave a viable offspring [33]. Other studies [29, 34] did not confirm these results, since the two strains crossed successfully in both directions. While performing backcrosses, Pashley and Martin found that the RC hybrid females mated with low success with their brothers but not at all with males of either parental strain [33]. The same result was obtained by Whitford *et al.* (1988)[29] and was recently confirmed by Groot *et al.* (2010)[27]. In the wild, hybrids between the two strains (identified as containing mt DNA from one strain and nuclear from the other) have been found at a rate reaching 16% [18]. Recently, Velasquez-Velez *et al.* (2011) [35] have found postzygotic isolation for several life-history traits in both strains. Furthermore, they have identified a reduction of the number of hybrid females and a reduction in hybrid fertility in *S. frugiperda*, suggesting a Haldane's rule (1922) [36], which corresponds to the decrease of selective value of the heterogametic sex in hybrid progeny from an interspecies cross. It has often been observed in early mechanisms of the process of speciation [37].

There is evidence of the continuous nature of speciation [38] with numerous studies [39-44] indicating that the divergence during this process varies continuously. For example, the strength of reproductive isolation can vary quantitatively along the "continuum" of speciation and groups differing by discrete levels of differentiation

can often be identified between populations and well-defined species, like host-races in the framework of ecological sympatric speciation [45-49]. Host-races were defined as "[...] genetically differentiated, sympatric populations of parasites that use different hosts and between which there is appreciable gene flow [45]. This definition seems partially congruent with features exhibited by *Spodoptera frugiperda* and led us to question: (i) the ability of the two strains to mate and reproduce, and (ii) the genetics of the resulting hybrid progeny.

Using laboratory strains of both variants, we measured mating efficiency in crosses between females of the corn strain and males of the rice strain, and vice versa. We present the first genetic analysis of reciprocal crosses between the two *S. frugiperda* strains, by following the segregation pattern of a set of microsatellite markers [2] in F2 populations. Only few markers showed a Mendelian inheritance. The distorted ones were evaluated in within-strain crosses. When distortion could be assigned specifically to inter-strain crosses, the genomic regions surrounding the markers were identified and analyzed in an attempt to infer the genetic basis of inter-strain incompatibilities. The possibility that the two host-strains correspond to two sister species is discussed.

Material and methods

We used two strains of *S. frugiperda*. Those strains have been seeded with 30 to 50 pupae ten and four years ago for the corn and rice strain, respectively. Since then, they were reared in laboratory conditions (on Poitout artificial diet [50], at 24°C with a 16:8 photoperiod and hygrometry of 40 %). The individuals that seeded the corn strain came

from Guadeloupe whereas those that seeded the rice strain came from Florida.

Measure of inter-strain versus within-strain mating efficiency and of sex-ratio of the progeny

A total of 30 isolated couples were constituted for each type of crosses: both CR (one corn female with one rice male) and RC (one rice female with one corn male) inter-strain crosses directions and both RR (one rice female with one rice male) and CC (one corn female with one corn male) within-strain crosses. In order to avoid paternity ambiguity, the sex of pupae was anatomically determined before emergence, and pupae of both sexes were reared separately. Virgin females were collected at emergence, and allowed to mate with a single male for five days. The number of couple with hatched larvae was counted and progeny of three couples for each kind of cross were reared in laboratory up to the pupal stage in order to determine sex of pupae.

In order to compare CR and RC F1 mating efficiencies, the Fisher-exact test [51] was performed using R version 2.15.1.

Between strains crosses for microsatellite markers segregation analysis

Two reciprocal F1 intercrosses were obtained (Figure 1): females from CR crosses x males from the same CR cross; and the reciprocal RC x RC cross. The number of progeny sampled from each F1 intercross was respectively 80 and 50 for the CR and RC progeny. Finally a total of 120 larvae were typed for markers, as described below. In order to avoid paternity ambiguity, sex of pupae was determined before emergence, and pupae of both sexes were reared separately. Virgin females were collected at emergence, and allowed to mate with a single male for five days. Eggs were collected and progeny was reared until the L6 larval instar

from which total genomic DNA was extracted by grinding up whole bodies using DNAeasy Blood and Tissue Kit (Qiagen). DNA quality was then assessed using a spectrophotometer (Nanodrop, Wilmington, DE, USA).

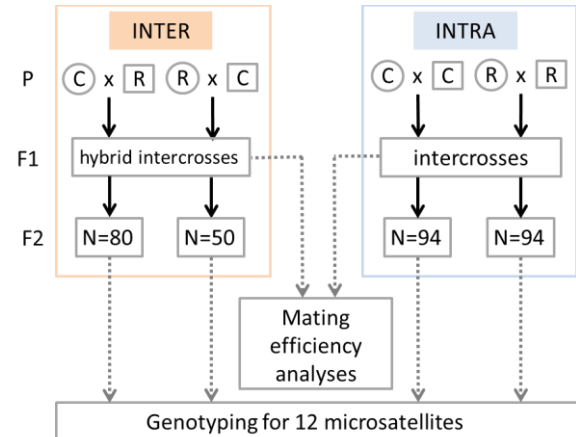


Figure 1: Crossing protocol used to follow microsatellite markers segregation patterns in F2 populations within- and inter-strains.

We used a set of 12 molecular markers recently characterized [2] to follow their segregation in two reciprocal interspecific crosses (starting from either C female x R male crosses or R female x C male at G0) at the F2 generation issued from crosses between F1 brothers and sisters. 80 individuals were genotyped for the CR population, 50 for the RC one.

Within strains crosses for microsatellite markers segregation analysis

Since polymorphism is scarcer between individuals of the same strain, this control experiment required a great number of crosses, in order to obtain G0 and F1 parents carrying different alleles of the microsatellites studied. At least 23 pairs of individuals were allowed to mate, and were genotyped at the adult stage after mating, in order to increase the chance of seeing segregation of different alleles of the microsatellites markers in the descendants. Progeny from pairs the most polymorphic at the different markers analyzed were reared until adulthood (corresponding to five pairs

(labeled A, B, C, D, E) for the CC G0 crosses and four pairs (labeled F, G, H, I) for the RR G0 crosses). At the adult stage, 9, 9, 8, 7, 1 F1 intercrosses issued from pairs A to E, respectively, for the corn strain, and 8, 8, 7, 6 F1 intercrosses issued from pairs F to I, respectively, for the rice strain were performed and genotyped. 94 descendants were genotyped from one of the F1 intercross (rice strain), 63 from a second one.

Genotyping of a set of microsatellite markers in F2 laboratory populations (from F1 brotherhood intercross).

Twelve microsatellite markers previously characterized [2] were typed. Adapting the method of Schuelke (2000), the 10 µl PCR reactions contained 4 to 10 ng of template DNA diluted in sterile Millipore water, 2 µl of 5x GoTaq reaction buffer, 1.5 mM MgCl₂, 0.2 mM of each dNTP, 0.04 µM of the forward primer (which included a 19 nucleotide tail corresponding to a sequence of the M13 bacteriophage), 0.15 µM of the reverse primer, 0.15 µM of the fluorescence marked M13 primer and 0.1 µl units of recombinant GoTaq-polymerase (5u/µl, Promega). The PCR protocol included an initial denaturation step at 94°C for 4 min, followed by 12 cycles involving denaturation at 94°C for 30 s; annealing at 60°C for 1 min and extension at 72°C for 30 s; then 25 cycles of denaturation at 94°C for 30 s; annealing at 52°C for 1 min and extension at 72°C for 30 s; with a final extension step at 72°C for 10 min. Four 96-well PCR products were simultaneously pooled and diluted to a ratio of 1:75. 2 µl of each PCR were mixed with 0.1 µl of a fluorescent size ladder (GeneScan 500 LIZ) and 15 µl of Hidi Formamide (Applied Biosystems). Electrophoresis and allele detection were carried out on an ABI 3130 automated sequencer. Output was analyzed with Genemapper v.3.7 software (Applied Biosystems, USA). All marker data were verified manually by visual inspection to

eliminate errors that may result from the automatic allele assignment procedure following [52]. To estimate genotype frequencies among members of F2 populations, chi-squared tests were performed using R version 2.15.1.

Analysis of genomic regions surrounding distorted and non-distorted markers.

Corn strain genomic scaffolds matching with microsatellite sequences were identified by BLASTN. The FGENESH program (with parameters used for the *Drosophila melanogaster* model system was then used for gene annotation (<http://linux1.softberry.com>). Illumina reads obtained from whole genomic sequencing of the rice strain were mapped on the corn reference scaffold by BOWTIE software [53]. The alignments were subsequently visualized in TABLET [54] in order to highlight nucleotidic substitutions between the two strains.

Results

Mating efficiencies in F1 crosses between and within corn and rice strains

The mating success of inter-strain crosses was measured and compared with those of within strain crosses (Figure 2). Mating was considered successful when a single pair composed of one female and one male gave rise to a viable progeny. Over 30 CC crosses, 97 % were fertile, while 83 % of RR crosses led to viable progeny (over 30 crosses). Rice females mated with corn males at a rate of success of 70% over 30 crosses, while corn females mated significantly less efficiently with rice males (p-value of 0.008), having a rate success of only 30% over 27 crosses. Except from the longer time before egg hatching, we do not notice any aberrant feature in the development of these CR larvae. Thought this asymmetry in mating efficiency has already been described by Pashley-Prowell

(1988) [33], it was not recovered by Whitford (1988)[29].

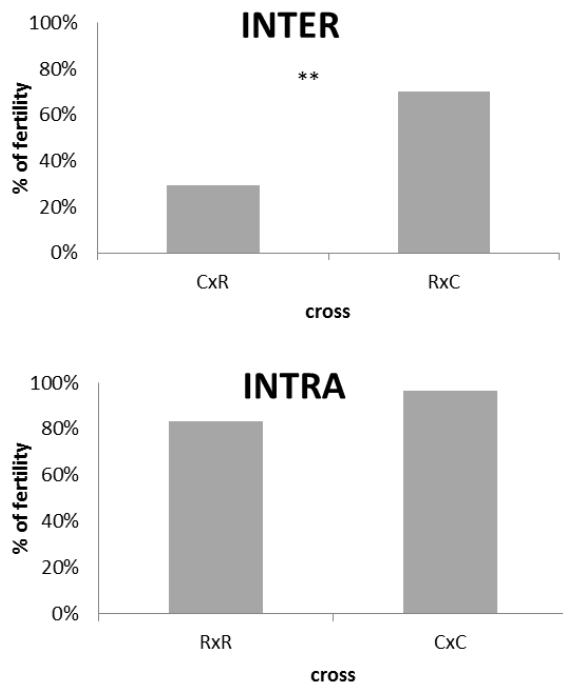


Figure 2: Mating efficiencies in inter-strain crosses in both CR (female Corn with male Rice) and RC (female Rice with male Corn) cross direction and in control within-strain crosses.

Segregation pattern of a set of microsatellite markers in F2 generation in crosses between and within corn and rice strains.

Obtaining a progeny does not preclude from the fact that some genotypes may be absent from it. We checked this hypothesis by following segregation of a set of markers (See Material and method section). The corresponding results are shown on table S1. The markers can be classified into three classes according to the way they segregate. The first four ones i.e. *Sfrugi2*, *Sfrugi33*, *Sfrugi43* and *Sfrugi76* show a Mendelian segregation for both direction of strain crossing. *Sfrugi6* also show a Mendelian segregation in RC progeny but the appearance of a highly frequent mutation for this marker in CR progeny (1.8 10⁻¹/F₂ progeny) prevent the estimation of its segregation ratio. Three of the markers (*Sfrugi37*, *Sfrugi38*, *Sfrugi50*)

departed from Mendelian expectations in one direction of the crosses only, one, *Sfrugi37* in the RC direction and the remaining ones in the reverse CR cross. Three other markers (*Sfrugi11*, *Sfrugi21*, and *Sfrugi29*) showed non Mendelian segregation in the two reciprocal crosses. The last one *Sfrugi25* could not be genotyped unambiguously.

Table 1: Synthetic table showing segregation patterns (Mendelian in green, distorted from Mendelian expectation in red) for 12 microsatellites markers within F₂ progeny from within-strain crosses (CC: female corn with male corn and RR: female rice with male rice) and inter-strain crosses in both CR (female Corn with male Rice) and RC (female Rice with male Corn) cross direction.

Marker	F2_INTER		F2_INTRA	
	CR	RC	Corn	Rice
<i>Sfrugi2</i>	✓	✓		
<i>Sfrugi33</i>	✓	✓		
<i>Sfrugi76</i>	✓	✓		
<i>Sfrugi43</i>	✓	✓	✓	✓
<i>Sfrugi6</i>	?	✓		
<i>Sfrugi37</i>	✓	×		✓
<i>Sfrugi50</i>	×	✓	✓	✓
<i>Sfrugi38</i>	×	✓	monomorphic	ambiguous
<i>Sfrugi11</i>	×	×	✓	✓
<i>Sfrugi21</i>	×	×	monomorphic	monomorphic
<i>Sfrugi29</i>	×	×	monomorphic	monomorphic
<i>Sfrugi25</i>	?	?	?	?
N total	80	50	94	94

Because a rate of 45% distorted markers at a p value < 0.05 was not expected in inter-strains crosses, we decided to test segregation of these distorted markers in intra-strains crosses, in order to check if these transmission ratio distortion (TRD) was inherent of the markers, or could reflect some inter-strains incompatibilities. We managed to control unambiguously 3 markers, *Sfrugi43*, *Sfrugi50* and *Sfrugi11* that showed a Mendelian segregation in both CC and RR crosses. *Sfrugi37* showed a Mendelian segregation in the RR cross, but was monomorphic in all CC crosses. *Sfrugi21* and *Sfrugi29* were also monomorphic in all crosses

genotyped (see Materials and Methods for details). For these latter markers, despite the fact that we could not assess their segregation pattern in intra-strain crosses, we could at least verify that the microsatellites could unambiguously be genotyped in all individuals, excluding genotyping errors as the source of segregation distortion. As a conclusion, for at least two of the markers (*Sfrugi11* and *Sfrugi50*), we can conclude that the TRD is due to inter-strain genetic incompatibilities (Table 1).

