

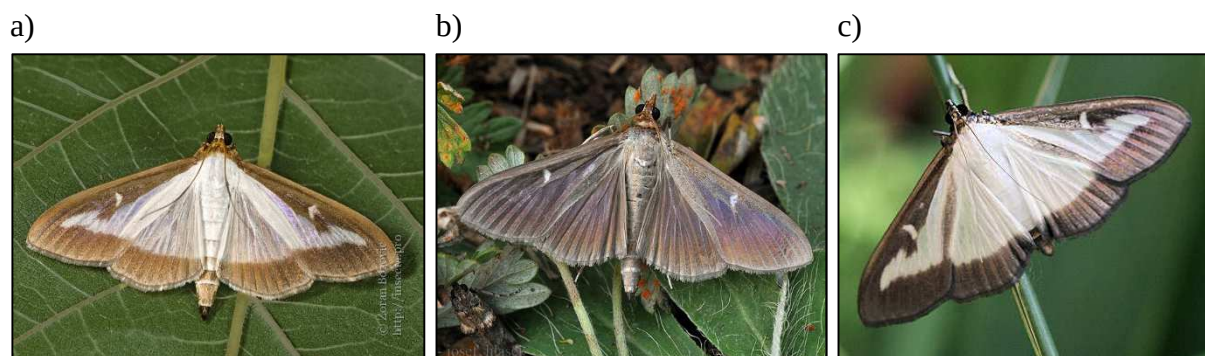
***Cydalima perspectalis*, la pyrale du buis.**

Cydalima perspectalis (Walker, 1859) est un lépidoptère de la famille des Crambidae et de la sous-famille des Spilomelinae. Avant d'en arriver à cette taxonomie, la pyrale du buis s'est vu attribuer différents noms de genre depuis son arrivée en Europe : *Palpita* Hübner, 1808, *Glyphodes* Guenée, 1854 et *Neoglyphodes* Streltsov, 2008. Finalement, en 2010, Mally et Nuss ont clarifié la situation suite à des études morphologique et la pyrale du buis fut classée dans le genre *Cydalima*.

1.1. La biologie de *Cydalima perspectalis*.

Ce papillon nocturne est facilement reconnaissable à l'âge adulte. Il présente des ailes blanches ou brunes avec des irisations dorées et violacées. Son envergure moyenne est de 36 mm et de 44 mm pour son maximum. Les deux sexes se ressemblent, le mâle ayant seulement à l'extrémité de son abdomen une coloration brune plus étendue. Il possède également des touffes de longues écailles odoriférantes dévaginables. L'imago de *C. perspectalis* ne peut être confondu avec d'autres espèces locales en raison de sa taille et de ses motifs caractéristiques blanc et brun. L'espèce présente trois formes distinctes (figure 1) : une forme classique blanche avec les extrémités des ailes et du corps brun ; une variante de la première forme avec en plus une bande brune sur le bord inférieur des ailes antérieures ; enfin, il existe une forme entièrement brune hormis une tache ovale blanche sur les ailes antérieures (Brua, 2014). C'est sous la forme imago que la pyrale du buis effectue sa propagation naturelle.

Fig. 1 : Imago de *Cydalima perspectalis* sous forme a) classique blanche, b) brune et bande brune. (Photo de Zoran Bozovic, Josef Hlasek et inconnu)



La femelle peut pondre jusqu'à 600 œufs en une semaine, près de 1100 œufs sur l'ensemble de sa vie, chaque ponte comportant une 20ème d'œufs sous forme de petits paquets (Tabone *et al*, 2015). Pendant ses 15 jours de vie imaginaire où elle ne s'accouple qu'une fois, la femelle peut pondre plusieurs fois. Les œufs sont ronds et aplatis et se chevauchent comme

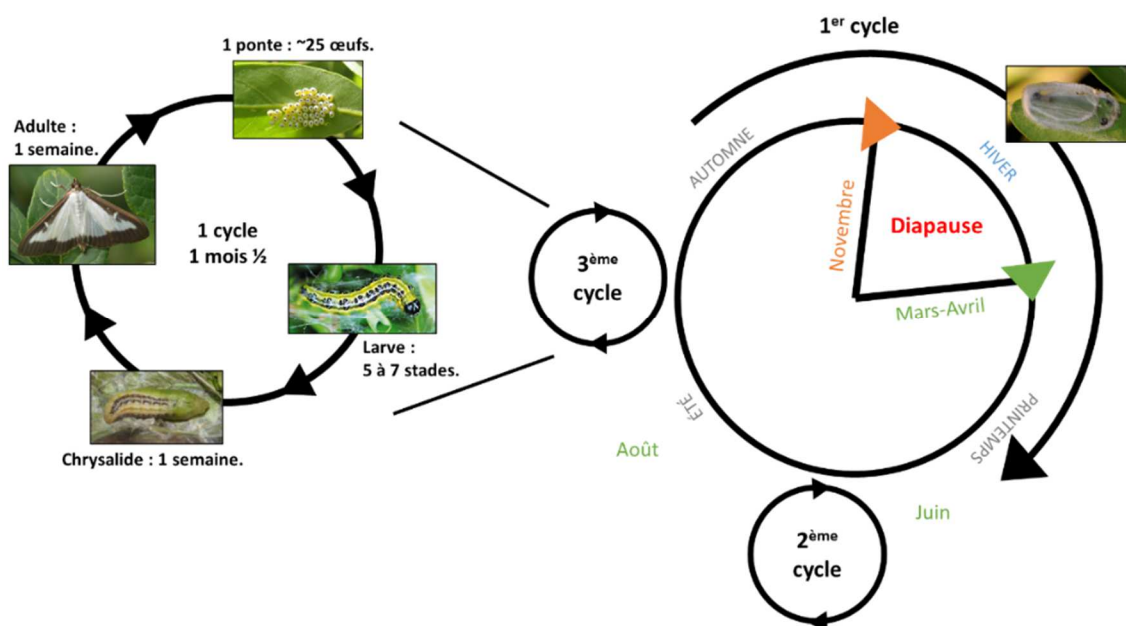
des écailles de poisson. A travers leur surface jaune et translucide, on peut observer un point noir, qui est la capsule céphalique de la chenille, juste avant l'éclosion.

Les larves possèdent 5 à 7 stades. Au dernier stade, la chenille peut faire jusqu'à 35-40 mm de long. Son corps est à dominance vert clair, strié longitudinalement de vert foncé, de jaune clair et de blanc (voir photo de couverture). Elle possède également des verrues noires et des longs poils blancs isolés. Au niveau de la partie thoracique, la chenille possède 3 paires de pattes jaunes et 5 paires de fausses pattes abdominales. Sa tête noire et luisante se démarque du reste du corps. A l'automne, les jeunes chenilles de dernière génération vont tisser un petit cocon entre deux feuilles de buis pour y faire leur diapause. Le développement s'arrête jusqu'aux premières chaleurs du printemps où elles reprennent leur activité et terminent leur développement.

Au stade nymphal, les chenilles à maturité tissent un cocon de soie. La chrysalide atteint 21 mm de long et est observable sur les feuilles de buis. Comme la chenille, elle est de couleur vert clair et jaune clair avec quatre lignes brun foncé séparé par des lignes claires au niveau dorsale. Sur l'arête dorsale s'y retrouve une bande brun orangé. La cuticule est translucide tout comme l'exuvie. Si elle est dérangée, elle agite vigoureusement son extrémité abdominale.

La pyrale du buis est un insecte polyvoltin et peut avoir 2 à 4 générations par an (site Ephytia). En France, elle effectue en général trois cycles de développement accompagné de 3 pics de dégâts : au début du printemps (mars - avril), en été (mi-juin - juillet) et au début de l'automne (septembre - novembre) (Brua, 2013) (figure 2). Après l'éclosion, le développement larvaire est rapide et peut durer entre 17 et 87 jours selon la température (15-30°C) et la source de nourritures (fiche technique Cabi). Les larves entrent en diapause l'hiver, celle-ci étant induite par les jours qui se raccourcissent (Wan, 2014).

Fig. 2 : Cycle biologique de la pyrale du buis. La chenille qui possède trois cycles dont une a la particularité de faire une diapause pendant l'hiver.



1.2. Les hôtes, les origines potentielles et les aires de propagation connues de *Cydalima perspectalis*.

L'origine de la pyrale du buis est asiatique, l'espèce étant présente naturellement en Chine, en Corée et au Japon. En 2016, Bras *et al.* ont montré que les haplotypes mitochondriaux retrouvés en France sont en faveur d'une origine chinoise, malgré une faible structuration génétique de la pyrale en Asie (et l'absence d'échantillons japonais). L'utilisation de marqueurs microsatellites et le recours à la méthode DIY ABC ont montré qu'il est vraisemblable qu'elle provienne de provinces côtières de Chine orientale (autour de Shangaï) qui font partie des principales régions chinoises productrices de buis. Elles ont également montré qu'il a eu au moins trois événements indépendants d'introduction à partir de cette seule région source. Dans son aire d'origine, la pyrale du buis fait son développement sur une quinzaine d'espèces appartenant au genre *Buxus* dont *Buxus microphylla*, très présent en Chine et utilisée en Asie comme plante ornementale dans les jardins et les parcs ; mais elle a également été signalée sur 3 autres espèces de ligneux : le houx à feuilles pourpres (*Ilex purpurea*), le fusain du Japon (*Euonymus japonicus*) et le fusain ailé (*Euonymus alatus*) (Casteels *et al.*, 2011). Les larves peuvent défolier complètement un buis et en raison de leur activité pendant toute la saison de croissance, et en cas d'attaques répétées, les plantes n'ont pas le temps de se rétablir et peuvent en mourir (Van der Straten *et Muus*, 2010).

C'est en mai 2007, dans une agglomération allemande proche de la Suisse et de la France : Weil-am-Rhein, que *C. perspectalis* a été signalée pour la première fois (Krüger, 2008). La pyrale fut sujette à une expansion fulgurante en Europe, elle est désormais présente dans une quinzaine de pays (Bella, 2013) : Suisse, Pays-Bas (2007), Slovénie, Autriche, France, Grande-Bretagne (2008), Belgique, Liechtenstein, Italie (2010), Turquie, Hongrie, République Tchèque, Roumanie (2011), Croatie (2012), Slovaquie, Danemark et Russie (2013) (figure 3).

En France, elle a été aperçue en premier à Saint-Louis en août 2008, proche de Weil am Rhein, puis rapidement à Strasbourg en octobre 2008 (Feldtrauer *et al.*, 2009).

En Europe, la pyrale du buis n'a pas été observée ailleurs que sur les buis ; de plus, le buis commun *B. sempervirens* est très utilisé en ornement (Brua, 2014). En France, *C. perspectalis* a mis moins de sept ans pour s'implanter avec certitude dans 17 régions et 52 départements (figure 4), démontrant la grande dynamique d'expansion du papillon.

Fig. 3 : Carte restituant les signalements de la pyrale du buis, *Cydalima perspectalis*, en Europe. L'étoile correspond au premier signalement à Weil-am-Rhein (Rousselet, 2018 ; Bella, 2013).

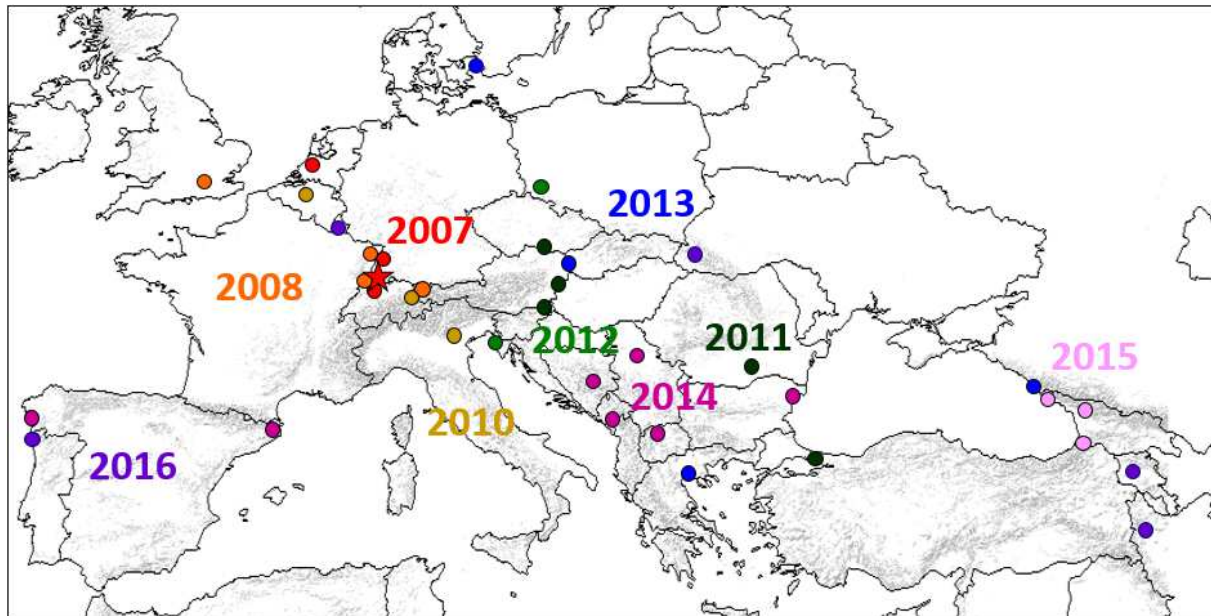
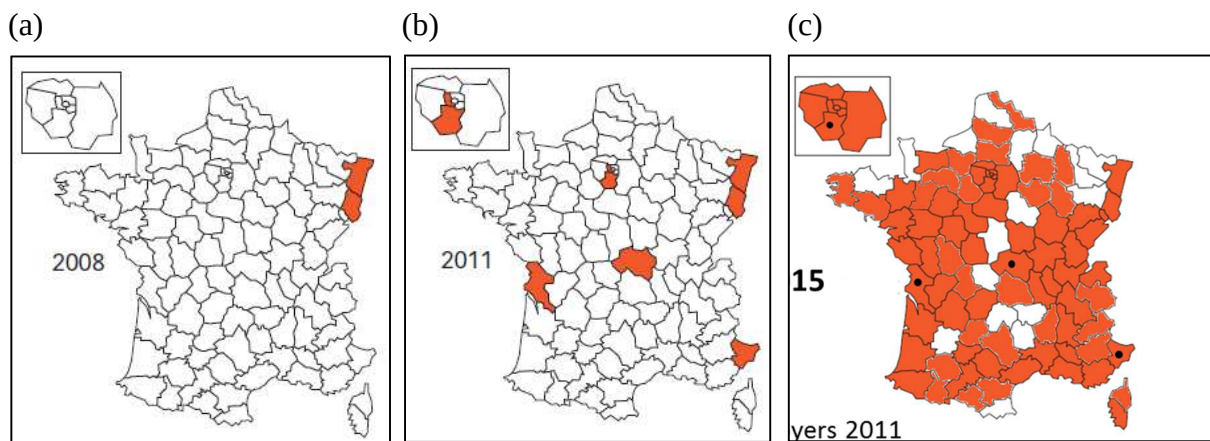


Fig. 4 : Cartes de présence certaine de la pyrale du buis *C. perspectalis* depuis sa première détection en France en 2008 jusqu'en 2016 (Brua, 2014). a) Première observation de la pyrale en France en Alsace 2008 ; b) En 2011, 4 nouvelles observations éloignées les unes des autres ; c) En 2016, plus de 79 départements sont envahis, montrant une invasion fulgurante en même pas dix ans.



2. Comprendre l'invasion biologique pour comprendre l'expansion de *Cydalima perspectalis*.

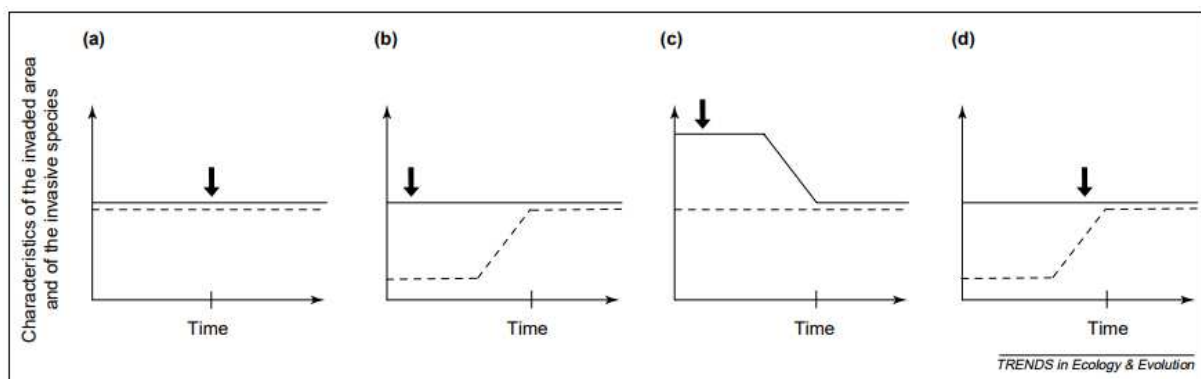
2.1. Quelques rappels de base sur les invasions biologiques.