Attempt to infer the genetic basis of inter-strain incompatibility by analyzing genomic regions surrounding distorted markers

Distorted microsatellite markers may be genetically linked to genomic regions involved in Dobzhansky and Muller incompatibilities (DMI) between the two host-plant variants, which would explain their transmission ratio distortion. Using a first whole genome sequence assembly of *S. frugiperda* (The FAW Consortium, in preparation)), we used the BLASTN algorithm for homology search to identify scaffolds matching with the distorted microsatellites markers. We looked for markers for which distortion is clearly linked to inter-strain incompatibilities. These scaffolds could be identified without ambiguity, excluding the fact that these microsatellites are linked to repeated elements ([55]) (Figure 3). *Sfrugi11*, which showed TRD in both crosses direction is carried by a 10kb genomic scaffold and overlaps an intervening sequence in the gene encoding a homolog of derailed 2 (*drl2*) of *D. melanogaster* (gene prediction by FGESH). *Sfrugi37*, is part of a 19,5kb scaffold in the vicinity of a gene encoding a hypothetical KGM protein of *Danaus plexippus*. *Sfrugi50* is carried by a large (115 kb) scaffold devoid of predictions for proteins of known function, except transposable elements proteins. These three genomic regions are not genetically

linked as inferred by synteny with *Bombyx mori* where they correspond to linkage groups 5, 14, 15, respectively. The comparison between *S. frugiperda* corn and rice strains in these genomic regions, in the closest predicted gene from the microsatellite sequence, showed presence of 12 synonymous substitutions throughout the 429 bp of the coding sequence of *drl2* gene. In order to assess the importance of this substitution rate (2.8%), we conducted a similar analysis for the non-distorted microsatellite marker *Sfrugi76*. We identified, using the same analysis pipeline as previously, the scaffold matching to this microsatellite. *Sfrugi76* is carried by a 116 kb genomic scaffold and overlaps an intervening sequence in the gene encoding a homolog of *mangetout E* of *D. melanogaster*. For this gene, the comparison between *S. frugiperda* corn and rice strains showed presence of four synonymous substitutions throughout the 762 bp of the coding sequence, representing a substitution rate of 0.5%.

Discussion

In order to clarify the evolutionary status of the two host plant strains of *S. frugiperda*, we analyzed criteria that possibly reflect the level of post-zygotic isolation: F1 hybrid lethality and sterility in addition to patterns of meiotic segregation of hybrids in reciprocal second generation (F2) as compared to the meiosis of both parental strains. We found a significant reduction of mating success in F1 in CR cross as compared to RC, RR, CC crosses and a high level (45% in CR cross with $p < 0.05$, 36.6% in RC cross) of markers showing transmission ratio distortion (TRD) in the F2 progeny obtained by F1 hybrid intercrosses.

Though this bias in mating success against CR cross has been already reported by Pashley (1987) [33], it has also not been evidenced by

other studies [29, 34]. These discrepancies may be explained by some heterogeneity in populations that have been used for experimental crosses. In our study, the rice strain originates from Florida while the corn strain comes from French Guadeloupe. We cannot neglect the fact that geographic distance may increase, by drift, genetic distance between the two isolates, however we think that this geographic effect is minor since i) *Spodoptera frugiperda* is a long-

distance migrator which yearly flies from the South to the North of the USA ii) Atmospheric trajectories are favorable for the northward transport from the Caribbean to the south-East of the USA iii) The Florida main haplotypes ratio h4/h2 in *COI* gene is conserved in the Caribbean at Puerto-Rico [56]. Additional phylogeographic studies are required to definitively compare populations in these localities.

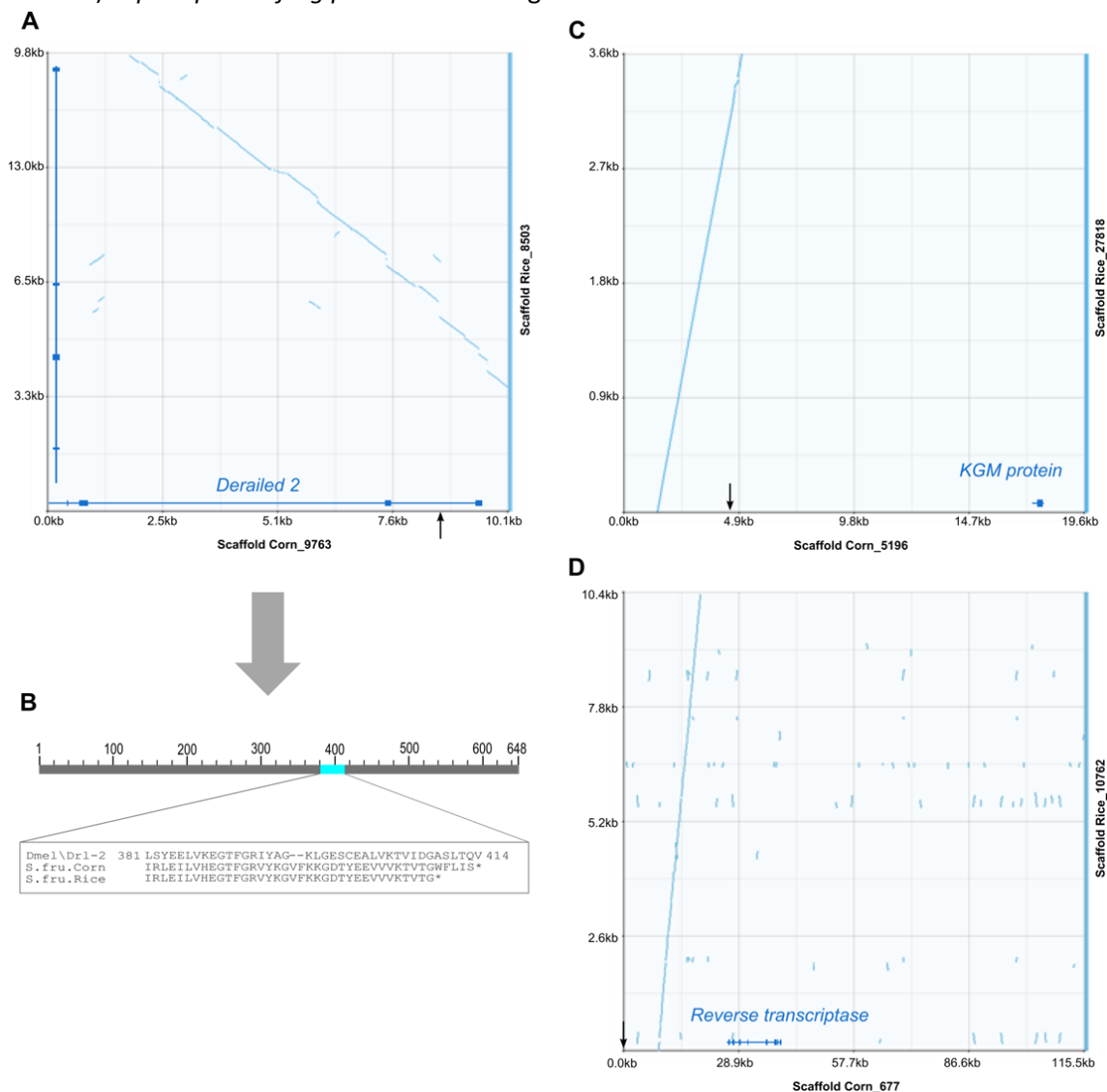


Figure 3: Dotplots resulting from alignments between rice and corn orthologous genomic regions containing distorted microsatellites *Sfrugi11*, *Sfrugi37* and *Sfrugi50* (A, C and D respectively). Black arrows indicate the position of the microsatellites. Comparison of the Derailed2 protein amino acids sequence between *S. frugiperda* Corn or Rice strain and *D. melanogaster* (B).

Moreover, a significant level of reproductive isolation between the two strains [35], has been found within individuals from natural populations sampled in central Colombia. Thus, rice and corn strains in this case are very close geographically. The decrease in mating success measured in our study reflects partial embryonic inviability of F1 hybrids obtained in the CR cross and may result from partial hybrid incompatibility due to asymmetric parental contribution. For instance maternal inheritance of mitochondria, mRNAs, proteins, and noncoding small RNAs through the maternal cytoplasm may create imbalance in hybrids with the paternally inherited genome.

Since F1 hybrids have been obtained in the two reciprocal crosses, we tried to obtain F2 generations through F1 intercross. F1 hybrids were fertile and gave rise to F2 progenies which developed normally showing no obvious phenotypic degeneracy. Nevertheless, since we obtained progeny from the two reciprocal crosses, we have followed the segregation pattern of a set of markers in order to check whether some genotypes would be absent or overrepresented. The high rate of TRD that we found is comparable to the amount of segregation distortion that has been observed within inter-species crosses in other taxa (e.g. *Nasonia spp.*, 29% of markers in adult males [57]; *Arabidopsis lyrata*, 50% of markers [58]; *Lepomis spp.* 36.8% of markers [59]). Transmission Ratio Distortion usually occurs at a lower rate in intraspecific than in interspecific crosses [60] although some exception to the rule has been documented (48% of distorted markers when crossing highly divergent populations within *Mimulus guttatus* species [61]). Absence of TRD when crossing the two *S. frugiperda* strains would have argued in favor of absence of F2 degeneracy, while the fact that we found a high level of TRD is consistent with some

hybrid incompatibility between the strains at the F2 generation

The level of TRD is known to increase with genetic distance [62] ; [63, #642]. Divergence between the two *S. frugiperda* host-plant strains has been estimated to 2.09% on average in the COI gene (Kimura 2-parameter distance) by [1]. As a comparison, 1.4% of base substitution has been found between human and chimpanzee DNA [64], between which taxa, divergence raises 4.8% when including *indels*. A rate of nucleotidic divergence of 1.9% Between *O. scapularis* and *O. nubilalis*, has been identified [65], and the two sibling species are completely reproductively isolated [66]. The amount of divergence found between the two strains of *S. frugiperda* is equivalent to the amount displayed by pairs of differentiated species in the *Spodoptera* genus ([1]; Dumas *et al.*; submitted). This divergence, in addition to prezygotic barriers to gene flow (reviewed in [27]) plus partial F1 hybrid inviability and indirect evidence of F2 hybrid degeneracy through high level of TRD make these two *S. frugiperda* strains more likely “differentiated species” than “host plant races”.

Within species, TRD can result from competition among male gametes, where sperm with a particular genotype manages to disrupt or outperform their competitors (as in the mouse t-haplotype system and the segregation distorter system in *Drosophila*, ([67]; [68])). In females, the principal opportunity for prezygotic distortion occurs during meiosis, when each primary oocyte produces one functional gamete and three polar bodies. This asymmetry provides scope for cheater genotypes to subvert the segregation process in order to improve their chances of appearing in the functional gamete. Finally, after fertilization, embryonic mortality can also lead to transmission distortion even if the rate of loss depends on the genotype. In interspecific crosses, TRD

may also result from competition among gametes, due for instance to defects in chromosome segregation during hybrid meiosis [69] [70]. TRD can also be due to inviability of embryos due to hybrid incompatibilities. Molecular basis involved in hybrid incompatibility can result from Dobzhansky-Muller diverged genes, chromosome rearrangements, sequence divergence, dosage imbalance and/or transposable elements and non-coding repeats, as recently reviewed in [12] .

Dobzhansky-Muller diverged genes model can explain the fact that F1 hybrids retain their fitness while F2 ones do not as observed with *S. frugiperda*, if one considers the fact that derived alleles are recessive compared to ancestral alleles [71]. In a two loci model, if the ancestral population *aabb* splits into two sub-populations, one acquiring allele *A* at locus *a*, that becomes fixed *AAbb*, and the others acquiring allele *B* at locus *b*, that becomes fixed *aaBB*. F1 hybrids will be *AaBb*. If *A* is incompatible with *B* but recessive, F1 hybrids will be viable, but some individuals of the F2 progeny will not, due to recombination that renders them homozygous for the two derived incompatible alleles. This model fits well with our observations of high level of TRD in F2 since it provides an explanation to absence of some genotypes in F2. Since we did not detect sex ratio bias in F1, we suppose that the incompatible loci are carried by autosomes. Conversely as Velasquez-Velez *et al.* (2011) [35], sex-ratios were not biased and we did not observe a Haldane's rule in F1 hybrid progeny.

TRD loci often cluster on regions of chromosomes that contain hybrid incompatibility genes: An approach for finding incompatibility genes consists in looking for deviation from Mendelian ratio of parental alleles in BC or F2 population. This method has been applied widely in seed-bearing plants [72], [73], and *Nasonia* wasps [74], [57].

Therefore, taking advantage of the availability of a first assembly of *S. frugiperda* genome (The FAW Consortium, in preparation), we have mapped distorted microsatellite markers. All of them could be assigned unambiguously to a unique genomic location. We report genomic location of the three markers 11, 37 and 50 for which within-strain polymorphic control crosses could be obtained and showed a Mendelian segregation, which led us to conclude that distortion was due to inter-strain incompatibilities. Marker *Sfrugi11* maps to an intervening sequence of a gene homolog to the *derailed2* gene of *D. melanogaster*. We wondered whether this gene could be involved in hybrid incompatibility, and looked at its function in early neuronal development. Neurons extend axons over comparatively vast distances to make synaptic connections with their targets. One recently uncovered axon guidance signaling pathway involves interactions between the Wnt (Wingless Integration site) signaling and the Receptor Tyrosine kinase-related tyrosine kinase (Ryk)-like trans-membrane receptor proteins. DRL is a receptor for the Wnt protein, WNT5. DRL binds WNT5 and *drl* and *Wnt5* interact during the formation of the embryonic central nervous system (see [75] for review on these receptors). DRL-2 is another receptor which competes with DRL for WNT5 binding at least in the antennal-lobes, but probably not only there since it is expressed in other cell types during early development in *Drosophila*. *Dr2* and DRL cooperate to establish the olfactory circuitry in *Drosophila* [76]. If this gene is involved in hybrid incompatibility, we expect that the two host-plant strains orthologous proteins should diverge or that the regulatory regions of the gene diverge. When we compared their predicted exonic sequences, we found only synonymous mutations. The scaffold is short (10 kb) and the regulatory region is not available. The rate of

synonymous substitution we found between the two orthologs was 2,8 %. When we compared between the two strains, the closest predicted gene in the vicinity of the non-distorted microsatellite marker *Sfrugi76*, we found a rate of 0.5%. Nucleotidic divergence is thus higher in the vicinity of the distorted marker. Further work is required to show DRL-2 may be involved in hybrid incompatibility. Marker *Sfrugi37* maps to a scaffold encoding the homolog of a hypothetical protein of *Danaus plexippus*, and marker *Sfrugi50* is part of a scaffold carrying unknown genes and a Reverse Transcriptase gene from *Danaus plexippus*. The on-going *Spodoptera frugiperda* genome project, including genomic comparison of the two host-plant strains will bring further light on these genomic regions and their level of differentiation between the two strains.

References

1. Kergoat GJ, Prowell DP, Le Ru BP, Mitchell A, Dumas P, Clamens A-L, Condamine FL, Silvain J-F: **Disentangling dispersal, vicariance and adaptive radiation patterns: A case study using armyworms in the pest genus *Spodoptera* (Lepidoptera: Noctuidae)**. *Molecular Phylogenetics and Evolution* 2012, **65**(3):855-870.
2. Molecular Ecology Resources Primer Development C, Arias MC, Arnoux E, Bell JJ, Bernadou A, Bino G, Blatrix R, Bourguet D, Carrea C, Clamens A-L *et al*: **Permanent Genetic Resources added to Molecular Ecology Resources Database 1 December 2011 – 31 January 2012**. *Molecular Ecology Resources* 2012, **12**(3):570-572.
3. Mayr E (ed.): **Systematics and the origin of species, from the viewpoint of a zoologist**. New York; 1942.
4. Dobzhansky T (ed.): **Genetics of the Evolutionary process**. New York: Columbia University; 1970.
5. Coyne J, Orr HA (eds.): **Speciation**. Massachusetts: Sunderland; 2004.
6. Coyne JA, Orr HA: **Patterns of Speciation in *Drosophila***. *Evolution* 1997, **51**(1):295-303.
7. Dobzhansky T: **Genetics and the Origin of Species** In. Edited by Press CU. New York; 1937.
8. Muller HJ: **Isolating mechanisms, evolution and temperature**. *Biol Symp* 1942, **6**(71).
9. Moyle LC, Graham EB: **Genome-Wide Associations Between Hybrid Sterility QTL and Marker Transmission Ratio Distortion**. *Molecular Biology and Evolution* 2006, **23**(5):973-980.
10. Harushima Y, Nakagahra M, Yano M, Sasaki T, Kurata N: **A genome-wide survey of reproductive barriers in an intraspecific hybrid**. *Genetics* 2001, **159**(2):883-892.
11. Li Z, Pinson SR, Paterson AH, Park WD, Stansel JW: **Genetics of hybrid sterility and hybrid breakdown in an intersubspecific rice (*Oryza sativa* L.) population**. *Genetics* 1997, **145**(4):1139-1148.
12. Maheshwari S, Barbash DA: **The genetics of hybrid incompatibilities**. *Annu Rev Genet* 2011, **45**:331-355.
13. Myburg AA, Vogl C, Griffin AR, Sederoff RR, Whetten RW: **Genetics of postzygotic isolation in Eucalyptus: whole-genome analysis of barriers to introgression in a wide interspecific cross of *Eucalyptus grandis* and *E. globulus***. *Genetics* 2004, **166**(3):1405-1418.
14. Barros EM, Torres JB, Ruberson JR, Oliveira MD: **Development of *Spodoptera frugiperda* on different hosts and damage to reproductive structures in cotton**. *Entomologia Experimentalis et Applicata* 2010, **137**(3):237-245.
15. Pogue MG: **World revision of the genus *Spodoptera* Guenée (Lepidoptera: Noctuidae)**. *Memoirs of the American Entomological Society* 2002, **43**:1-202.
16. Meagher R, Nagoshi R, Stuhl C: **Oviposition Choice of Two Fall**

- Armyworm (Lepidoptera: Noctuidae) Host Strains.** *Journal of Insect Behavior* 2011;1-11.
17. Pashley DP: **Host-associated genetic differentiation in fall armyworm (Lepidoptera: Noctuidae): a sibling species complex?** *Annals of the Entomological Society of America* 1986, **79**:898-904.
 18. Prowell DP, McMichael M, Silvain J-F: **Multilocus genetic analysis of host use, introgression, and speciation in host strains of fall armyworm (Lepidoptera: Noctuidae).** *Annals of the Entomological Society of America* 2004, **97**(5):1034-1044
 19. Levy HC, Garcia-Maruniak A, Maruniak JE: **Strain identification of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) insects and cell line: PCR-RFLP of cytochrome oxidase C subunit I gene.** *The Florida Entomologist* 2002, **85**(1):186-190.
 20. Lu YJ, Adang MJ: **Distinguishing fall armyworm (Lepidoptera: Noctuidae) strain using a diagnostic mitochondrial DNA marker.** *Florida Entomologist* 1996, **79**(1):48-55.
 21. Lu YJ, Kochert GD, Isenhour DJ, Adang MJ: **Molecular characterization of a strain-specific repeated DNA sequence in the fall armyworm *Spodoptera frugiperda* (Lepidoptera: Noctuidae).** *Insect Molecular Biology* 1994, **3**(2):123-130.
 22. McMichael M, Pashley Prowell DP: **Differences in amplified fragment-length polymorphisms in fall armyworm (Lepidoptera: Noctuidae) host strain.** *Annals of the Entomological Society of America* 1999, **92**(2):175-181.
 23. Meagher RL, Gallo-Meagher M: **Identifying host strains of fall armyworm (Lepidoptera : Noctuidae) in Florida using mitochondrial markers.** *Florida Entomologist* 2003, **86**(4):450-455.
 24. Nagoshi RN, Meagher RL: **FR Tandem-Repeat Sequence in Fall Armyworm (Lepidoptera: Noctuidae) Host Strains.** *Annals of the Entomological Society of America* 2003, **96**(3):329-335.
 25. Machado V, Wunder M, Baldissera VD, Oliveira JV, Fiúza LM, Nagoshi RN: **Molecular Characterization of Host Strains of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in Southern Brazil.** *Annals of the Entomological Society of America* 2008, **101**(3):619-626.
 26. Pair SD, Raulston JR, Sparks AN, Westbrook JK, Douce GK: **Fall Armyworm Distribution and Population Dynamics in the Southeastern States.** *The Florida Entomologist* 1986, **69**(3):468-487.
 27. Groot AT, M. M, Heckel DG, Schöfl G: **The roles and interactions of reproductive isolation mechanisms in fall armyworm (Lepidoptera: Noctuidae) host strains.** *Ecological Entomology* 2010, **35**(1):105-118.
 28. Pashley DP: **Current Status of Fall Armyworm Host** *Florida Entomologist* 1988, **71**(3):227-234.
 29. Whitford F, Quisenberry SS, Riley TJ, JW L: **Oviposition preference, mating compatibility, and development of two fall armyworm strains.** *The Florida Entomologist* 1988, **71**(3):234-243.
 30. Meagher RL, Nagoshi RN, Stuhl C, Mitchell ER: **Larval development of Fall armyworm (Lepidoptera: Noctuidae) on different cover crop plants.** *Florida Entomologist* 2004, **87**(4):454-460.
 31. Schöfl G, Dill A, Heckel DG, Groot AT: **Allochronic Separation versus Mate Choice: Nonrandom Patterns of Mating between Fall Armyworm Host Strains.** *The American Naturalist* 2011, **177**(4):470-485.
 32. Groot AT, Marr M, Schöfl G, Lorenz S, Svatos A, Heckel DG: **Host strain specific sex pheromone variation in *Spodoptera frugiperda*.** *Frontiers in Zoology* 2008, **5**(1):20.
 33. Pashley Prowell DP, Martin JA: **Reproductive incompatibility between host strains of the Fall armyworm (Lepidoptera: Noctuidae).**

- Annals of the Entomological Society of America* 1987:731-733.
34. Quisenberry SS: **Fall armyworm (Lepidoptera: Noctuidae) host strain reproductive compatibility.** *Florida Entomologist* 1991, **74**(2):194-199.
 35. Velásquez-Vélez MI, Saldamando-Benjumea CI, Ríos-Diez JD: **Reproductive Isolation between Two Populations of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) Collected in Corn and Rice Fields from Central Colombia.** *Annals of the Entomological Society of America* 2011, **104**(4):826-833.
 36. Haldane J: **Sex ratio and unisexual sterility in hybrid animals.** *Journal of Genetics* 1922, **12**(2):101-109.
 37. Presgraves DC: **Patterns of postzygotic isolation in Lepidoptera.** *Evolution* 2002, **56**(6):1168-1183.
 38. Nosil P (ed.): **Ecological Speciation.** Oxford: Oxford University; 2012b.
 39. de Queiroz K: **Species concepts and species delimitation.** *Syst Biol* 2007, **56**:879-886.
 40. Funk DJ, Nosil P, Etges WJ: **Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(9):3209-3213.
 41. Nosil P, Harmon L, Seehausen O: **Ecological explanations for (incomplete) speciation.** *Trends Ecol Evol* 2009, **24**:145 - 156.
 42. Nosil P, Vines TH, Funk DJ: **Reproductive isolation caused by natural selection against immigrants from divergent habitats.** *Evolution* 2005, **59**(4):705-719.
 43. Peccoud J, Ollivier A, Plantegenest M, Simon J-C: **A continuum of genetic divergence from sympatric host races to species in the pea aphid complex.** *Proceedings of the National Academy of Sciences* 2009, **106**(18):7495-7500.
 44. Rundle HD, Nosil P: **Ecological speciation.** *Ecology Letters* 2005, **8**:336-352.
 45. Berlocher SH, Feder JL: **Sympatric speciation in phytophagous insects: moving beyond controversy.** *Annual Review in Entomology* 2002, **47**:773-815.
 46. Blair CP, Abrahamson WG, Jackman JA, Tyrrell L: **Cryptic speciation and host-race formation in a purportedly generalist tumbling flower beetle.** *Evolution* 2005, **59**(2):304-316.
 47. Drès M, Mallet J: **Host races in plant-feeding insects and their importance in sympatric speciation.** *The Royal Society* 2002, **357**:471-492.
 48. Matsubayashi KW, Ohshima I, Nosil P: **Ecological speciation in phytophagous insects.** *Entomologia Experimentalis et Applicata* 2010, **134**(1):1-27.
 49. Thomas Y, Bethenod M-T, Pelozuelo L, Frérot B, Bourguet D: **Genetic isolation between two sympatric host-plant races of the european corn borer, *Ostrinia nubilalis* Hubner I. sex pheromone, moth emergence timing, and parasitism.** *Evolution* 2003, **57**(2):261-273.
 50. Poitout S, Bues R: **Linolenic acid requirements of lepidoptera Noctuidae Quadrifinae Plusiinae: *Chrysodeixis chalcites* Esp, *Autographa gamma* L., *Macdunnoughia confusa* Stph., *Trichoplusia ni* Hbn. reared on artificial diets.** *Annales de la nutrition et de l'alimentation* 1974, **28**:173-187.
 51. Fisher R: **Statistical Methods and Scientific Induction.** *Journal of the Royal Statistical Society Series B (Methodological)* 1955, **17**(1):69-78.
 52. Piffaretti J, Vanlerberghe-Masutti F, Tayeh A, Clamens A-L, D'Acier AC, Jousselin E: **Molecular phylogeny reveals the existence of two sibling species in the aphid pest *Brachycaudus helichrysi* (Hemiptera: Aphididae).** *Zoologica Scripta* 2012, **41**(3):266-280.
 53. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA**

- sequences to the human genome. *Genome Biol* 2009, **10**(R25).
54. Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, Shaw PD, Marshall D: **Using Tablet for visual exploration of second-generation sequencing data.** *Briefings in Bioinformatics* 2013, **14**(2):193-202.
55. Tay WT, Behere GT, Batterham P, Heckel DG: **Generation of microsatellite repeat families by RTE retrotransposons in lepidopteran genomes.** *BMC Evol Biol* 2010, **10**:144.
56. Nagoshi RN, Meagher RL, Hay-Roe M: **Inferring the annual migration patterns of fall armyworm (Lepidoptera: Noctuidae) in the United States from mitochondrial haplotypes.** *Ecology and Evolution* 2012b, **2**(7):1458-1467.
57. Niehuis O, Judson AK, Gadau J: **Cytoneuclear genic incompatibilities cause increased mortality in male F2 hybrids of *Nasonia giraulti* and *N. vitripennis*.** *Genetics* 2008, **178**(1):413-426.
58. Kuitinen H, de Haan AA, Vogl C, Oikarinen S, Leppala J, Koch M, Mitchell-Olds T, Langley CH, Savolainen O: **Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *A. thaliana*.** *Genetics* 2004, **168**(3):1575-1584.
59. Lopez-Fernandez H, Bolnick DI: **What causes partial F1 hybrid viability? Incomplete penetrance versus genetic variation.** *PLoS One* 2007, **2**(12):e1294.
60. Liu X, Guo L, You J, Liu X, He Y, Yuan J, Liu G, Feng Z: **Progress of segregation distortion in genetic mapping of plants.** *Research journal of agronomy* 2010, **4**(4):78-83.
61. Hall MC, Willis JH: **Transmission ratio distortion in intraspecific hybrids of *Mimulus guttatus*: implications for genomic divergence.** *Genetics* 2005, **170**(1):375-386.
62. Leppala J, Bokma F, Savolainen O: **Investigating Incipient Speciation in *Arabidopsis lyrata* from Patterns of Transmission Ratio Distortion.** *Genetics* 2013, **194**(3):697-708.
63. Matsubara K, Ebana K, Mizubayashi T, Itoh S, Ando T, Nonoue Y, Ono N, Shibaya T, Ogiso E, Hori K *et al*: **Relationship between transmission ratio distortion and genetic divergence in intraspecific rice crosses.** *Mol Genet Genomics* 2011, **286**(5-6):307-319.
64. Britten RJ: **Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels.** *Proc Natl Acad Sci U S A* 2002, **99**(21):13633-13635.
65. Gschloessl B, Beyne E, Audiot P, Bourguet D, Streiff R: **De novo transcriptomic resources for two sibling species of moths: *Ostrinia nubilalis* and *O. scapularis*.** *BMC Research Notes* 2013, **6**(1):73.
66. Malausa T, Dalecky A, Ponsard S, Audiot P, Streiff R, Chaval Y, Bourguet D: **Genetic structure and gene flow in French populations of two *Ostrinia* taxa: host races or sibling species?** *Molecular Ecology* 2007, **16**(20):4210-4222.
67. Lyttle TW: **Segregation distorters.** *Annu Rev Genet* 1991, **25**:511-557.
68. Montchamp-Moreau C, Ogereau D, Chaminade N, Colard A, Aulard S: **Organization of the sex-ratio Meiotic Drive Region in *Drosophila simulans*.** *Genetics* 2006, **174**(3):1365-1371.
69. Henikoff S, Ahmad K, Malik HS: **The centromere paradox: stable inheritance with rapidly evolving DNA.** *Science* 2001, **293**(5532):1098-1102.
70. Henikoff S, Malik HS: **Centromeres: selfish drivers.** *Nature* 2002, **417**(6886):227.
71. Turelli M, Orr HA: **Dominance, Epistasis and the Genetics of Postzygotic Isolation.** *Genetics* 2000, **154**(4):1663-1679.
72. Harushima Y, Nakagahra M, Yano M, Sasaki T, Kurata N: **Diverse variation of reproductive barriers in three intraspecific rice crosses.** *Genetics* 2002, **160**(1):313-322.