L'invasion biologique est un phénomène qui se décompose en trois phases : l'introduction, l'établissement et l'expansion. Chaque phase agit comme un filtre, car dépend de la capacité des individus à franchir chacune des barrières rencontrées. Ainsi, ce processus peut échouer et l'introduction d'espèce exotique devenant envahissante est minoritaire en cas

d'absence ou de rareté en plantes-hôtes, de conditions climatiques défavorables, voire de présence de prédateurs, de pathogènes ou de compétiteurs dans la nouvelle aire de répartition. En absence de ces facteurs de régulation, l'espèce exotique se trouve favorisée par rapport aux autres espèces locales et peut connaître une expansion démographique et spatiale (annexe 3).

L'invasion biologique présente plusieurs scénarios, décrit par Facon *et al.* en 2006 (figure 5). Ils décrivent comment le succès des invasions biologiques peut être affecté par différents facteurs tels que la migration de l'organisme envahissant, les paramètres environnementaux de la zone envahie et l'évolution interne de l'organisme envahissant. Il existe trois scénarios d'invasions théoriques possibles : le changement de migration, le changement de l'environnement et le changement évolutif. Ces théories ont été basées sur des extrêmes, et des cas intermédiaires peuvent très bien se produire.

Fig. 5 : Scénarios d'invasion théorique (Facon *et al.*, 2006). Les lignes en pointillées représente les caractéristiques de la zone envahie, tandis que la ligne continue sont celles des espèces envahissantes. Le premier contact (par migration) potentiel entre l'espèce et la zone envahie et le début des processus d'invasion se font au niveau de la flèche. (a) Scénario 1 : l'invasion de fait par simple changement migratoire de l'espèce. (b) Scénario 2 : Le milieu subit des changements permettant l'installation d'espèce envahissante. (c) Scénario 3 : Une espèce subie des changements génétiques la rendant envahissante pour un milieu donné. (d) Scénario 1+2 : scénario mixte entre le 1 et le 2, il faut un changement dans la migration et le milieu pour déclencher une invasion.



2.2. Le rôle de la génétique dans l'invasion d'un nouveau milieu.

Le succès ou l'échec d'une invasion biologique est aussi influencé par des facteurs génétiques. En général, on considère que plus il y a de diversité génétique, plus l'espèce a de chances de s'adapter à son nouveau milieu, mais l'espèce, ou les individus introduits, peuvent aussi être, par chance, préadaptés à leur nouveau milieu et prédisposés à être envahissants. Cependant, il existe des paradoxes génétiques en biologie de l'invasion (Allendorf et Lundquist, 2003). Une population introduite est généralement caractérisée par une faible abondance et présente une diminution de diversité génétique causé par un goulot d'étranglement. Ce petit effectif peut donc provoquer une dérive génétique de la population introduite, mais également engendrer de la consanguinité. En parallèle, ces populations introduites vont souvent devoir côtoyer les espèces autochtones du nouveau milieu qui sont déjà adaptées aux conditions

locales. Cependant, l'absence de co-évolution avec les espèces introduites peut être un avantage majeur. Dans leurs nouvelles aires de répartition, une espèce exotique ne rencontrera pas forcément ses régulateurs naturels, favorisant leur expansion démographique (Colautti *et al.*, 2004).

Pour faire face à ces paradoxes, les populations introduites doivent avoir un effectif minimal assurant leur viabilité (Allendorf et Lundquist, 2003), ce qui peut se produire en cas d'introduction massive (rarement le cas) ou d'introductions multiples apportant du matériel génétique et contrebalançant l'effet de la dérive génétique (Sakai *et al.*, 2001). Lorsque ces premières introductions sont réussies, ces populations installées peuvent aussi servir de source pour l'infestation de nouvelles régions, tel que Lombaert et al, 2010 le décrit pour *Harmonia axyridis*, la coccinelle asiatique.

3. Les microsatellites : leur intérêt.

3.1. Les propriétés des microsatellites.

Les microsatellites sont composés de motifs d'un à six nucléotides répétés un grand nombre de fois dans les génomes nucléaires de la plupart des taxons. Ils sont donc aussi appelés : séquences répétées en tandem ou SSR (Simple Sequence Repeats). Un locus de microsatellite varie généralement d'une longueur entre 5 et 40 séquences répétées, même s'il est possible d'observer des chaînes plus longues de répétitions. Les répétitions de dinucléotides, trinucléotides et tétranucléotides sont les choix les plus courants pour les études génétiques moléculaires ; d'autant plus que les répétitions dinucléotidiques représentent la majorité des microsatellites pour de nombreuses espèces (Li *et al.* 2002). Les répétitions trinucléotides et hexanucléotidiques sont les classes de répétition plus susceptibles d'apparaître dans les régions codantes, car ils ne provoquent pas un cadre de lecture (Toth *et al.* 2000). Pour finir, les répétitions mononucléotidiques sont moins fiables en raison de problèmes avec l'amplification (Li *et al.* 2002). La variabilité étant la « matière première » du généticien, les microsatellites avec leur grand nombre d'allèles fournissent un haut degré d'information.

Ce qui rend les microsatellites intéressants est qu'ils possèdent un fort taux de mutation caractéristique (Kelkar *et al.*, 2010). Ce taux de mutation élevé rend les microsatellites fortement polymorphes et repose sur le nombre d'unités de répétition constituant le SSR, changeant d'un individu à l'autre. Ils mutent souvent par des erreurs de glissement et lors des checkpoints correctifs lors de la réplication de l'ADN ; ceux-ci modifient principalement le nombre de répétitions et donc la longueur de la chaîne de répétition. De ce fait, les allèles diffèrent en longueur et ils peuvent être distingués par électrophorèse sur gel à haute résolution ou dans un séquenceur à capillaires. De nombreux microsatellites ont des grands taux de mutation (entre 10^{-2} and 10^{-6} mutations par locus par génération, et en moyenne 5.10^{-4}) qui génèrent des niveaux élevés de diversité allélique nécessaire pour les études génétiques des processus agissant sur des échelles de temps écologiques (Schlötterer, 2000). Les marqueurs microsatellites sont plus variables que les séquences mitochondriales utilisés dans les études phylogéographiques, et contrairement à eux, ils sont généralement isolés à partir du génome

nucléaire et sont donc à hérédité biparentale. Contrairement à d'autres marqueurs moléculaires comme les AFLP³, ils présentent l'avantage d'être codominants (les deux allèles sont détectables chez un individu hétérozygote). Généralement plus abondants dans des régions non codantes, ils sont neutres vis-à-vis de la sélection et constituent de bons marqueurs pour inférer l'histoire des populations. Pour détecter le brassage génétique, les microsatellites sont assez efficaces (Clayton *et al.*, 1998 ; Gill, 2001) et sont particulièrement adaptés pour l'étude de l'*admixture*, quand deux populations précédemment isolées et génétiquement différenciées se mélangent (Vignal *et al.*, 2002). De plus, les microsatellites sont plus utiles pour détecter les expansions récentes de la population ; en effet, l'accumulation de nouvelles mutations, qui renseigne sur l'expansion de la population, nécessite des périodes courtes pour faire évoluer rapidement loci (Morin *et al.*, 2004).

3.2. Quels sont les limites des microsatellites ?

Les microsatellites présentent certains inconvénients : une phase de développement long et coûteux et un débit relativement faible en raison de difficultés pour la gestion de l'automatisation et des données. Par conséquent, l'utilisation continue de microsatellites dépendra probablement de la possibilité de surmonter certaines de ces limitations (Guichoux, 2011). Pour commencer, le grand nombre d'allèles par locus microsatellites, implique que pour une estimation précise des fréquences alléliques, il est nécessaire d'avoir des échantillons de grande taille. Ensuite, il y a plus de risque d'avoir beaucoup de mutations spontanées avec les microsatellites dans une lignée donnée, ce qui peut rendre difficile la reconstruction de la filiation génétique (Ellegren, 2000). A cela, le taux élevé de mutation récurrente des microsatellites, avec des risques d'homoplasie⁴ de taille, font d'eux de mauvais indicateurs de l'histoire des populations à une grande échelle de temps (Li *et al.*, 2002 ; Ellegren, 2004 ; Morin *et al.*, 2004 ; Schlötterer, 2004). De plus, la variabilité au niveau des marqueurs microsatellites hautement polymorphes peut ne pas exactement refléter la diversité génomique sous-jacente (Vali *et al.*, 2008 ; Ljungqvist *et al.*, 2010). Il peut également exister des allèles dits « nuls » du fait de mutation au niveau du site d'accrochage de l'amorce empêchant l'amplification par PCR de certains allèles. Les hétérozygotes allèle nul / allèle amplifié apparaissent homozygote pour l'allèle amplifié, ce qui fausse l'estimation de l'hétérozygotie et crée des écarts à l'équilibre d'Hardy-Weinberg artificiels.

La seule méthode utilisée pour le dosage des microsatellites est l'électrophorèse sur gel capillaire couplée à une détection à base de fluorescence (Butler *et al.*, 2001 ; Koumi *et al.*, 2004). Les amplifications PCR utilisés pour les SSR rendent plus difficile l'étude d'échantillons d'ADN fortement dégradés, tels que des matières fécales ou des échantillons mal conditionnés (Seddon *et al.*, 2005 ; Morin & McCarthy, 2007).