73. Xu Y, Zhu L, Xiao J, Huang N, McCouch SR: **Chromosomal regions associated with segregation distortion of molecular markers in F2, backcross, doubled haploid, and recombinant inbred populations in rice (*Oryza sativa* L.)**. *Mol Gen Genet* 1997, **253**(5):535-545.
74. Gadau J, Page RE, Jr., Werren JH: **Mapping of hybrid incompatibility loci in *Nasonia***. *Genetics* 1999, **153**(4):1731-1741.
75. Fradkin LG, Dura JM, Noordermeer JN: **Ryks: new partners for Wnts in the developing and regenerating nervous system**. *Trends Neurosci* 2010, **33**(2):84-92.
76. Sakurai M, Aoki T, Yoshikawa S, Santschi LA, Saito H, Endo K, Ishikawa K, Kimura K, Ito K, Thomas JB *et al*: **Differentially expressed Drl and Drl-2 play opposing roles in Wnt5 signaling during *Drosophila* olfactory system development**. *J Neurosci* 2009, **29**(15):4972-4980.

Supplementary

Table S1

Cross #	Parents		Progeny			Expected ratio	p-value
	Male	Female	AA	AB	BB		
CR	AB	BB	-	24	52	1:1	0,001
RC	AA	BB	18	15	6	1:2:1	0,009
IntraC	AB	AB	13	49	28	1:2:1	0,058
IntraR	AB	AA	28	35	-	1:1	0,378

Cross #	Parents		Progeny			Expected ratio	p-value
	Male	Female	AA	AB	BB		
CR	AB	AB	4	5	63	1:2:1	<2,2e-16
RC	AB	AB	4	24	12	1:2:1	0,091
IntraR	CC	CC	-	-	-	94	1

Cross #	Parents		Progeny			Expected ratio	p-value
	Male	Female	AA	AB	BB		
CR	AB	AB	-	16	57	1:1	<2,2e-16
RC	BB	BB	8	12	21	1:2:1	0,0004

Sfrugi6 Cross #	Parents		Progeny							Expected ratio	p-value	
	Male	Female	AA	AB	BB	AC	DE	DB	EB			EE
	CR	AB	AB	8	31	22	14	-	-			-
RC	DE	EB	-	-	-	-	12	10	14	5	1:1:1:1	0,225

Sfrugi37 Cross #	Parents		Progeny		Expected ratio	p-value
	Male	Female	AB	BB		
	CR	AB	BB	32		
RC	AB	BB	9	33	1:1	0,0002
IntraC	BB	BB	-	94	1	1
IntraR	AB	BB	49	45	1:1	0,680

Sfrugi50 Cross #	Parents		Progeny							Expected ratio	p-value	
	Male	Female	AC	AB	BC	BB	CC	AA	DA			DE
	CR	AB	CB	12	4	13	21	-	-			-
RC	CB	CB	-	-	22	12	5	-	-	-	2:1:1	0,207
IntraC	DA	DE	-	-	-	-	-	-	56	38	1:1	0,063
IntraR	CA	CA	35	-	-	-	19	18	-	-	2:1:1	0,959

Sfrugi38 Cross #	Parents		Progeny				Expected ratio	p-value	
	Male	Female	AB	AC	BB	BC			CC
	CR	AB	BC	4	12	28			30
RC	BC	BC	-	8	-	15	-	1:1:2	0,12
IntraC	CC	CC	-	-	-	-	29	1	1

Sfrugi11

Cross #	Parents		Progeny			Expected ratio	p-value
	Male	Female	AA	AB	BB		
CR	AB	BB	-	24	52	1:1	0,001
RC	AA	BB	18	15	6	1:2:1	0,009
IntraC	AB	AB	13	49	28	1:2:1	0,058
IntraR	AB	AA	28	35	-	1:1	0,378

Sfrugi21

Cross #	Parents		Progeny			Expected ratio	p-value
	Male	Female	AA	AB	BB		
CR	AB	AB	4	5	63	1:2:1	<2,2e-16
RC	AB	AB	4	24	12	1:2:1	0,091
IntraR	CC	CC	-	-	-	1	1

Partie 4 : Conclusion Générale et Perspectives

PARTIE 4 : Conclusion générale et Perspectives

La question centrale de ce travail était d'évaluer le niveau de différenciation génétique entre les deux variants adaptatifs de la noctuelle *Spodoptera frugiperda* afin d'essayer de clarifier le statut d'espèce pour ces deux variants. Les résultats obtenus à partir des différentes échelles d'analyses et présentés précédemment (Partie 2) ont permis de mettre en avant un niveau élevé de différenciation génétique entre les deux variants de *S. frugiperda*, qui suggère que ces deux entités se situent à une étape avancée dans le continuum de la spéciation. L'ensemble des résultats sur populations naturelles (Partie 3, Chapitre 1) supporte l'existence de deux groupes génétiques bien distincts correspondants aux deux variants de *S. frugiperda*. Ainsi, les analyses phylogénétiques pour plusieurs marqueurs mitochondriaux et nucléaires mettent en évidence la présence de deux lignées distinctes, ayant potentiellement divergé il y a plus de 2 millions d'années. Les analyses de délimitations moléculaires d'espèces vont encore plus loin, suggérant que ces deux lignées correspondent à des espèces putatives. Les analyses de phylogéographie et de génétique des populations vont également dans ce sens, via la mise en évidence d'un niveau de structuration statistiquement significatif entre populations naturelles des deux variants. Ces résultats sont soutenus par ceux obtenus sur les populations d'élevages (Partie 3, chapitre 2). En effet la présence d'un isolement reproducteur a été identifié entre les deux variants, ce qui va dans le sens des études précédentes menées sur le sujet (Pashley Prowell et Martin 1987; Groot *et al.* 2010; Velásquez-Vélez *et al.* 2011). De plus, pour la première fois chez *S. frugiperda*, une étude de génétique sur la ségrégation de marqueurs a été réalisée (article 5), laquelle met en évidence des distorsions de ségrégation dans les descendances hybrides reflétant sans ambiguïté une incompatibilité hybride. Toutefois, même si l'ensemble de ces résultats permettent de clarifier le statut des deux variants, ils soulèvent par ailleurs un certain nombre de questions en ce qui concerne l'origine de la divergence ainsi que sur la nature des mécanismes impliqués dans ce processus. Pour l'instant, la plupart de ces questions restent sans réponse et seules des hypothèses peuvent être proposées.

Une de ces questions concerne l'origine de la divergence entre les deux variants. La première étude sur la phylogénie du genre *Spodoptera* (article 1) supporte l'hypothèse selon laquelle la divergence des deux variants remonte à plus de 2 millions d'années. Cette date suggère ainsi que les deux variants auraient divergé bien avant l'introduction du riz ou la

domestication du maïs. Il est donc difficile d'imaginer un mécanisme de spéciation impliquant un changement de plante hôte vers une plante nouvellement introduite, comme cela est généralement supposé chez les insectes phytophages (Berlocher et Feder 2002; Drès et Mallet 2002; Matsubayashi *et al.* 2010) et comme cela semble être le cas pour les races d'hôtes du lépidoptère *Ostrinia* (Malausa *et al.* 2005). Deux scénarios complémentaires peuvent être imaginés. Le premier envisage une spéciation en allopatrie il y a plus de 2 millions d'années avec un contact secondaire correspondant à la sympatrie que l'on observe aujourd'hui, comme c'est le cas par exemple pour les deux races d'hôte du diptère *Rhagoletis* (Feder *et al.* 2005). Le deuxième scénario fait l'hypothèse d'une spéciation par renforcement (Blair *et al.* 2005) (Figure 27), impliquant plusieurs étapes successives. La première étape correspond à l'existence initiale d'incompatibilités post-zygotiques entre les deux populations (pour *S. frugiperda* ce serait le variant riz et le variant maïs). Cette divergence pouvant se produire en sympatrie ou en allopatrie, ce qui n'exclue pas le premier scénario. La deuxième étape correspond au renforcement à proprement parler et débute par un isolement pré-zygotique faible ou inexistant. Dans une zone de sympatrie (secondaire ou non), les individus des deux populations peuvent se reproduire aléatoirement et comme il existe un début d'isolement post-zygotique, les croisements entre les individus des deux populations vont produire des hybrides présentant un désavantage sélectif par rapport aux individus parentaux. Cette sélection contre les hybrides induit de façon indirecte une sélection contre l'hybridation favorisant alors une sélection des individus homogames. Un déséquilibre de liaison peut alors s'établir entre les gènes responsables de l'incompatibilité hybride et ceux responsables de la reproduction homogame. Au cours du temps la fréquence des individus homogames au sein des deux populations va alors augmenter, induisant une augmentation de l'isolement pré-zygotique. Le phénomène peut alors se poursuivre dans le temps et mener à la dernière étape, qui correspond à un isolement reproducteur complet entre les deux populations. Cette hypothèse pourrait s'appliquer au cas de *S. frugiperda* avec un renforcement de l'isolement pré-zygotique au niveau par exemple de la reconnaissance des phéromones sexuelles ou bien au niveau du choix de la plante hôte lors de l'oviposition. Dans le cas de *S. frugiperda* la dernière étape ne serait pas encore atteinte, et une étude permettant de regarder si des grandes familles de gènes tels que les gènes de l'olfaction ou de détoxification ou encore de digestion concernés par ces traits - reconnaissance sexuelle et choix de la plante hôte - divergent entre les variants.

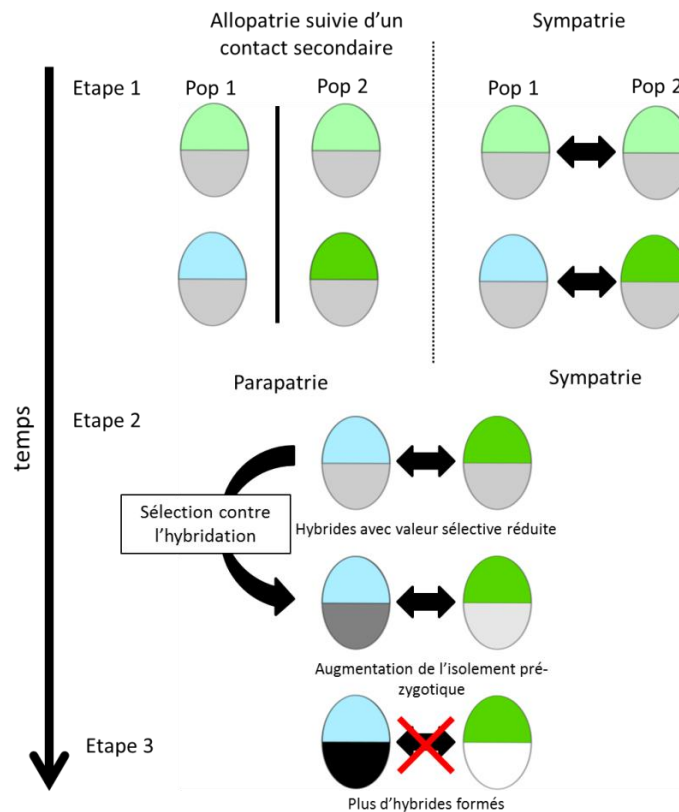


Figure 27 : Représentation schématique de la spéciation par renforcement (d'après Thomas *et al.* 2010, pp107).

Ce dernier scénario amène à se poser une seconde question, qui concerne les hybrides. Contrairement, au cas du lépidoptère *Ostrinia* - où moins de 5% d'hybrides ont été retrouvés dans la nature et pour lequel le flux de gènes semble négligeable (Malausa *et al.* 2007) - pour *S. frugiperda*, des études ont mis en évidence un pourcentage élevé d'hybrides dans la nature : 14% au Brésil (Prowell *et al.* 2004) et 16% en Amérique du Nord et en Amérique Centrale (Nagoshi *et al.* 2009), ce qui laisse penser que l'hybridation est répandue et donc le flux de gènes important entre les deux variants dans la nature, ce qui apparaît contradictoire avec les résultats de génétique des hybrides exposés dans cette thèse (article 5). Cependant aucune étude sur la valeur sélective de ces hybrides de populations naturelles n'a été menée. Quel est le devenir de ces hybrides ? S'ils arrivent au stade adulte, sont-ils fertiles ? Et arrivent-ils à migrer aussi bien que les individus parentaux ? en effet la question de la capacité des hybrides à migrer aussi bien que les individus de variant « pur » a été soulevée récemment après l'observation d'un pourcentage très faible voir nul d'hybrides

dans les populations de Floride connues pour être un des lieux de fin de migration des adultes (Nagoshi *et al.* 2012b). Il semble donc capital d'orienter les recherches sur l'étude de ces hybrides, notamment via des analyses de flux de gènes entre des populations proche de *S. frugiperda* variant riz et maïs afin de mieux comprendre la nature de l'hybridation entre les deux variants.

D'autres hypothèses sur la nature de l'incompatibilité génétique entre les deux variants peuvent être émises, telle que la présence d'incompatibilités cyto-nucléaire liées à la présence de bactéries endosymbiotiques différentes entre les deux variants et incompatibles chez les hybrides ou encore par la présence d'éléments transposables dérégulés chez les hybrides, ou par dérégulation de l'expression des gènes. Dans ce contexte, la comparaison des transcriptomes et des génomes de *S. frugiperda* riz et maïs constitue un outil essentiel et puissant pour vérifier ces hypothèses. L'analyse des génomes va permettre d'identifier les éléments transposables, et l'analyse comparée des transcriptomes entre variants et hybrides permettra de voir si certains sont dérégulés. De même, l'analyse comparée des transcriptomes des hybrides et des variants parentaux devrait permettre d'identifier chez les hybrides un défaut de régulation de gènes, autres que ceux des éléments transposables.