Stage

1. Matériels et méthodes

1.1. Matériel biologique.

Des chenilles de *Cydalima perspectalis* prélevées sur différents sites de la France (annexe 4) ont été utilisées pour les extractions des ADNs, ainsi que des imagos venant de Legnano et Lucques en Italie (30 individus chacun), de Delémont en Suisse (21), de Andel au Pays-Bas (28) et de Weil-am-Rhein (Allemagne) et Saint-Louis (France) regroupés en une population (30).

Des chenilles de différents stades larvaires ont été prélevées et conservées dans de l'alcool à 90°. La taille des échantillons était de 1 à 30 individus selon les sites. Les chenilles des premiers stades, petites, ont été utilisées en entier, tandis qu'un segment a été prélevé sur les plus grosses. Dans la mesure du possible, les chenilles génotypées provenaient de collectes sur des pieds de buis différents afin de minimiser les risques d'apparentement. Toutefois, pour certaines populations à basse densité, deux à quatre individus du même pied ont pu dans de rares cas être utilisés. Pour les adultes, une patte ou du muscle thoracique a été prélevée.

1.2. Extraction et purification de l'ADN.

L'extraction et la purification de l'ADN ont été faites avec le kit QIAGEN. Les échantillons de chenilles ont été lavés à l'eau distillée ; puis ont été placées dans des tubes Eppendorf contenant 180µl de tampon d'extraction (tampon de lyse Buffer ATL), avant d'être broyées manuellement. 20µl de protéase K ont été rajoutée et les tubes ont été incubés pendant toute une nuit au bain-marie à 56°C.

Le lendemain, 200µl de Buffer AL ont été rajoutées dans chaque tubes Eppendorf avant d'être incubées pendant 10 minutes au bain-marie à 56°C. 200µl d'éthanol à 96% ont été ensuite rajouté. Les solutions ont été pipetées et placées dans des colonnes DNeasy Mini spin avant d'être centrifugé pendant 1 minutes à 6000g. Les colonnes ont été récupérées et placées dans de nouveaux tubes collecteurs, les anciens contenant les déchets sont jetés. Ensuite, 500µl de Buffer AW1 ont été rajoutées puis les colonnes ont été centrifugées. Après récupération des colonnes, 500µl de Buffer AW2 ont été rajoutées. Les colonnes ont été centrifugées pendant 3 minutes à 20000g. Les colonnes ont été transférées dans des tubes Eppendorf puis 200µl de Buffer AE (solution d'élution) ont été rajouté. Après une incubation d'une minute à température ambiantes, les solutions ont été centrifugées pendant 1 minute à 6000 g. Les solutions dans les tubes Eppendorf contenant l'ADN ont été récupérées et les colonnes jetées. Puis l'ADN a été quantifié à l'aide d'un nanodrop

Le nanodrop est un spectropfluorimètre et il permet de vérifier la qualité et la quantité d'ADN et ainsi savoir si la solution d'ADN doit être diluée avant PCR. Pour cela, il utilise une longueur d'onde de 260 nm qui est la zone d'absorbance maximale des acides nucléiques. Une seconde mesure à 280 nm permet de contrôler la pureté d'extraction, c'est-à-dire la présence de protéines résiduelles dans la solution d'ADN en faisant le rapport de l'absorbance 280 nm sur

l'absorbance 260 nm. La machine a été initialisée avec 1,5µl d'eau distillée pour faire un blanc. Les échantillons d'ADN sont mesurés un par un avec un nettoyage à l'eau pure entre chaque mesure. Si la solution contient plus de 50 ng/µl d'ADN, elle sera diluée lors de la mise en plaque. Le rapport de pureté 260/280 doit être compris entre 1,80 et 2.

1.3. PCR en plaque.

Pour l'optimisation de la PCR, le kit Quiagen contenant une solution Mix MULTIPLEX PCR, une solution de RNase et de Q solution a été utilisé. Dans le cas de la pyrale du buis, 15 marqueurs ont été développés et répartis dans 3 mix en fonction de leur température d'hybridation (annexe 5). Une fois le mix préparé, 115µl ont été déposés par puits dans une barrette, puis à l'aide d'une pipette multicanaux, 9µl ont été déposés par puits sur la plaque. Les puits ont été complétés avec 1µl d'ADN à l'aide d'une plaque ADN préparé précédemment avant d'être passé au thermocycleur.

Sous la hotte, 960 µl de formamide + 28,8 µl de marqueur de taille GeneScan 600LIZ ont été mis dans un tube. La solution a été répartie dans une barrette avec environ 110 µl par puits. Puis 9 µl par puits ont été réparties sur la plaque et ensuite 1 µl d'ADN de chaque PCR. Un tapis blanc a été mis sur la plaque avant de sortir de la hotte. Ensuite, la plaque a été dénaturée pendant 3 minutes à 96°C puis placée directement sur la glace. Le tapis blanc a été remplacé par un tapis gris et conservé à 4°C pour envoi au séquenceur ou mis au congélateur en attendant de la passer au séquenceur. Les tailles de fragments obtenues après le séquençage, ont été lues à l'aide du logiciel Genemapper®.

1.4. Analyse des données génétiques.

L'analyse des données génétiques a pour objectif de caractériser la structure génétique des populations et d'estimer les flux de gènes entre celles-ci. Ainsi, une différenciation génétique forte signifie peu d'échanges d'individus entre les populations.

1.4.1. Approche population-centrée.

Pour cela, le logiciel Genetix (v. 4.05) (Belkhir et al 1996-2004) a été utilisé. Il permet de calculer les valeurs des indices F de Wright (Wright, 1969 ; Weir & Cockerham, 1984) dont Fis et Fst. Ces indices s'intéressent à la distribution de la diversité génétique entre différents niveaux hiérarchiques que sont les individus (niveau I), les sous-populations (niveau S pour *subpopulation* dans la nomenclature de Wright ; on parle plus facilement habituellement de populations en langage commun) et la population théorique totale (niveau T). Ces indices sont calculés à l'échelle de la population totale ou à l'échelle des (sous-)populations à partir d'échantillons de population composés de plusieurs individus qui doivent être en nombre suffisant pour être représentatif de celle-ci (approche population-centrée). De ce fait, dans le cadre de l'étude, seuls les Fis et Fst des populations avec un effectif ≥ 20 ont été calculés.

Lorsque la distribution de la variabilité génétique s'écarte de celle attendue sous l'hypothèse d'Hardy-Weinberg, cela signifie que des régimes de reproduction (homogamie, hétérogamie, autofécondation, consanguinité, ...) ou des forces évolutives (migration, sélection,

dérive, mutation) ont modifié les fréquences alléliques et/ou les fréquences génotypiques (appréhendées synthétiquement via le degré d'hétérozygotie, c'est à-dire la fréquence relative des hétérozygotes et des homozygotes considérée à différents niveaux hiérarchiques).

1.4.1.1. Structuration de la diversité génétique au sein des populations.

L'indice F_{IS} renseigne sur les écarts entre hétérozygotie observée et celle attendue selon l'équilibre de Hardy-Weinberg au sein des sous-populations. Il considère donc comment se distribuent les allèles d'une population entre les individus ($F_{IS} = [\overline{H_S}] - H_I / \overline{H_S}$). Cet indice varie de -1 à 1 et donne des informations en particulier sur le degré de consanguinité existant au sein d'une population ($F_{IS} > 0$ traduisant un excès d'homozygotes par rapport à H-W, ce qui est par exemple le cas lors d'une reproduction entre apparentés génétiquement proches) ou encore sur l'importance d'immigration récente ($F_{IS} < 0$ traduisant un excès d'hétérozygotes). Un $F_{IS} > 0$ peut également être dû à des marqueurs qui ne sont pas à l'équilibre notamment du fait de la présence d'allèles nuls. Un F_{IS} nul signifie que les conditions d'Hardy-Weinberg sont globalement respectées (panmixie, taille efficace $\sim \infty$, ...) et la population est dite à l'équilibre (*i.e.* les fréquences génotypiques sont celles attendues sous l'équilibre d'Hardy-Weinberg). La significativité de l'écart à zéro des valeurs de F_{IS} est testée au moyen de tests statistiques non paramétriques (qui ne supposent pas que les données suivent une loi de distribution connue avec un nombre de paramètres définis à l'avance). A partir du jeu de donnée réel initial, des jeux de données sont simulés en permutant les allèles à l'intérieur de chaque population, créant ainsi des génotypes au hasard, et une nouvelle valeur de F_{IS} est calculée pour chaque jeu simulé. La valeur de F_{IS} observée sur les données réelles est ensuite comparée à la distribution des valeurs simulées, ce qui permet d'obtenir une estimation de la probabilité d'obtenir une valeur supérieure ou égale à la valeur observée sous l'hypothèse nulle (équilibre d'H-W). Si la proportion de valeurs simulées supérieures à la valeur observée sur données réelles est inférieure au seuil de 5%, la valeur de F_{IS} est considérée significativement différente de zéro et l'hypothèse nulle est rejetée. Inversement, si la proportion de valeur est supérieure à 5%, l'écart est non significatif et la population est considérée à l'équilibre. Trivialement, cela signifie que les permutations au hasard, qui normalement devraient le plus souvent « casser » une structure qui existerait, ont été capables de créer dans plus de 5% des cas des jeux de données avec plus de structuration que les observations, ce qui indique en fait l'absence de structuration dans le jeu de donnée réel. Nous avons réalisé ici 5000 permutations.

1.4.1.2. Structuration de la diversité génétique entre populations.

L'indice F_{ST} quantifie le degré de différenciation des sous-populations résultant d'une subdivision de l'ensemble, à travers la quantification des écarts entre l'hétérozygotie observée dans les sous-populations et celle attendue dans une population théorique totale à l'équilibre d'H-W ($F_{ST} = [H_T - \overline{H_S}] / H_T$). Il considère donc comment se distribue l'ensemble des allèles entre les différentes sous-populations. Cet indice varie de 0 à 1. Dans cette étude, l'estimateur de Weir & Cockerham (1984) a été utilisé pour des données multilocus multiallèles sans considérer d'hypothèses particulières en matière de nombre de populations, de taille des échantillons et de fréquence des hétérozygotes. Les valeurs de F_{ST} ont été estimées par paire de

populations et présentées sous forme d'une demi-matrice donnant les valeurs de F_{ST} entre les différentes populations. La significativité de l'écart à zéro des valeurs de F_{ST} observées, indiquant ici une structuration spatiale, a également été testée par permutations sur le même principe que les F_{IS} , mais cette fois par permutation globale des individus (permutations sur les génotypes et non sur les allèles). Nous avons réalisé ici 1000 permutations.

1.4.2. Approche individu-centrée.

Ce type d'approche se base, non sur un échantillon de population (composé d'un nombre minimal d'individus) à partir duquel on calcule des indices (F statistiques) traduisant la façon dont se répartit la diversité génétique entre individus au sein des populations (F_{IS}) et entre populations (F_{ST}), mais sur le génotype multilocus des individus eux-mêmes pour les assigner à des groupes dont les membres partagent des patrons similaires de variation génétique. Un des intérêts est notamment de pouvoir travailler sur des échantillons de très petite taille pour un site donné (jusqu'à l'individu lui-même seul). Il est donc possible d'intégrer également dans son jeu de données des individus de sites où la densité de population est très faible.

Nous avons ici utilisé le logiciel Structure (Pritchard, 2000) qui, pour former ces groupes, analyse les différences dans la distribution des variants génétiques entre populations avec un algorithme itératif bayésien (Pritchard, 2000 ; Porras-Hurtado, 2013). Pour un nombre de *clusters*⁵ (K) fixé, l'algorithme cherche à constituer K groupes en minimisant le déséquilibre d'Hardy-Weinberg (groupes génétiquement les plus homogènes possible). Les individus sont assignés à ces *clusters* avec un coefficient d'appartenance à chaque *cluster* (*membership*). Dans les approches population-centrées, on utilise plusieurs marqueurs surtout pour éviter de raconter l'histoire d'un locus en particulier (qui n'est pas forcément représentative de l'évolution neutre de l'ensemble du génome) et on tire partie de l'information associée à différents degrés de polymorphisme des marqueurs. Chaque locus contribue à l'information globale, mais est considéré indépendamment. Dans cette approche individu-centrée, on tire cette fois pleinement partie de l'information multilocus en cherchant également à minimiser le déséquilibre de liaison. Du « métissage » (*admixture*) entre des populations précédemment isolées et différenciées génétiquement se traduit par des associations préférentielles d'allèles à différents loci qui mettent plus ou moins de générations à se résorber. Contrairement aux méthodes multivariées, c'est une méthode qui repose sur un modèle explicite dont il faut respecter les hypothèses sous-jacentes (Pritchard, 2000 ; Porras-Hurtado, 2013) ; Jay, 2011). L'identification de la structure hiérarchique avec Structure peut notamment poser problème en cas d'isolement par la distance (IBD) mais dans notre cas on s'attend à ce que la structuration de la diversité génétique soit fortement et principalement modelée par les événements d'introduction. Avec l'algorithme de Structure, on explore différents niveaux de *clusterisation* (et donc la structure hiérarchique des populations) et on cherche à en interpréter la pertinence biologique et à déterminer le nombre de *clusters* optimal rendant compte au mieux du jeu de données. Il s'agit donc de détecter le niveau hiérarchique de structure le plus élevé pour les scénarios testés. Pour chaque K , nous avons effectué 10 *runs* avec les paramètres suivants : 50 000 mcmc (*Markov Chain Monte Carlo*) 50 000 burnins, *admixture* model (pour le choix de l'*ancestry model*),

correlated allele frequencies model (pour le choix de l'*allele frequency model*) (Pritchard et al, 2009 ; Falush *et al.*, 2003). Pour chaque simulation (*run*), une valeur de probabilité postérieure (conditionnelle) est calculée (« *estimated Ln Prob of data* ou $\ln \Pr(X|K)$ où X sont les données et K le nombre de *clusters*). Pour déterminer le nombre de K optimal, nous avons commencé par considérer l'évolution du ratio $\Pr(X|K)$ pour un cas donné sur la somme de celui-ci pour tous les K. Par exemple pour K=1, nous avons : $Posterior\ Prob\ K1 = \frac{e^{\ln \Pr(X|K1)}}{e^{\ln \Pr(X|K1)} + e^{\ln \Pr(X|K2)} + \dots + e^{\ln \Pr(X|Kmax)}}$ Mais nous avons ici principalement utilisé le critère ΔK d'Evanno (2005) à l'aide du programme Structure Harvester (Earl & von Holdt, 2012 ; <http://taylor0.biology.ucla.edu/structureHarvester/>) qui permet une meilleure estimation du K optimal. La statistique ΔK est basée sur le taux de changement dans le logarithme de la probabilité des données entre valeurs successives de K.

2. Résultats.

Au total, 571 individus ont été génotypés parmi les 49 localités échantillonnées. Certains individus ont été mal ou pas amplifiés en raison, sans doute, d'une mauvaise conservation lors de la collecte ou de l'envoi. Mais ils restent très minoritaires en comparaison au nombre d'individus traités, ainsi, tous les échantillons de population disponibles ont pu être analysés.

2.1. Approche population-centrée.

2.1.1. Structuration de la diversité génétique au sein des (sous-)populations.

Les valeurs de Fis sont présentées dans le tableau 1 (le détail par type de marqueurs est présenté dans l'annexe 6). La majorité des populations ont des Fis proches de 20%, ce qui très élevé. Nantes possède la valeur la plus élevée (34%) alors que Blois (4%) et Pyrénées (9%) ont les valeurs les plus basses.

Tableau 1 : Valeurs de Fis pour les échantillons de population dont l'effectif est supérieur ou égal à 20 individus (toutes sont significatives au seuil de 5%).

Population	Fis
Nice	0.19516
Strasbourg	0.25448
Saintes	0.23716
La Rochelle	0.22747
Legnaro	0.29251
Paris	0.18891
Creuzier-le-Vieux	0.23723
Orléans	0.23422
Nantes	0.34455
Chartres	0.11308
Tours	0.16573
Blois	0.04464
Pyrénées	0.09919
Lucques	0.17815
Weil-Am-Rhein	0.14829
Andel	0.18484
Délémont	0.27219

En regardant les Fis des marqueurs population par population (annexe 6), certains se sont révélés très déséquilibrés dans tous les échantillons de population ou presque (en particulier BTM15, BTM22 et BTM40). En les retirant, on améliore nettement les valeurs Fis qui se rapprochent de l'équilibre cependant, cela ne change pas énormément les Fst et le *clustering* obtenu. De ce fait, le choix a été fait de tous les garder en première approche.

2.1.2. Structuration de la diversité génétique entre populations.

Toutes les valeurs de Fst sont significativement différentes de zéro au seuil de 5% (tableau 2). Les valeurs sont très élevées entre les primo-foyers de Creuzier-le-Vieux, Paris et Strasbourg (de 18 à 28% selon les comparaisons deux à deux), entre Saintes et Strasbourg (22,5%) et entre Nice et Strasbourg (20,8%). Elles sont modérées entre Nice et Creuzier-le-Vieux ou Paris (respectivement 8,5% et 9,5%) et faible entre Nice et Saintes (3,5%).