Enfin, la comparaison de génome hybrides avec les génomes du variant riz et du variant maïs apporterait sans doute des informations essentielles pour répondre à certaines des questions sur l'architecture génomique des hybrides, et surtout sur la distribution de la divergence le long du génome des deux variants parentaux. Les zones de différenciation impliquent-elles des régions bien définies (îlots de différenciation) ou concernent-elles plutôt de grandes régions du génome (continents) ? De plus, une dernière question se pose sur l'impact du génome holocentrique de *S. frugiperda* sur les réarrangements. En effet, lorsqu'il y a flux de gènes entre deux populations différentes, les génomes vont tendre à être homogénéisés par la recombinaison, qui devient alors un frein à la divergence. Cependant certaines régions moins accessibles du génome sont moins soumises à cette recombinaison, comme par exemple les régions centromériques. Pour les organismes holocentriques, le centromère est réparti le long du chromosome et donc plusieurs régions de ce chromosome pourraient être préservées de la recombinaison, créant alors des régions de différenciation en « mosaïque ». Dans ce contexte, il peut être fait

l'hypothèse que cette architecture génomique particulière puisse influencer le patron de distribution de la divergence entre les populations.

Au cours de ces trois années de travail, la nature intégrative de cette thèse, en appréhendant la question sous différents angles de vues et en croisant les différentes sources d'informations, a permis d'apporter des réponses nouvelles sur le modèle de *Spodoptera frugiperda*. Toutefois, de nombreuses interrogations ont été soulevées et différentes hypothèses ont été émises en ce qui concerne la nature et les mécanismes évolutifs impliqués dans la différenciation de ces deux variants et il semble évident que le chemin est encore long avant de pouvoir répondre à ces questions.

Annexes

Protocoles de biologie moléculaire utilisés pour caractériser génétiquement les variants de *S. frugiperda*.

Extraction d'ADN

L'ADN génomique total est extrait à partir d'une patte d'individu unique pour les adultes et à partir d'une 50 mg environ de tissus broyés pour les larves, grâce au kit d'extraction DNeasy 96 Blood & Tissue Kit (Qiagen®) en plaques 96 puits dans 100 µl de tampon d'extraction. Tous les échantillons d'ADN sont conservés à – 20°C.

Amplification de la séquence *FR* (*fall armyworm Rice strain*)

La réaction de PCR de la séquence *FR* est réalisée dans un volume final de 50 µL contenant 10µL de tampon de réaction 10x contenant du MgCl₂ (Promega, Madison, WI), 1 µL dNTP à 10mM (Promega, Madison, WI), 5µL d'amorce à 5µM, 0.25 µL de GoTaq DNA polymérase (Promega, Madison, WI) et de 3 µL d'ADN génomique concentré à 4ng/µL. L'amplification de la séquence *FR* et des autres fragments utilise les amorces *EA126* (5'-GAATTCGTGTAACGTAACGTTTC-3') et *EA127* (5'-GACATAGAAGAGCACGTTTT-3') et débute par une incubation initiale à 94°C (2 min), suivit de 40 cycles de dénaturation à 94°C (30 sec), hybridation à 56°C (45 sec) et élongation à 72°C (30 sec), puis une élongation finale à 72°C (2 min). 20µL du produit de PCR sont déposés sur gel d'agarose à 2%. Les échantillons possédant un faible nombre de la répétition *FR* (*FR*⁰) produisent de 0 à 3 bandes relativement faibles (Nagoshi et Meagher, 2003). La présence de bande supérieure à 500pb est caractéristiques des échantillons *FR*⁺, c'est-à-dire des individus du variant riz.

Amplification du fragment *COI*-RFLP

L'amplification du fragment du gène *COI* est réalisée dans un volume final de 25µL contenant 2.5µL de tampon de réaction 10x (Thermo Scientific), 0.5 µL de MgCl₂ 25mM (Qiagen), 1µL de dNTP à 2.5mM (Qiagen), 0.25µL d'amorces à 100 µM, 0.2 µL de *DreamTaq* DNA polymérase (Thermo Scientific) et de 2µL d'ADN génomique concentré à 40 ng/µL. L'amplification du fragment *COI* utilise les amorces *JM76* (5'-GAGCTGAATTAGG(G/A)ACTCCAGG-3') et *JM77* (5'-ATCACCTCC(A/T)CCTGCAGGATC-3') et débute par une incubation initiale à 94°C (2 min), suivit de 35 cycles de dénaturation à 94°C

(1 min), hybridation à 54°C (1 min) et élongation à 72°C (1 min), puis une élongation finale à 72°C (10 min). 2 µL du produit de PCR sont ensuite prélevés pour être utilisés dans la réaction de digestion enzymatique par MspI (Promega, Madison, WI), qui est réalisée dans un volume final de 20 µL où est ajouté 2 µL de tampon 10x, 0.2µL de BSA et 1µ d'enzyme de restriction MspI (Promega, Madison, WI). Chaque réaction est incubée 2h à 37°C puis 15 min à 60°C pour inactiver l'enzyme et arrêter la digestion. 5µL de réaction sont alors déposés sur gel d'agarose à 1.5%. Les échantillons correspondant au variant riz présenteront une bande PCR de 569 pb, alors que les échantillons correspondant au variant maïs présenteront une bande de 497 pb et une autre de 72 pb.

Amplification des marqueurs microsatellites par PCR marqué avec la technique des queues M13.

Les amorces permettant l'amplification des microsatellites (Dumas *et al.* 2012) ont été redessinées afin de pouvoir utiliser la technique des queues M13 (Schuelke 2000). Le principe de la méthode est illustré par la figure ci-dessous (figure S1). La réaction de PCR utilise alors 3 amorces différentes (tableau S1a et b). Le volume final de la réaction est de 10 µL contenant 2 µL de tampon 5x (Promega, Madison, WI), 0.6 µL de MgCl₂ à 25mM, 0.2 µL de dNTP à 10mM (Promega, Madison, WI), 0.2µL d'amorce spécifique sens à 2µM, 0.75 µL d'amorce spécifique anti-sens à 2 µM, 0.75 µL d'amorce marquée pour l'un des quatre fluorochromes à 2µM, 0.1 µL de GoTaq DNA polymérase (Promega, Madison, WI) et 1µL d'Adn génomique concentré à 40 ng/µL. Les produits de PCR sont ensuite multiplexés par quatre (un de chaque fluorochrome) pour le passage sur électrophorèse capillaire.

Lecture des profils par électrophorèse capillaire

Deux microlitres de produit de digestion sont ajoutés à 15 µL de Formamide et 0.2 µL de marqueur de taille Liz500. Les mix sont analysés sur un séquenceur à capillaires ABI par la plateforme SeqGen (CeMEB Labex, Montpellier, France). La lecture de la taille et l'identification des fragments est effectuée avec GENEMAPPER v4.0 (Applied Biosystems), les résultats générés par le logiciel sont revérifiés à l'œil.

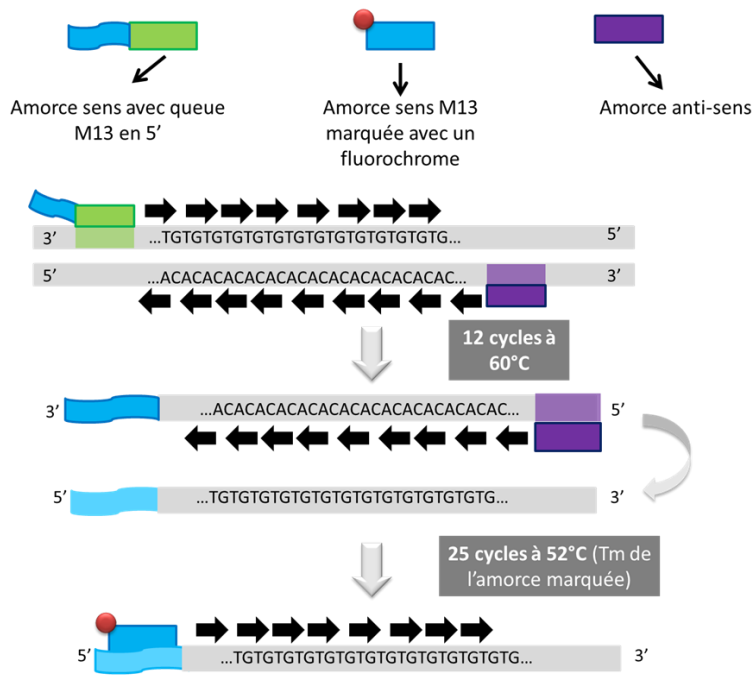


Figure S1 : Schéma présentant le principe de la PCR avec amorces queue M13

Tableau S1a : Séquences des amorces utilisées pour amplifier les différents microsatellites grâce à la méthode des queues M13.

Nom du microsatellite	amorces spécifique Sens	amorces spécifique Anti-sens	N°accesion GeneBank
<i>Sfrugi2</i>	TGTA AACGACGGCCAGT GTACGGCTAGGGGT GAAATGTAA	GCGTTGAATATAATGCG GCT	JN796474
<i>Sfrugi5</i>	TGTA AACGACGGCCAGT GTAGCTAGGAGTTT TCGGGAAAGTA	CGTCCATATTACCCCGAG TG	JN790256
<i>Sfrugi6</i>	TGTA AACGACGGCCAGT GTATGGAATATGTG CAGCGAAAA	TCGTTATGTCATACACCA TTCAGA	JN790257
<i>Sfrugi21</i>	TGTA AACGACGGCCAGT GTACCTTGGGTAAG GTGAGAGGG	CAAATCCAATGATGACC ACG	JN790258
<i>Sfrugi24</i>	TGTA AACGACGGCCAGT GTATTACACATTTCGT TTGTCACGC	GAATGGAACAGACAGTC GTTGT	JN790259
<i>Sfrugi25</i>	TGTA AACGACGGCCAGT GTATGTCATTGATGC TGCAAACA	CGGCAACTACGTACTGT ATGGTTA	JN790260
<i>Sfrugi29</i>	TGTA AACGACGGCCAGT GTAGCAAACGGCGA TCAAAGTAT	CGAAATGTGTTTAAGTCA CTACCT	JN790261
<i>Sfrugi30</i>	TGTA AACGACGGCCAGT GTAGGTGTAGGTTG TCTTGCGGT	ACTCATTGACCCAGCGTA CC	JN790262
<i>Sfrugi33</i>	TGTA AACGACGGCCAGT GTAATTCCAAAGAG GCACCATAG	GCAGCATGTTCAACCCAC TA	JN790263
<i>Sfrugi34</i>	TGTA AACGACGGCCAGT GTATCATCATTTGCA TGACCAACT	AGGGAGGCAAGCAAAG AAAT	JN790264
<i>Sfrugi37</i>	TGTA AACGACGGCCAGT GTAGTAAGTAAATC CTGGCAACACGC	TGTCTGGAGCAGTTATG CATT	JN790265
<i>Sfrugi38</i>	TGTA AACGACGGCCAGT GTAAAATCATTCCC AGAAACAAACA	GACAGACCATTGACGG TGA	JN790266
<i>Sfrugi39</i>	TGTA AACGACGGCCAGT GTAGAGCTCGTTCTC CCTCACTG	GCCATTCTGACTGAAGC AT	JN790267
<i>Sfrugi43</i>	TGTA AACGACGGCCAGT GTAACCTTGAAAAA GGCGTCGAA	GCCCTCAACCAGACCAAT AA	JN790268
<i>Sfrugi50</i>	TGTA AACGACGGCCAGT GTAACCACCATGCG GTAGCTATT	CACCTACCTCGTATCAGC GT	JN790269
<i>Sfrugi 58</i>	TGTA AACGACGGCCAGT GTAATCGATGTGAC ACCACTCA	TTAGTTGTGTGTGTAGA GTAAGC	JN790270
<i>Sfrugi60</i>	TGTA AACGACGGCCAGT GTATGTTGCGCACTT CTACTGCT	AATTGTTGAGGTTTGCCT CA	JN790271
<i>Sfrugi76</i>	TGTA AACGACGGCCAGT GTATCGTTACCA AGCCGTGC	ACCCTTATTGGCAATC GAAA	

Tableau S1b : Séquences des quatre amorces sens marquées par un fluorochrome différents.

VIC-TGT AAA ACG ACG GCC AGT
NED-TGT AAA ACG ACG GCC AGT
FAM-TGT AAA ACG ACG GCC AGT
PET-TGT AAA ACG ACG GCC AGT

Références

- Abbott R., *et al.* (2013). **Hybridization and speciation**. *Journal of Evolutionary Biology* **26**(2): 229-246.
- Abbott R. J., *et al.* (2010). **Homoploid hybrid speciation in action**. *Taxon* **59**(5): 1375-1386.
- All J. N., *et al.* (1996). **Controlling fall armyworm infestations in whorl stage corn with genetically modified *Bacillus thuringiensis* formulations**. *Florida Entomologist* **79**: 311-317.
- Alstrom P., *et al.* (2008). **Species delimitation based on multiple criteria: the Spotted Bush Warbler *Bradypterus thoracicus* complex (Aves:Megaluridae)**. *Zoological Journal of the Linnean Society* **154**: 291-307.
- Arbuthnott D., *et al.* (2010). **Divergent patterns of diversification in courtship and genitalic characters of *Timema walking-sticks***. *Journal of Evolutionary Biology* **23**(7): 1399-1411.
- Astrin J. J., *et al.* (2012). **Exploring diversity in cryptorhynchine weevils (Coleoptera) using distance-, character- and tree-based species delineation**. *Molecular Phylogenetics and Evolution* **63**(1): 1-14.
- Avise J. and Ball R. (1990). **Principles of genealogical concordance in species concepts and biological taxonomy**. *Surveys in Evolutionary Biology* **7**: 45-67.
- Barluenga M. K. N. S. W. M. A. (2006). **Sympatric speciation in Nicaraguan crater lake cichlid fish**. *Nature* **439**(7077): 719-723.
- Barraclough T. G., *et al.* (2003). **Diversification in sexual and asexual organisms**. *Evolution* **57**(9): 2166-2172.
- Barros E. M., *et al.* (2010). **Development of *Spodoptera frugiperda* on different hosts and damage to reproductive structures in cotton**. *Entomologia Experimentalis et Applicata* **137**(3): 237-245.
- Berlocher S. H. and Feder J. L. (2002). **Sympatric speciation in phytophagous insects: moving beyond controversy**. *Annual Review in Entomology* **47**: 773-815.
- Bickford D., *et al.* (2007). **Cryptic species as a window on diversity and conservation**. *Trends in Ecology & Evolution* **22**(3): 148-155.
- Birky C. W., Jr., *et al.* (2005). **Speciation and Selection without Sex**. *Hydrobiologia* **546**(1): 29-45.
- Blair C. P., *et al.* (2005). **Cryptic speciation and host-race formation in a purportedly generalist tumbling flower beetle**. *Evolution* **59**(2): 304-316.
- Bolnick D. I. and Fitzpatrick B. M. (2007). **Sympatric Speciation: Models and Empirical Evidence**. *Annual Review of Ecology, Evolution, and Systematics* **38**(1): 459-487.
- Brucker R. M. and Bordenstein S. R. (2011). **The roles of host evolutionary relationships (Genus:*Nasonia*) and development in structuring microbial communities**. *Evolution* **66**(2): 349-362.
- Brucker R. M. and Bordenstein S. R. (2012). **Speciation by symbiosis**. *Trends in Ecology & Evolution* **27**(8): 443-451.
- Brucker R. M. and Bordenstein S. R. (2013). **The Hologenomic Basis of Speciation: Gut Bacteria Cause Hybrid Lethality in the Genus *Nasonia***. *Science* **341**(6146): 667-669.
- Bryson R. W., *et al.* (2013). **Biogeography of scorpions in the *Pseudouroctonus minimus* complex (Vaejovidae) from south-western North America: implications of ecological specialization for pre-Quaternary diversification**. *Journal of Biogeography*.
- Budar F., *et al.* (2003). **The Nucleo-Mitochondrial Conflict in Cytoplasmic Male Sterilities Revisited**. *Genetica* **117**(1): 3-16.
- Busato G. R., *et al.* (2004). **Analysis of the molecular structure and diversity of *Spodoptera frugiperda* (JE smith) (Lepidoptera : Noctuidae) populations associated to the corn and rice crops in Rio Grande do Sul State, Brazil**. *Neotropical Entomology* **33**(6): 709-716.
- Bush G. (1969). **Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera, Tephritidae)**. *Evolution* **23**: 237 - 251.
- Butlin R. K. (2005). **Recombination and speciation**. *Molecular Ecology* **14**(9): 2621-2635.
- Butlin R. K. (2008). **Population genomics and speciation**. *Genetica* **138**: 409-418.