Tableau 2 : Demi-matrices représentant respectivement les FST entre les populations des primo-foyers de pyrale du buis et les distances géographiques euclidiennes entre eux.

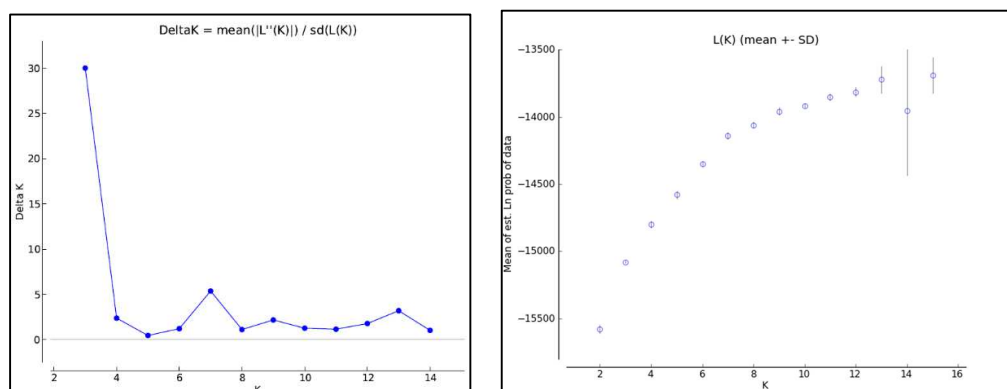
FST / km	Nice	Strasbourg-Kehl	Saintes	Paris	Creuzier	WeilaR
Nice	-	0.20837	0.03458	0.09474	0.08579	0.12099
Strasbourg-Kehl	545 km	-	0.22520	0.27822	0.18211	0.11363
Saintes	665 km	708 km	-	0.07430	0.07238	0.15556
Paris	688 km	397 km	413 km	-	0.17637	0.18912
Creuzier	408 km	422 km	318 km	311 km	-	0.13415
WeilaR	435 km	111 km	663 km	415 km	356 km	-

2.2. Approche individu-centrée.

2.2.1. Choix du meilleur niveau hiérarchique.

Les résultats de Structure et de Structure Harvester ont donné l'optimum à K=3. La valeur la plus élevée de ΔK est obtenue pour ce nombre de *clusters* (figure 6a). En parallèle, l'évolution des estimations du $\ln P(X|K)$ montre un gain d'information jusqu'à K=7, valeur de K à partir de laquelle on tend à atteindre un plateau (et à partir de 13 une forte variabilité entre *runs*) (figure 6b). Le nombre de *clusters* considéré principalement pour l'interprétation des résultats a donc été K=3.

Fig. 6 : a) Evolution de ΔK en fonction de K b) Estimation du $\ln P(X|K)$ en fonction de K.



2.2.2. Structuration spatiale pour K=3.

A l'échelle de la France (figure 7), les trois *clusters* ainsi définis sont spatialement très structurés. Le *cluster* en rouge sur la figure 7 a une distribution centrée sur l'Est du pays, le long du Rhin au niveau des frontières franco-allemande et franco-suisse. Dans cette région, les individus sont en très grande majorité très proches génétiquement les uns des autres et les populations sont donc très homogènes, composées presque uniquement d'individus assignés à ce seul *cluster*. A Weil-Am-Rhein et Saint-Louis (le premier foyer européen de 2007 en Allemagne qui s'est étendu en France en 2008), 28 individus sur 30 ont un coefficient de *membership*⁶ supérieur à 70% pour ce *cluster* dont 25 sur 30 supérieur à 90% (et seulement $2/30 \leq 50\%$). A Strasbourg (et Kehl) où la pyrale a été signalée en 2008, 29 individus sur 30 ont un coefficient de *membership* supérieur à 90% pour le *cluster* « rouge » (et seulement $1/30$ un *membership* < 20%), A Délémont en Suisse, on trouve 19 individus sur 21 avec des valeurs de *membership* supérieures à 80% (dont $17/21 > 90\%$ et seulement $2/21 < 5\%$). A l'exception de Nice où le génotype de 15 individus sur 30 est à plus de 50% assigné à ce *cluster*, on le retrouve dans les autres populations avec un niveau d'affiliation (*membership*) très faible ($83/571$ individus > 5% ; $31/571$ entre 10 et 50%, et seulement $22/571$ entre 50 et 70%).

La distribution du *cluster* en vert est centrée sur l'Ile-de-France. La plupart des individus autour du premier foyer d'IdF (observés dans l'Essonne en 2009) sont assignés en totalité à ce *cluster* (29 individus sur 40 sur l'ensemble Paris-Rungis-Soisy avec un *membership* > à 90%, et seulement $5/40 < 50\%$ et $3/40 < 80\%$). Ce *cluster* se rencontre dans les autres régions de France et d'Europe (dont Nantes, La Rochelle, Saintes, Pyrénées, Nice, Weil-am-Rhein, Délémont, Italie, Pays-Bas) mais avec des contributions faibles (voir carte 10), à l'exception de sites de la région Centre Val de Loire (Chinon, Montrichard, Valençay, Issoudun, Le Blanc, Blois, Menetou-Salon) où un grand nombre d'individus présentent un coefficient de *membership* supérieur à 50% (39 sur 47) et même supérieur à 90% (29 sur 47 ; voir carte 11).

Le *cluster* « bleu » est fortement représenté dans sud et l'ouest de la France ainsi qu'en Italie et aux Pays-Bas. C'est à Creuzier-le-Vieux et Saintes (deux primo-foyers français) et à Lucques (près de la principale zone d'import-export italienne) que la proportion d'individus présentant un coefficient de *membership* supérieur à 90% est la plus élevée (21 individus sur 30, 20 individus sur 30 et 25 individus sur 30 respectivement). Nice, et dans une moindre mesure Andel aux Pays-Bas, sont les localités où les 3 *clusters* co-occurrent le plus.

A l'échelle de la Région Centre (figure 8), les *clusters* vert et bleu dominant mais ne sont pas structurés en deux groupes spatialement distincts. Un grand nombre d'individus ne sont pas assignés en totalité à l'un ou l'autre des *clusters* (voir figure 8). 145 sur 268 de l'ensemble de la région possèdent moins de 80% d'assignation au *cluster* vert et 177/268 au *cluster* bleu. Il y a un effet de « patchwork ».

Fig. 7 : Distribution spatiale à l'échelle de l'Europe de l'Ouest des coefficients individuels d'assignation (membership) pour $K=3$ (seuls 24 sites d'échantillonnage sont représentés)

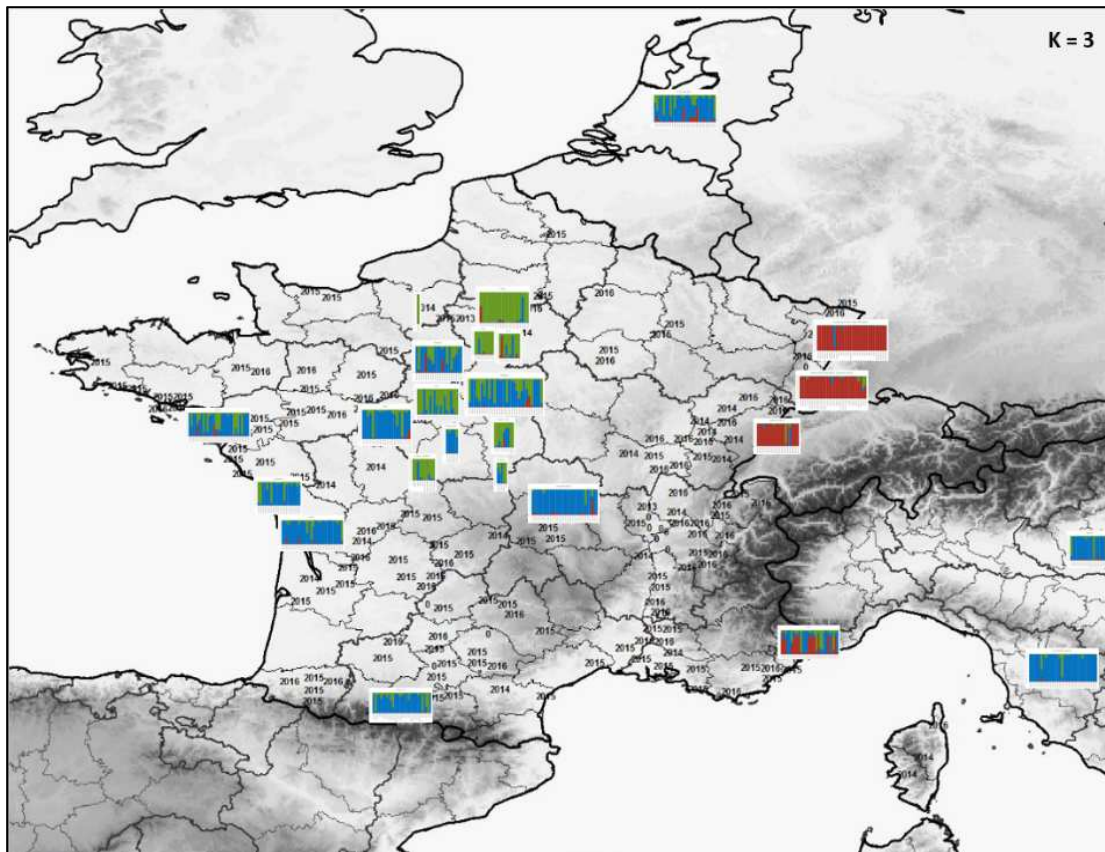
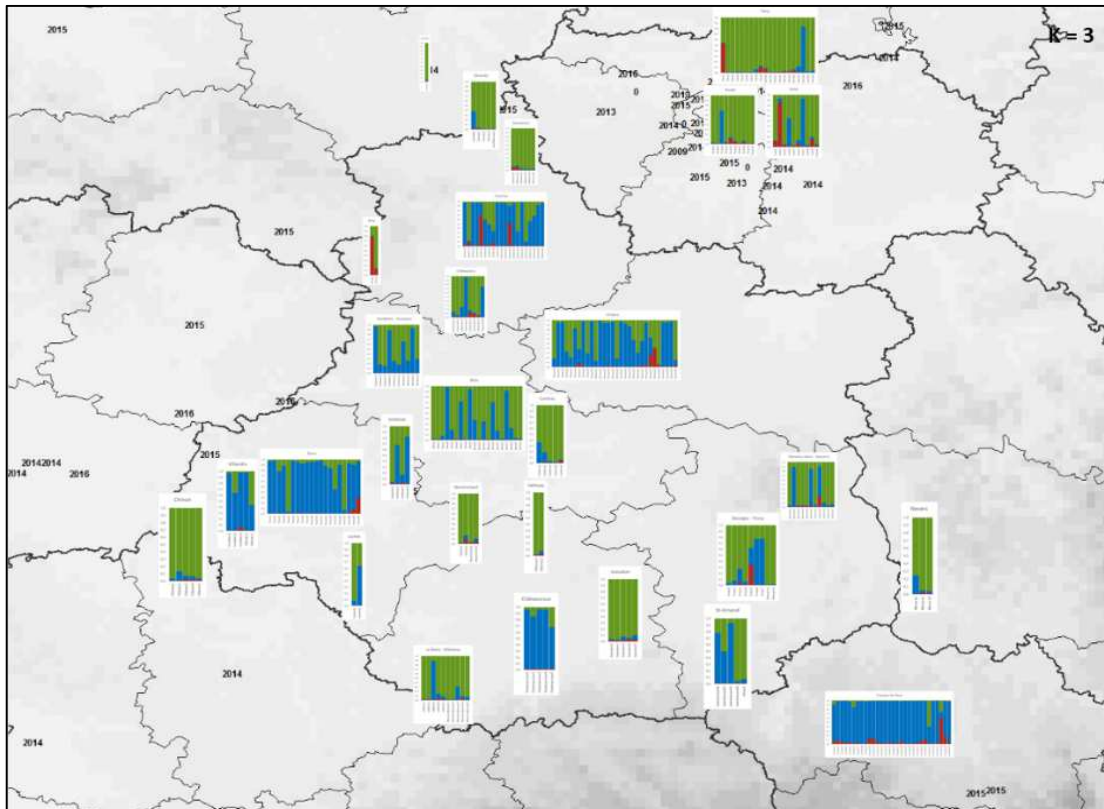


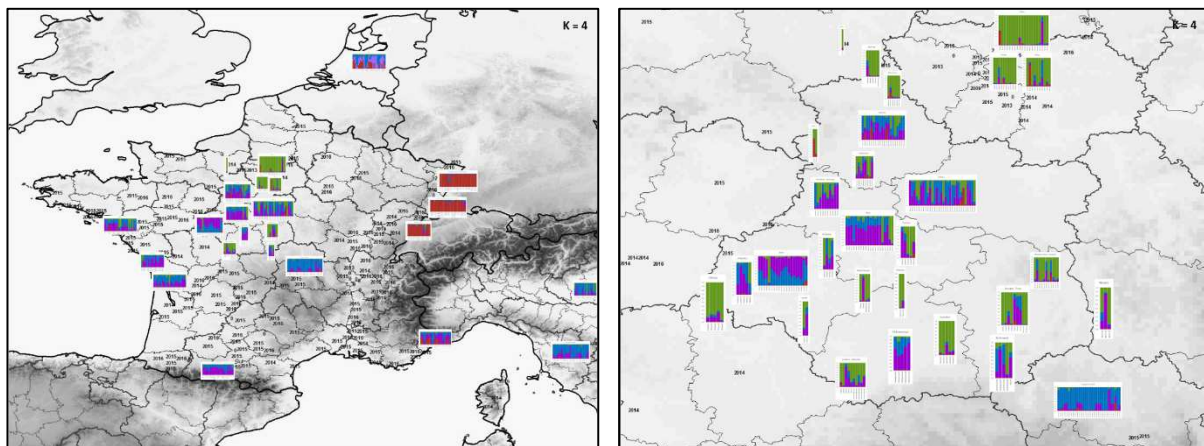
Fig. 8 : Distribution spatiale à l'échelle de la région Centre Val de Loire des coefficients individuels d'assignation (membership) pour $K=3$ (tous les échantillons sont représentés)



2.2.3. Structuration spatiale pour $K > 3$.

Pour $K=4$ (figure 9), seuls 4 individus sur les 571 de l'ensemble du jeu de données sont assignés à plus de 90% au nouveau *cluster* et seulement 28/571 à plus de 80%. Il est fortement représenté uniquement dans les populations qui présentaient des signatures d'*admixture* pour $K=3$. Pour les trois autres *clusters*, leur distribution correspond respectivement pour l'essentiel aux régions où les *clusters* à $K=3$ étaient presque exclusivement ou fortement représentés. Les résultats sont similaires pour $K=5$ et $K=6$ (voir annexe 7ab).

Fig. 9 : Distribution à l'échelle de la France et de la Région Centre VdL des coefficients individuels d'assignation (*membership*) pour $K=4$.



3. Discussion.

3.1. Structure hiérarchique des populations en Europe et en France : plusieurs introductions.

Bras et al (2016 ; en préparation) ont montré à l'aide de marqueurs mitochondriaux que les populations de la zone d'introduction étaient principalement structurées en trois grands ensembles : l'Europe de l'Ouest où l'insecte a été observé pour la première fois et où la diversité haplotypique était la plus élevée, l'Europe Centrale et les Balkans, et l'Europe orientale et le Moyen-Orient. L'analyse de la structure hiérarchique des populations d'Europe de l'ouest réalisée ici montre également une forte structuration spatiale de la diversité génétique au sein de ce territoire qui n'avait pas été mise en évidence avec les marqueurs mitochondriaux.

La structure des populations indique qu'il existe au moins 3 événements indépendants d'introduction sur ou à proximité territoire français. Le critère d'Evanno (2005), ΔK , indique en effet que le nombre optimal de *clusters* est $K=3$ pour ce jeu de données (figure 6a). Des événements d'introductions indépendants, accompagnés de fortes distorsions de fréquences alléliques dues au goulot d'étranglement (effet de fondation), peuvent rendre compte des valeurs élevées de l'indice de fixation F_{st} (qui mesure le degré de différenciation) entre les trois paires de *clusters* inférés par Structure (de 10 à 13% environ selon les paires). Il est par ailleurs à noter que les valeurs de F_{st} calculées à l'aide de Genetix entre les premiers foyers français

caractéristiques de ces trois *clusters* atteignent même 28% entre Strasbourg/Kehl et Paris, 18% entre Weil-am-Rhein et Paris, 18 % entre Strasbourg/Kehl et Creuzier-le-Vieux, 16% entre Weil-am-Rhein et Saintes, et 17% entre Paris et Creuzier-le-Vieux. Ces processus d'introduction et de goulot d'étranglement peuvent également rendre compte de valeurs de *F_{is}* parfois élevées, avec des populations non encore revenues à l'équilibre, mais au vu des capacités de dispersion, du nombre de génération par an, et de la prolifération rapide de l'insecte, ces dernières pourraient en fait être liés à des problèmes avec certains marqueurs ou certains échantillonnages.

Tableau 2 : Demi-matrices représentant respectivement les *F_{ST}* entre les populations des primo-foyers de pyrale du buis et les distances géographiques euclidiennes entre eux.