- Butlin R. K., *et al.* (2008). **Sympatric, parapatric or allopatric: the most important way to classify speciation?** *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**(1506): 2997-3007.
- Capinera J. L. (1999). **Fall Armyworm, *Spodoptera frugiperda* (J.E. Smith) (Insecta: Lepidoptera: Noctuidae).** Featured Creatures from the Entomology and Nematology Department, Florida Cooperative Extension Service **EENY-098**.
- Capinera J. L. (2008). Maize (Corn) Pests and Their Management. *Encyclopedia of Entomology*. J. L. Capinera, Springer Netherlands: 2266-2272.
- Cardoso A. and Vogler A. P. (2005). **DNA taxonomy, phylogeny and Pleistocene diversification of the *Cicindela hybrida* species group (Coleoptera: Cicindelidae).** *Molecular Ecology* **14**(11): 3531-3546.
- Carson H. L. and Kaneshiro K. Y. (1976). ***Drosophila* of Hawaii: Systematics and Ecological Genetics.** *Annual Review of Ecology and Systematics* **7**: 311-345.
- Chapman J. W., *et al.* (1999). **Fitness consequences of cannibalism in the fall armyworm, *Spodoptera frugiperda*.** *Behavioral Ecology* **10**(3): 298-303.
- Chen H., *et al.* (2010). **Statistical Parsimony Networks and Species Assemblages in *Cephalotrichid Nemerteans* (Nemertea).** *Plos One* **5**(9): e12885.
- Colborn J., *et al.* (2001). **The evolutionary enigma of bonefishes (*Albula* spp.): cryptic species and ancient separations in a globally distributed shorefish.** *Evolution* **55**(4): 807-820.
- Cordaux R., *et al.* (2011). **The impact of endosymbionts on the evolution of host sex-determination mechanisms.** *Trends in Genetics* **27**(8): 332-341.
- Coyne J. and Orr H. A., Eds. (2004). **Speciation.** Massachusetts, Sunderland.
- Coyne J. A. (1994). **Ernst Mayr and the Origin of Species.** *Evolution* **48**(1): 19-30.
- Coyne J. A. and Orr H. (1998). **The evolutionary genetics of speciation.** *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **353**(1366): 287-305.
- Craig T. P., *et al.* (1993). **Behavioral Evidence for Host-Race Formation in *Eurosta solidaginis*.** *Evolution* **47**(6): 1696-1710.
- Cruz I., *et al.* (1999). **Damage of *Spodoptera frugiperda* (Smith) in different maize genotypes cultivated in soil under three levels of aluminium saturation.** *International Journal of Pest Management* **45**(4).
- d'Alençon E., *et al.* (2004). **A genomic BAC library and a new BAC-GFP vector to study the holocentric pest *Spodoptera frugiperda*.** *Insect Biochemistry and Molecular Biology* **34**: 331-341.
- Darwin C., Ed. (1859). **The Origins of Species by Means of Natural Selection, or the Preservation of Favored Races in the Struggle for Life.** New York, Modern Library.
- Dayrat B. (2005). **Toward integrative taxonomy.** *Biol J Linn Soc* **85**: 407-415.
- de Queiroz K. (2007). **Species concepts and species delimitation.** *Syst Biol* **56**: 879-886.
- Delsuc F. and Douzery E. J. P. (2004). **Avenir et pertinence des méthodes d'analyse en phylogénie moléculaire.** *Biosystema* **22**: 59-86.
- DeSalle R., Ed. (1995). **Molecular approaches to biogeographic analysis of Hawaiian Drosophilidae** *Hawaiian biogeography: Evolution of a hotspot archipelago.* Washington D.C., Smithsonian Institution.
- Dobzhansky T. (1937). **Genetics and the Origin of Species** C. U. Press. New York.
- Dobzhansky T., Ed. (1970). **Genetics of the Evolutionary process.** New York, Columbia University.
- Doebeli M. and Dieckmann U. (2003). **Speciation along environmental gradients.** *Nature* **421**(6920): 259-264.
- Dopman E. B., *et al.* (2004). **Genetic Mapping of Sexual Isolation Between E and Z Pheromone Strains of the European Corn Borer (*Ostrinia nubilalis*).** *Genetics* **167**(1): 301-309.
- Drès M. and Mallet J. (2002). **Host races in plant-feeding insects and their importance in sympatric speciation.** *The Royal Society* **357**: 471-492.
- Ellison C. K. and Burton R. S. (2008). **Interpopulation hybrid breakdown maps to the mitochondrial genome.** *Evolution* **62**(3): 631-638.

- Ellison C. K., *et al.* (2008). **Hybrid breakdown and mitochondrial dysfunction in hybrids of *Nasonia parasitoid* wasps**. *Journal of Evolutionary Biology* **21**(6): 1844-1851.
- Emelianov I., *et al.* (2004). **Genomic evidence for divergence with gene flow in host races of the larch budmoth**. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271**(1534): 97-105.
- Ence D. D. and Carstens B. C. (2011). **SpedeSTEM: a rapid and accurate method for species delimitation**. *Molecular Ecology Resources* **11**(3): 473-480.
- Engelstädter J. and Hurst G. D. D. (2009). **The Ecology and Evolution of Microbes that Manipulate Host Reproduction**. *Annual Review of Ecology, Evolution, and Systematics* **40**(1): 127-149.
- Fang S., *et al.* (2002). **A Mutation in the Promoter of Desaturase 2 Is Correlated With Sexual Isolation Between *Drosophila* Behavioral Races**. *Genetics* **162**(2): 781-784.
- Feder J., *et al.* (1998). **Sympatric Host-Race Formation and Speciation in *Rhagoletis* (Diptera: Tephritidae): A Tale of Two Species for Charles D. Genetic Structure and Local Adaptation in Natural Insect Populations**. S. Mopper and S. Strauss, Springer US: 408-441.
- Feder J., *et al.* (2003). **Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis***. *Proceedings of the National Academy of Sciences USA* **100**(18): 10314-10319.
- Feder J., *et al.* (2005). **Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in *Rhagoletis***. *Proceedings of the National Academy of Sciences USA* **102**: 6573-6580.
- Feder J. L., *et al.* (1994). **Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly**. *Proceedings of the National Academy of Sciences* **91**(17): 7990-7994.
- Felsenstein J. (1973). **Maximum Likelihood and Minimum-Steps Methods for Estimating Evolutionary Trees from Data on Discrete Characters**. *Systematic Biology* **22**(3): 240-249.
- Felsenstein J. (1981). **Evolutionary trees from DNA sequences: A maximum likelihood approach**. *Journal of Molecular Evolution* **17**(6): 368-376.
- Fernández-Flores S., *et al.* (2013). **DNA barcoding species inventory of *Microgastrinae* wasps (Hymenoptera, Braconidae) from a Mexican tropical dry forest**. *Molecular Ecology Resources*.
- Fitzpatrick B. M., *et al.* (2008). **What, if anything, is sympatric speciation?** *Journal of Evolutionary Biology* **21**(6): 1452-1459.
- Fitzpatrick B. M., *et al.* (2009). **Pattern, process and geographic modes of speciation**. *Journal of Evolutionary Biology* **22**(11): 2342-2347.
- Fontaneto D., *et al.* (2011). **Cryptic diversity in the genus *Adineta* Hudson & Gosse, 1886 (Rotifera: Bdelloidea: Adinetidae): a DNA taxonomy approach**. *Hydrobiologia* **662**(1): 27-33.
- Fujita M. K., *et al.* (2012). **Coalescent-based species delimitation in an integrative taxonomy**. *Trends in Ecology & Evolution* **27**(9): 480-488.
- Funk D. J. (1998). **Isolating a Role for Natural Selection in Speciation: Host Adaptation and Sexual Isolation in *Neochlamisus bebbianae* Leaf Beetles**. *Evolution* **52**(6): 1744-1759.
- Funk D. J. (2012). **Of "Host Forms" ; and Host Races: Terminological Issues in Ecological Speciation**. *International Journal of Ecology*.
- Funk D. J., *et al.* (2002). **Herbivorous insects: model systems for the comparative study of speciation ecology**. *Genetica* **116**: 251-267.
- Gavrilets S. (2003). **Perspective: models of speciation: what have we learned in 40 years?** *Evolution* **57**(10): 2197-2215.
- Gavrilets S., *et al.* (2000). **Patterns of parapatric speciation**. *Evolution* **54**(4): 1126-1134.
- Gebiola M., *et al.* (2012). **Integration of molecular, ecological, morphological and endosymbiont data for species delimitation within the *Pnigalio soemius* complex (Hymenoptera: Eulophidae)**. *Molecular Ecology* **21**(5): 1190-1208.
- Gerbal M., *et al.* (2000). **Adaptation of an insect cell line of *Spodoptera frugiperda* to grow at 37°C: Characterization of an endodiploid clone**. *In Vitro Cellular & Developmental Biology - Animal* **36**(2): 117-124.
- Giraldo N., *et al.* (2008). **Two sisters in the same dress: *Heliconius* cryptic species**. *BMC Evolutionary Biology* **8**(1): 324.

- Giraud T. (2006). **Speciation: Selection against migrant pathogens: the immigrant inviability barrier in pathogens.** *Heredity* **97**(5): 316-318.
- Grant V., Ed. (1981). **Plant speciation.** New York, Columbia University
- Groot A. T., *et al.* (2010). **The roles and interactions of reproductive isolation mechanisms in fall armyworm (Lepidoptera: Noctuidae) host strains.** *Ecological Entomology* **35**(1): 105-118.
- Groot A. T., *et al.* (2008). **Host strain specific sex pheromone variation in *Spodoptera frugiperda*.** *Frontiers in Zoology* **5**(1): 20.
- Haldane J. (1922). **Sex ratio and unisexual sterility in hybrid animals.** *Journal of Genetics* **12**(2): 101-109.
- Hamback P., *et al.* (2013). **Bayesian species delimitation reveals generalist and specialist parasitic wasps on *Galerucella* beetles (Chrysomelidae): sorting by herbivore or plant host.** *BMC Evolutionary Biology* **13**(1): 92.
- Harlan J. R. and De Wet J. M. J. (1963). **The Compilospecies concept.** *Evolution* **17**: 497-501.
- Hebert P., *et al.* (2003). **Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species.** *Proceedings of the Royal Society B: Biological Sciences* **270** Suppl 1: 96-99.
- Hebert P., *et al.* (2004). **Identification of Birds through DNA Barcodes.** *PLoS Biol* **2**(10): e312.
- Hoffmann A. A. and Rieseberg L. (2008). **Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation?** *The Annual Review of Ecology, Evolution, and Systematics* **39**: 21-42.
- Hosken D. J. and Stockley P. (2004). **Sexual selection and genital evolution.** *Trends in Ecology & Evolution* **19**(2): 87-93.
- Hurst G. D. D. and Jiggins F. M. (2005). **Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts.** *Proceedings of the Royal Society B: Biological Sciences* **272**(1572): 1525-1534.
- Itami J., *et al.* (1998). **Factors Affecting Gene Flow between the Host Races of *Eurosta solidaginis*. Genetic Structure and Local Adaptation in Natural Insect Populations.** S. Mopper and S. Strauss, Springer US: 375-407.
- Jiggins C. D., *et al.* (2001). **Reproductive isolation caused by colour pattern mimicry.** *Nature* **411**(6835): 302-305.
- Johnson N. A. (2010). **Hybrid incompatibility genes: remnants of a genomic battlefield?** *Trends in Genetics* **26**(7): 317-325.
- Joron M., *et al.* (2011). **Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry.** *Nature* **477**: 203-206.
- Juárez M. L., *et al.* (2012). **Host Association of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) Corn and Rice Strains in Argentina, Brazil, and Paraguay.** *Journal of Economic Entomology* **105**(2): 573-582.
- Kergoat G. J., *et al.* (2011). **Phylogenetics, species boundaries and timing of resource tracking in a highly specialized group of seed beetles (Coleoptera: Chrysomelidae: Bruchinae).** *Molecular Phylogenetics and Evolution* **59**(3): 746-760.
- Kingman J. F. C. (1982). **The coalescent.** *Stochastic Processes and their Applications* **13**(3): 235-248.
- Kirkpatrick M. and Ravigné V. (2002). **Speciation by Natural and Sexual Selection: Models and Experiments.** *The American Naturalist* **159**(S3): S22-S35.
- Kitching I. and Rawlins J., Eds. (1998). **Handbook of Zoology Vol. IV Arthropoda: Insecta. Lepidoptera, Moths AND Butterflies. Vol. 1: Evolution, Systematics AND Biogeography.** Berlin, Walter de Gruyter & Co.
- Knowlton N., *et al.* (1993). **Divergence in Proteins, Mitochondrial DNA, and Reproductive Compatibility Across the Isthmus of Panama.** *Science, New Series* **260**(5114): 1629-1632.
- Knowlton N. and Weigt L. A. (1998). **New dates and new rates for divergence across the Isthmus of Panama.** *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265**(1412): 2257-2263.

- Kobayashi Y. and Telschow A. (2011). **The concept of effective recombination rate and its application in speciation theory.** *Evolution* **65**(3): 617-628.
- Kronforst M. R., *et al.* (2006). **Parallel Genetic Architecture of Parallel Adaptive Radiations in Mimetic Heliconius Butterflies.** *Genetics* **174**(1): 535-539.
- Laurie C. C. (1997). **The Weaker Sex Is Heterogametic: 75 Years of Haldane's Rule.** *Genetics* **147**(3): 937-951.
- Lessios H. A., *et al.* (2001). **Population structure and speciation in tropical seas: global phylogeography of the sea Urchin *diadema*.** *Evolution* **55**(5): 955-975.
- Lessios H. A., *et al.* (1998). **Massive gene flow across the world's most potent marine biogeographic barrier.** *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265**(1396): 583-588.
- Levy H. C., *et al.* (2002). **Strain identification of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) insects and cell line: PCR-RFLP of cytochrome oxidase C subunit I gene.** *The Florida Entomologist* **85**(1): 186-190.
- Li S. (1996). **Phylogenetic reconstruction using Markov chain Monte carlo.** Columbus, Ohio State University. **PhD Dissertation.**
- Löfstedt C., *et al.* (1991). **Sex pheromones and their potential role in the evolution of reproductive isolation in small ermine moths (*Yponomeutidae*).** *Chemoecology* **2**: 20-28.
- Lu Y. J. and Adang M. J. (1996). **Distinguishing fall armyworm (Lepidoptera: Noctuidae) strain using a diagnostic mitochondrial DNA marker.** *Florida Entomologist* **79**(1): 48-55.
- Lu Y. J., *et al.* (1994). **Molecular characterization of a strain-specific repeated DNA sequence in the fall armyworm *Spodoptera frugiperda* (Lepidoptera: Noctuidae).** *Insect Molecular Biology* **3**(2): 123-130.
- Lumley L. and Sperling F. (2010). **Integrating morphology and mitochondrial DNA for species delimitation within the spruce budworm (*Choristoneura fumiferana*) cryptic species complex (Lepidoptera: Tortricidae).** *Systematic Entomology* **35** (3): 416-428.
- Lumley L. M. and Cusson M. (2013). **Linking adaptation, delimitation of evolutionarily significant units (ESUs), and gene function: a case study using hemlock looper ecotypes.** *Systematic Entomology* **38**(2): 428-439.
- Mable B. K., *et al.* (2011). **Genome duplication in amphibians and fish: an extended synthesis.** *Journal of Zoology* **284**(3): 151-182.
- Machado V., *et al.* (2008). **Molecular Characterization of Host Strains of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in Southern Brazil.** *Annals of the Entomological Society of America* **101**(3): 619-626.
- Mäkinen H. S., *et al.* (2008). **Hitchhiking Mapping Reveals a Candidate Genomic Region for Natural Selection in Three-Spined Stickleback Chromosome VIII.** *Genetics* **178**(1): 453-465.
- Malausa T., *et al.* (2005). **Assortative Mating in Sympatric Host Races of the European Corn Borer.** *Science* **308**: 258-260.
- Malausa T., *et al.* (2007). **Genetic structure and gene flow in French populations of two *Ostrinia* taxa: host races or sibling species?** *Molecular Ecology* **16**(20): 4210-4222.
- Mallet J., *et al.* (2007). **Natural hybridization in heliconiine butterflies: the species boundary as a continuum.** *BMC Evolutionary Biology* **7**(1): 28.
- Marenco R. J., *et al.* (1992). **Sweet corn response to fall armyworm (Lepidoptera: Noctuidae) damage during vegetative growth.** *Journal of Economic Entomology* **85**: 1285-1292.
- Matsubayashi K. W. and Katakura H. (2009). **Contribution of multiple isolating barriers to reproductive isolation between a pair of phytophagous ladybird beetles.** *Evolution* **63**(10): 2563-2580.
- Matsubayashi K. W., *et al.* (2010). **Ecological speciation in phytophagous insects.** *Entomologia Experimentalis et Applicata* **134**(1): 1-27.
- Mau B. (1996). **Bayesian phylogenetic inference via markov chain Monte carlo methods.** PhD Dissertation. University of Wisconsin, Madison.

- Mayden R., Ed. (1997). **A hierarchy of species concepts: The denouement in the saga of the species problem**. *Species: The units of biodiversity*. London, Chapman & Hall
- Maynard Smith J. (1966). **Sympatric Speciation**. *The American Naturalist* **100**(916): 637-650.
- Mayr E., Ed. (1942). **Systematics and the origin of species, from the viewpoint of a zoologist**. Columbia University Press. New York.
- McBride C. S., *et al.* (2009). **Allopatric origin of cryptic butterfly species that were discovered feeding on distinct host plants in sympatry**. *Molecular Ecology* **18**: 3639-3651.
- McDermott S. R. and Noor M. A. F. (2010). **The role of meiotic drive in hybrid male sterility**. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**(1544): 1265-1272.
- McMichael M. and Pashley Prowell D. P. (1999). **Differences in amplified fragment-length polymorphisms in fall armyworm (Lepidoptera: Noctuidae) host strain**. *Annals of the Entomological Society of America* **92**(2): 175-181.
- McMillan W. O., *et al.* (1997). **What initiates speciation in passion-vine butterflies?** *Proceedings of the National Academy of Sciences* **94**(16): 8628-8633.
- Meagher R., *et al.* (2011). **Oviposition Choice of Two Fall Armyworm (Lepidoptera: Noctuidae) Host Strains**. *Journal of Insect Behavior*: 1-11.
- Meagher R. L. and Gallo-Meagher M. (2003). **Identifying host strains of fall armyworm (Lepidoptera : Noctuidae) in Florida using mitochondrial markers**. *Florida Entomologist* **86**(4): 450-455.
- Meagher R. L., *et al.* (2004). **Larval development of Fall armyworm (Lepidoptera: Noctuidae) on different cover crop plants**. *Florida Entomologist* **87**(4): 454-460.
- Melters D., *et al.* (2012). **Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis**. *Chromosome Research* **20**(5): 579-593.
- Michel A. P., *et al.* (2010). **Widespread genomic divergence during sympatric speciation**. *Proceedings of the National Academy of Sciences* **107**(21): 9724-9729.
- Mishler B. D. (1985). **The Morphological, Developmental, and Phylogenetic Basis of Species Concepts in Bryophytes**. *The Bryologist* **88**(3): 207-214.
- Monaghan M., *et al.* (2009). **Accelerated species inventory on Madagascar using coalescent-based models of species delineation**. *Syst Biol* **58**: 298 - 311.
- Monteiro A. and Pierce N. E. (2001). **Phylogeny of *Bicyclus* (Lepidoptera: Nymphalidae) inferred from COI, COII, and Ef-1 α gene sequences**. *Molecular Phylogenetic and Evolution* **18**: 264–281.
- Muller H. J. (1942). **Isolating mechanisms, evolution and temperature**. *Biol. Symp.* **6**(71).
- Nadeau N. J., *et al.* (2012). **Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing**. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**(1587): 343-353.
- Nagoshi R. N., *et al.* (2011). **Use of DNA Barcodes to Identify Invasive Armyworm Spodoptera Species in Florida**. *Journal of Insect Science* **11**(154): 1-11.
- Nagoshi R. N., *et al.* (2009). **Texas Is the Overwintering Source of Fall Armyworm in Central Pennsylvania: Implications for Migration Into the Northeastern United States**. *Environmental Entomology* **38**(6): 1546-1554
- Nagoshi R. N. and Meagher R. L. (2003). **Fall armyworm FR sequences map to sex chromosomes and their distribution in the wild indicate limitations in interstrain mating**. *Insect Molecular Biology* **12**(5): 453–458.
- Nagoshi R. N. and Meagher R. L. (2003). **FR Tandem-Repeat Sequence in Fall Armyworm (Lepidoptera: Noctuidae) Host Strains**. *Annals of the Entomological Society of America* **96**(3): 329-335.
- Nagoshi R. N., *et al.* (2006). **New restriction fragment length polymorphisms in the cytochrome oxidase I gene facilitate host strain identification of Fall armyworm (Lepidoptera: Noctuidae) populations in the southeastern United States**. *Journal of Economic Entomology* **99**(3): 671-677.

- Nagoshi R. N., *et al.* (2012b). **Inferring the annual migration patterns of fall armyworm (*Lepidoptera: Noctuidae*) in the United States from mitochondrial haplotypes.** *Ecology and Evolution* **2**(7): 1458-1467.
- Nagoshi R. N., *et al.* (2006). **Effects of Fall armyworm (*Lepidoptera: Noctuidae*) interstrain mating in wild populations.** *Environmental Entomology* **35**(2): 561-568
- Nagoshi R. N., *et al.* (2007). **Identification and comparison of Fall armyworm (*Lepidoptera: Noctuidae*) host strains in Brazil, Texas, and Florida.** *Annals of the Entomological Society of America* **100**(3): 394-402
- Nei M. and Nozawa M. (2011). **Roles of Mutation and Selection in Speciation: From Hugo de Vries to the Modern Genomic Era.** *Genome Biology and Evolution* **3**: 812-829.
- Nei M., *et al.* (1983). **Accuracy of estimated phylogenetic trees from molecular data.** *Journal of Molecular Evolution* **19**(2): 153-170.
- Noor M. A. F. (2002). **Is the biological species concept showing its age?** *Trends in Ecology & Evolution* **17**(4): 153-154.
- Noor M. A. F. and Feder J. L. (2006). **Speciation genetics: evolving approaches.** *Nat Rev Genet* **7**(11): 851-861.
- Noor M. A. F., *et al.* (2007). **Divergence Between the *Drosophila pseudoobscura* and *D. persimilis* Genome Sequences in Relation to Chromosomal Inversions.** *Genetics* **177**(3): 1417-1428.
- Noor M. A. F., *et al.* (2001). **Chromosomal inversions and the reproductive isolation of species.** *Proceedings of the National Academy of Sciences* **98**(21): 12084-12088.
- Nosil P. (2004). **Reproductive isolation caused by visual predation on migrants between divergent environments.** *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271**(1547): 1521-1528.
- Nosil P. (2008). **Ernst Mayr and the integration of geographic and ecological factors in speciation.** *Biological Journal of the Linnean Society* **95**(1): 26-46.
- Nosil P., Ed. (2012b). **Ecological Speciation.** Oxford, Oxford University.
- Nosil P., *et al.* (2002). **Host-plant adaptation drives the parallel evolution of reproduction isolation.** *Nature* **417**(6887): 440.
- Nosil P. and Feder J. L. (2012a). **Genomic divergence during speciation: causes and consequences.** *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**(1587): 332-342.
- Nosil P., *et al.* (2009). **Ecological explanations for (incomplete) speciation.** *Trends Ecol Evol* **24**: 145 - 156.
- Nosil P., *et al.* (2005). **Reproductive isolation caused by natural selection against immigrants from divergent habitats.** *Evolution* **59**(4): 705-719.
- O'Meara B. C., *et al.* (2006). **Testing for different rates of continuous trait evolution using likelihood.** *Evolution* **60**(5): 922-933.
- Orr H. A. (1995). **The population genetics of speciation: the evolution of hybrid incompatibilities.** *Genetics* **139**: 1805-1813.
- Orr H. A., *et al.* (2004). **Speciation genes.** *Current Opinion in Genetics & Development* **14**(6): 675-679.
- Ortiz-Barrientos D., *et al.* (2002). **Recombination and the divergence of hybridizing species.** *Genetics of Mate Choice: From Sexual Selection to Sexual Isolation.* W. J. Etges and M. A. F. Noor, Springer Netherlands. **9**: 167-178.
- Pair S. D., *et al.* (1986). **Fall Armyworm Distribution and Population Dynamics in the Southeastern States.** *The Florida Entomologist* **69**(3): 468-487.
- Pashley D. P. (1986). **Host-associated genetic differentiation in fall armyworm (*Lepidoptera: Noctuidae*): a sibling species complex?** *Annals of the Entomological Society of America* **79**: 898-904.
- Pashley D. P., Ed. (1989). **Host-associated differentiation in armyworms (*Lepidoptera: Noctuidae*): An allozymic and mitochondrial DNA perspective.** *Electrophoretic Studies on Agricultural Pests.* Oxford.
- Pashley D. P., *et al.* (1985). **Genetic population structure of migratory moths: the fall armyworm (*Lepidoptera: Noctuidae*).** *Annals of the Entomological Society of America* **78**: 756-762.

- Pashley D. P. and Ke L. D. (1992). **Sequence evolution in mitochondrial ribosomal and ND-1 genes in lepidoptera: implications for phylogenetic analyses.** *Molecular Biology and Evolution* **9**(6): 1061-1075.
- Pashley Prowell D. P., *et al.* (1995). **Host effect on developmental and reproductive traits in fall armyworm strain (Lepidoptera: Noctuidae).** *Annals of the Entomological Society of America* **88**(6): 748-755.
- Pashley Prowell D. P. and Martin J. A. (1987). **Reproductive incompatibility between host strains of the Fall armyworm (Lepidoptera: Noctuidae).** *Annals of the Entomological Society of America*: 731-733.
- Peccoud J., *et al.* (2009). **A continuum of genetic divergence from sympatric host races to species in the pea aphid complex.** *Proceedings of the National Academy of Sciences* **106**(18): 7495-7500.
- Peccoud J. and Simon J.-C. (2010). **The pea aphid complex as a model of ecological speciation.** *Ecological Entomology* **35**: 119-130.
- Petit R. J. and Excoffier L. (2009). **Gene flow and species delimitation.** *Trends in Ecology & Evolution* **24**(7): 386-393.
- Pogue M. G. (2002). **World revision of the genus *Spodoptera* Guenée (Lepidoptera: Noctuidae).** *Memoirs of the American Entomological Society* **43**: 1-202.
- Poitout S. and Bues R. (1974). **Linolenic acid requirements of lepidoptera Noctuidae Quadrifinae Plusiinae: *Chrysodeixis chalcites* Esp, *Autographa gamma* L.' *Macdunnoughia confusa* Stph., *Trichoplusia ni* Hbn. reared on artificial diets.** *Annales de la nutrition et de l'alimentation* **28**: 173-187.
- Pons J., *et al.* (2006). **Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects.** *Systematic Biology* **55**(4): 595-609.
- Presgraves D. (2010). **The molecular basis of species formation.** *Nature Rev Gen.*
- Presgraves D. C. (2002). **Patterns of postzygotic isolation in Lepidoptera.** *Evolution* **56**(6): 1168-1183.
- Presgraves D. C. (2007). **Does genetic conflict drive rapid molecular evolution of nuclear transport genes in *Drosophila*?** *BioEssays* **29**(4): 386-391.
- Prowell D. P., *et al.* (2004). **Multilocus genetic analysis of host use, introgression, and speciation in host strains of fall armyworm (Lepidoptera: Noctuidae).** *Annals of the Entomological Society of America* **97**(5): 1034-1044
- Quisenberry S. S. (1991). **Fall armyworm (Lepidoptera: Noctuidae) host strain reproductive compatibility.** *Florida Entomologist* **74**(2): 194-199.
- Rannala B. and Yang Z. (1996). **Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference.** *Journal of Molecular Evolution* **43**(3): 304-311.
- Ravigné V., *et al.* (2004). **Implications of habitat choice for protected polymorphisms.** *Evolutionary Ecology Research* **6**: 125-145.
- Reed R. D., *et al.* (2011). ***optix* Drives the Repeated Convergent Evolution of Butterfly Wing Pattern Mimicry.** *Science* **333**(6046): 1137-1141.
- Rice W. R. and Hostert E. E. (1993). **Laboratory Experiments on Speciation: What Have We Learned in 40 Years?** *Evolution* **47**(6): 1637-1653.
- Rieseberg L. H. (2001). **Chromosomal rearrangements and speciation.** *Trends in Ecology & Evolution* **16**(7): 351-358.
- Roelofs W. L. and Rooney A. P. (2003). **Molecular genetics and evolution of pheromone biosynthesis in Lepidoptera.** *Proceedings of the National Academy of Sciences* **100**(16): 9179-9184.
- Rosen D. E. (1979). **Fishes from the uplands and intermontane basins of Guatemala : Revisionary studies and comparative geography.** *Bulletin of the American Museum of Natural History* **162**: 267-376.
- Rougerie R., *et al.* (2012). **Morphology and Molecules Reveal Unexpected Cryptic Diversity in the Enigmatic Genus *Sinobirma* Bryk, 1944 (Lepidoptera: Saturniidae).** *Plos One* **7**(9): e43920.
- Rundle H. D. and Nosil P. (2005). **Ecological speciation.** *Ecology Letters* **8**: 336-352.

- Rundle H. D. and Whitlock M. C. (2001). **A genetic interpretation of ecologically dependent isolation.** *Evolution* **55**(1): 198-201.
- Saitou N. and Nei M. (1987). **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* **4**(4): 406-425.
- Salazar C., *et al.* (2010). **Genetic Evidence for Hybrid Trait Speciation in *Heliconius* Butterflies.** *PLoS Genet* **6**(4): e1000930.
- Samadi S. and Barberousse A. (2006). **The tree, the network, and the species.** *Biological Journal of the Linnean Society* **89**(3): 509-521.
- Sanmartin I., *et al.* (2001). **Patterns of animal dispersal, vicariance and diversification in the Holarctic.** *Biological Journal of the Linnean Society* **73**(4): 345-390.
- Scascitelli M., *et al.* (2010). **Genome scan of hybridizing sunflowers from Texas (*Helianthus annuus* and *H. debilis*) reveals asymmetric patterns of introgression and small islands of genomic differentiation.** *Molecular Ecology* **19**(3): 521-541.
- Schlick-Steiner B., *et al.* (2010). **Integrative taxonomy: a multisource approach to exploring Biodiversity.** *Annual Review of Entomology* **55**: 421-438.
- Schluter D. (2001). **Ecology and the origin of species.** *Trends in Ecology & Evolution* **16**(7): 372-380.
- Schöfl G., *et al.* (2009). **Time-shifted reproductive behaviours among fall armyworm (*Noctuidae: Spodoptera frugiperda*) host strains: evidence for differing modes of inheritance.** *Journal of Evolutionary Biology* **22**: 1447-1459.
- Schuelke M. (2000). **An economic method for the fluorescent labeling of PCR fragments.** *Nature Biotechnology* **18**(2): 233-234.
- Seehausen O. (2004). **Hybridization and adaptive radiation.** *Trends in Ecology & Evolution* **19**(4): 198-207.
- Servedio M., *et al.* (2009). **Reinforcement and learning.** *Evolutionary Ecology* **23**(1): 109-123.
- Seymour P. R., *et al.*, Eds. (1985). **Insects and other invertebrates found in plant material imported into England and Wales, 1984.** UK.
- Shaw K. L. (2002). **Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets.** *Proceedings of the National Academy of Sciences* **99**(25): 16122-16127.
- Simmons M. P. (2004). **Independence of alignment and tree search.** *Molecular Phylogenetics and Evolution* **31**(3): 874-879.
- Simon C., *et al.* (1994). **Evolution, Weighting and Phylogenetic Utility of Mitochondrial Gene Sequences and a Compilation of Covered Polymerase Chain Reaction Primers.** *Ann. Entomol. Soc. Am.* **87**(6): 651-701.
- Sites J. and Marshall J. (2004). **Operational criteria for delimiting species.** *Annu Rev Ecol Evol Syst* **35**: 199 - 227.
- Sites J. J. W. and Marshall J. C. (2003). **Delimiting species: a Renaissance issue in systematic biology.** *Trends in Ecology & Evolution* **18**(9): 462-470.
- Smadja C., *et al.* (2008). **Hitching a lift on the road to speciation.** *Molecular Ecology* **17**(19): 4177-4180.
- Smadja C. M. and Butlin R. K. (2011). **A framework for comparing processes of speciation in the presence of gene flow.** *Molecular Ecology* **20**(24): 5123-5140.
- Sota T. and Kubota K. (1998). **Genital Lock-and-Key as a Selective Agent against Hybridization.** *Evolution* **52**(5): 1507-1513.
- Sparks A. N. (1979). **A review of the biology of the Fall armyworm.** *The Florida Entomologist* **62**(2): 82-87.
- Stebbins G. L. (1959). **The Role of Hybridization in Evolution.** *Proceedings of the American Philosophical Society* **103**(2): 231-251.
- Supple M. A., *et al.* (2013). **Genomic architecture of adaptive color pattern divergence and convergence in *Heliconius* butterflies.** *Genome Research* **23**(8): 1248-1257.
- Swanson W. J. and Vacquier V. D. (2002). **The rapid evolution of reproductive proteins.** *Nature Reviews Genetics* **3**: 137-144.

- Symonds M. R. E. and Elgar M. A. (2004). **The mode of pheromone evolution: evidence from bark beetles**. Proceedings of the Royal Society of London. Series B: Biological Sciences **271**(1541): 839-846.
- Taylor J. W., *et al.* (2000). **Phylogenetic Species Recognition and Species Concepts in Fungi**. Fungal Genetics and Biology **31**(1): 21-32.
- Templeton A. R., *et al.* (1995). **Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum***. Genetics **140**(2): 767-782.
- The Marie Curie speciation Network (2012). **What do we need to know about speciation?** Trends in Ecology & Evolution **27**(1): 27-39.
- Thomas F., *et al.*, Eds. (2010). **Biologie Evolutive**. Bruxelles.
- Thomas Y., *et al.* (2003). **Genetic isolation between two sympatric host-plant races of the european corn borer, *Ostrinia nubilalis* Hubner I. sex pheromone, moth emergence timing, and parasitism**. Evolution **57**(2): 261-273.
- Turelli M. and Orr H. A. (2000). **Dominance, Epistasis and the Genetics of Postzygotic Isolation**. Genetics **154**(4): 1663-1679.
- Turner T. L., *et al.* (2005). **Genomic Islands of Speciation in *Anopheles gambiae***. PLoS Biol **3**(9): e285.
- Valen L. V. (1976). **Ecological Species, Multispecies, and Oaks**. Taxon **25**(2/3): 233-239.
- Velásquez-Vélez M. I., *et al.* (2011). **Reproductive Isolation between Two Populations of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) Collected in Corn and Rice Fields from Central Colombia**. Annals of the Entomological Society of America **104**(4): 826-833.
- Via S. (1999). **Reproductive Isolation between Sympatric Races of Pea Aphids. I. Gene Flow Restriction and Habitat Choice**. Evolution **53**(5): 1446-1457.
- Via S. (2001). **Sympatric speciation in animals: the ugly duckling grows up**. Trends in Ecology & Evolution **16**(7): 381-390.
- Via S. (2002). **The ecological genetics of speciation**. The American Naturalist **159**(supplement).
- Via S. (2009). **Natural selection in action during speciation**. Proceedings of the National Academy of Sciences **106**(Supplement 1): 9939-9946.
- Via S., *et al.* (2000). **Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments**. Evolution **54**(5): 1626-1637.
- Via S. and West J. (2008). **The genetic mosaic suggests a new role for hitchhiking in ecological speciation**. Molecular Ecology **17**(19): 4334-4345.
- Vuataz L., *et al.* (2011). **Toward a DNA Taxonomy of Alpine *Rhithrogena* (Ephemeroptera: Heptageniidae) Using a Mixed Yule-Coalescent Analysis of Mitochondrial and Nuclear DNA**. Plos One **6**(5): e19728.
- Walsh B. D. (1867). **The apple-worm and the apple maggot**. The Journal of Horticultural Science and Biotechnology **2**: 338-343.
- Wheeler Q. and Meier R., Eds. (2000). **Species Concepts and Phylogenetic Theory: A Debate**, Columbia University Press.
- Whitford F., *et al.* (1988). **Oviposition preference, mating compatibility, and development of two fall armyworm strains**. The Florida Entomologist **71**(3): 234-243.
- Whitworth T. L., *et al.* (2007). **DNA barcoding cannot reliably identify species of the blowfly genus *Protophormia* (Diptera: Calliphoridae)**. Proceedings of the Royal Society B: Biological Sciences **274**(1619): 1731-1739.
- Wicker-Thomas C. (2011). **Evolution of insect pheromones and their role in reproductive isolation and speciation**. Annales de la Société entomologique de France (N.S.) **47**(1-2): 55-62.
- Wiemers M. and Fiedler K. (2007). **Does the DNA barcoding gap exist? - a case study in blue butterflies (Lepidoptera: Lycaenidae)**. Frontiers in Zoology **4**(1): 8.
- Wilder J. A. and Hollocher H. (2003). **Recent radiation of endemic caribbean drosophila of the dunnii subgroup inferred from multilocus dna sequence variation**. Evolution **57**(11): 2566-2579.
- Wilson A. R., Ed. (1999). **Species : new interdisciplinary essays**, Lavoisier S.A.S.

- Wood T., *et al.* (2009). **The frequency of polyploid speciation in vascular plants**. Proceedings of the National Academy of Sciences USA **106**: 13875 - 13879.
- Wood T. K. and Sheldon I. G. (1982). **Ecological and Behavioral Basis for Reproductive Isolation in the Sympatric *Enchenopa binotata* Complex (Homoptera: Membracidae)**. Evolution **36**(2): 233-242.
- Wu C.-I. and Davis A. W. (1993). **Evolution of Postmating Reproductive Isolation: The Composite Nature of Haldane's Rule and Its Genetic Bases**. The American Naturalist **142**(2): 187-212.
- Xie X., *et al.* (2007). **Hawthorn-infesting populations of *Rhagoletis pomonella* in Mexico and speciation mode plurality**. Evolution **61**(5): 1091-1105.
- Yule G. U. (1925). **A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S.** Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character **213**: 21-87.

Résumé

Chez les insectes phytophages, l'adaptation à la plante hôte peut correspondre à l'une des premières étapes de la spéciation. Dans ce contexte, ce manuscrit s'intéresse à *Spodoptera frugiperda* (Lepidoptera: Noctuidae), un ravageur responsable de dommages importants sur de nombreuses cultures en Amérique et dans les Caraïbes. *Spodoptera frugiperda* présente deux variants, l'un adapté au riz et l'autre au maïs, et constitue un modèle biologique pertinent pour étudier ce mécanisme. En effet, les deux variants sont morphologiquement identiques, mais ils sont génétiquement différenciés et présentent des différences écologiques et comportementales. L'ensemble de ces caractéristiques suggèrent la présence d'un isolement reproducteur entre les deux variants, qui pourraient alors correspondre à deux espèces différentes. Cependant la présence d'hybrides dans la nature ainsi que des résultats controversés sur le succès d'accouplement entre les deux variants, rendent ambiguë le statut d'espèce de *S. frugiperda*. L'objectif de ce travail est donc de mieux estimer, à travers une étude intégrative, le niveau de différenciation génétique présent entre les deux variants de *S. frugiperda*. Grâce à un premier niveau d'analyse, menées à partir de population naturelles, il a été possible de mettre en évidence un niveau de différenciation génétique élevé entre les deux variants, qui est compatible à celui attendu entre deux espèces. Le deuxième niveau d'étude réalisé à partir de population de laboratoire nous a permis de mettre en évidence la présence d'un isolement reproducteur entre les deux variants, se traduisant notamment par une distorsion de ségrégation méiotique des marqueurs moléculaires chez les hybrides. L'ensemble de ces résultats supportent l'hypothèse selon laquelle le variant riz et maïs de *Spodoptera frugiperda* seraient des entités situées à une étape avancée dans le continuum de la spéciation.

Mots clés : bioagresseur, délimitation d'espèce, isolement reproducteur, Lépidoptère, polymorphisme

Integrative study of the status of the two host plant variants of *Spodoptera frugiperda* (Lepidoptera, Noctuidae)

Abstract

In phytophagous insects, adaptation to host plant could be the first step leading to speciation. In this context, this document focus on moth *Spodoptera frugiperda* (Lepidoptera: Noctuidae), a pest responsible for serious damages in several crops in the Western hemisphere. *Spodoptera frugiperda* consists of two host-plant strains, one adapted to rice and the other adapted to maize, which made it a relevant model to study this mechanism. Though the two variants are morphologically identical, they are nonetheless genetically distinguishable and present some ecological and behavioral differences. The species status of *S. frugiperda* is also highly controversial because hybrids naturally occur in the wild, not to mention the fact of the discrepancies among published results concerning mating success between the two strains. The aim of this thesis is thus to better estimate, through an integrative approach, the level of genetic differentiation between the two variants of *S. frugiperda*. Starting with natural populations, various phylogenetic methods allowed us to highlight a high level of genetic differentiation between the two variants, compatible with what is expected between distinct species. Furthermore, studies on laboratory populations, including crossing experiments, showed a significant unidirectional bias in inter-strain mating success and the presence of meiotic segregation distortion of molecular markers in hybrid progenies. These results as a whole support the assumption that the two strains of *S. frugiperda* are well-advanced in the continuum of speciation.

Key words: Lepidoptera, pest species, polymorphism, species delimitation, reproductive isolation