FST / km	Nice	Strasbourg-Kehl	Saintes	Paris	Creuzier	WeilaR
Nice	-	0.20837	0.03458	0.09474	0.08579	0.12099
Strasbourg-Kehl	545 km	-	0.22520	0.27822	0.18211	0.11363
Saintes	665 km	708 km	-	0.07430	0.07238	0.15556
Paris	688 km	397 km	413 km	-	0.17637	0.18912
Creuzier	408 km	422 km	318 km	311 km	-	0.13415
WeilaR	435 km	111 km	663 km	415 km	356 km	-

La distribution dans l'espace des trois *clusters* montre qu'il existe des régions où les populations sont nettement différenciées génétiquement et présentent très peu d'*admixture* entre les *clusters* identifiés. Ceci correspond à la signature génétique attendue de populations introduites indépendamment les unes les autres, et servant de source aux individus à l'origine d'*admixture* en se propageant vers d'autres foyers.

La première de ces régions correspond à celle du Rhin au niveau des frontières franco-allemande et franco-suisse et inclut Weil-am-Rhein / Saint-Louis, le premier foyer détecté en Europe (2007) et en France (2008), et Strasbourg / Kehl, le second détecté en France (2008). Un seul événement d'introduction suivi d'une propagation dans les environs par dispersion active (vol du papillon) est suffisant avec le présent jeu de données pour rendre compte de la distribution spatiale du *cluster* qui y prédomine très largement. Les distances entre le premier foyer et Délémont sont en effet compatibles avec les capacités de vol observées au laboratoire (même si celles-ci sont probablement surestimées par rapport aux conditions naturelles). En manège de vol, un papillon parcourt en moyenne 23 km sur l'ensemble de sa vie, avec un maximum enregistré à 71 km, ce qui représente par extrapolation, 75 km par an en moyenne, pour un maximum de 200 km pour les meilleurs disperseurs (Bras et al, en préparation). Toutefois, on ne peut pas exclure qu'il y ait eu plusieurs introductions plus ou moins indépendantes dans la région. Même s'ils appartiennent au même cluster (figure 7 et 9 pour $K=3-4$, annexe 7a pour $K=5-7$), les foyers de Weil-am-Rhein / Saint-Louis et Strasbourg / Kehl sont relativement fortement différenciés ($F_{st} = 12\%$) au vu de la faible distance géographique qui les sépare (111 km), et ils ont été détectés à seulement quelques mois d'intervalle, ce qui laisse peu de temps pour une propagation par dispersion active malgré la faible distance. Les

zones où prédomine le second *cluster* correspondent principalement à l'Ile de France où a été détecté un des 5 premiers foyers français en 2009 dans l'Essonne (à proximité duquel sont distribués nos échantillons de Rungis, Paris et Soisy). Celui-ci correspond donc bien à un autre événement d'introduction indépendant du premier, et non à un signalement partiel du processus de colonisation du territoire (avec une présence de l'insecte entre les premiers foyers distants les uns des autres mais une absence de signalements). Le troisième *cluster* a une distribution beaucoup plus vaste, et plusieurs sites dispersés sur le territoire français et européen montrent peu d'*admixture* avec donc une très forte prédominance de ce *cluster*. Les régions où il prédomine très largement rassemblent ainsi les Pyrénées, mais aussi, et surtout, deux des cinq premiers foyers français (Creuzier-le-Vieux et Saintes avec à proximité le port de La Rochelle), ainsi que l'Italie et les Pays-Bas (deux plaques tournantes du commerce ornemental en Europe).

Si l'on considère un nombre de K supérieur à 3, le gain de vraisemblance est plus faible pour atteindre un plateau à partir de $K=7$ (accompagné d'une plus forte variance entre *runs*) (figure 6b), et ceci ne permet pas de démontrer l'existence d'autres événements indépendants d'introduction (ayant pu conduire à des différenciations génétiques plus faibles entre foyers au sein de cette région, par exemple avec des introductions en masse depuis la même zone source). Toutefois, la distance géographique entre les foyers de Saintes et de Creuzier-le-Vieux, détectés la même année, n'apparaît pas compatible avec une dispersion active (qui ne peut expliquer que la propagation dans les départements voisins les années suivantes). Considérer un plus grand nombre de *clusters* semble essentiellement permettre de distinguer différents degrés d'*admixture* entre les trois *clusters* primordiaux mis en évidence avec le critère d'Evanno (les *clusters* pour un nombre de K plus élevé apparaissent dans les populations où l'*admixture* est forte pour $K=3$). Les principales zones d'*admixture* correspondent ainsi à la Région Centre Val de Loire considérée dans sa globalité, et à Nice, le dernier des cinq premiers foyers qui existait déjà en France en 2011. D'ailleurs, les valeurs de F_{st} entre Nice et les autres primo-foyers sont faibles (3,5%) à modérées (environ 9%), à l'exception de Strasbourg, le premier foyer (21%), alors qu'elles sont très élevées entre primo-foyers de *clusters* différents (18 à 28%). Bras et al (données non publiées) ont montré à l'aide de la méthode DIY ABC que les scénarios d'invasion les plus vraisemblables pour certaines des 8 populations européennes étudiées impliquaient des phénomènes « tête de pont » où un premier foyer sert par la suite de source à d'autres transports accidentels dans des régions encore indemnes ou déjà colonisées, avec donc parfois en outre de l'*admixture* (plusieurs foyers existants contribuant à la formation d'un nouveau) (voir par exemple Lombaert et al, 2010). Ceci pourrait être le cas pour le foyer de Nice, mais le sous-échantillonnage dans le sud-est ne permet pas de distinguer les contributions de la dispersion passive (transport accidentel via des plantes d'ornement) ou de la dispersion active (vol du papillon) dans la constitution de ce foyer avec le jeu de données actuel. Une fois décrits et sélectionnés les sites les plus pertinents en Europe, la méthode DIY ABC (limitée par la lourdeur des calculs) pourrait être appliquée également à une partie du jeu de données obtenu au cours de mon stage pour tester des scénarios d'invasion un peu plus complexe que ceux que l'on peut inférer avec cette première approche que j'ai mise en œuvre.

3.2. Structure des populations en Région Centre Val de Loire : une contribution vraisemblable de la dispersion passive via des « têtes de pont »

L'échantillonnage à une échelle fine en Région Centre Val de Loire suggère l'existence de phénomènes « tête de pont » à l'instar de ce qu'ont montré Bras et al (données non publiées) à large échelle en Europe.

Dans l'ensemble, les populations de la Région Centre Val de Loire sont caractérisées par une forte *admixture* entre deux des trois *clusters*, celui d'Ile-de-France et celui représenté au plus près par les « primo-foyers » de Creuzier-le-Vieux au sud, et de Saintes au sud-ouest (tous deux détectés en 2011). Comme pour Nice, les données génétiques actuelles et les analyses conduites ne permettent pas de distinguer à elles seules entre des scénarios faisant intervenir des contributions des différents *clusters* sources via de la dispersion passive ou de la dispersion active, d'autant que ces hypothèses ne sont pas exclusives. Au vu des capacités de vol de la pyrale du buis telles qu'elles ont été mesurées au laboratoire, une propagation de proche en proche et une rencontre des deux *clusters* dans la région pourraient apparaître suffisantes pour rendre compte du patron général observé à grande échelle (surtout plusieurs années après les premières détections). Néanmoins, localement, la structure génétique de certains sites en région Centre Val de Loire n'apparaît pas pouvoir faire intervenir cette seule explication. Par exemple, à Chinon, Valençay ou Issoudun les individus sont génétiquement très proches les uns des autres, sont très apparentés aux individus d'Ile-de-France tout en étant éloignés géographiquement, et sont fortement différenciés des individus des sites environnants (en particulier à Chinon). De la même façon, le patron d'*admixture*, et la forte contribution du *cluster* « Creuzier-le-Vieux / Saintes / Italie / Pays-Bas » à Chartres (qui est un des sites de la région les éloignés des régions où prédomine ce *cluster*) est difficilement explicable avec les données actuelles par la seule dispersion active de l'insecte depuis le sud, que ce soit Saintes ou Creuzier-le-Vieux. La structure génétique des populations en région Centre Val de Loire apparaît être en partie faite de *patches* et ne semble donc pas seulement résulter d'un gradient de « métissage » entre des foyers d'introduction qui lui seraient uniquement extérieurs. L'explication la plus vraisemblable semble ici être la dispersion passive avec plusieurs autres introductions apparentées aux premières (notamment via une ou des sources franciliennes de plantes d'ornement ou via de mêmes arrivages massifs et simultanés en plusieurs points du territoire depuis un *hub* commercial à l'étranger), et non une diffusion de proche en proche par dispersion active à partir de deux sources uniques. Il est à rappeler qu'il s'est vendu en France de 2011 à 2014, près de 2 millions de buis par an, dont une grande partie dans la grande distribution et les jardineries spécialisées (FranceAgrimer, 2012, 2013, 2014, 2015).

Conclusion.

Depuis son signalement en Europe en 2008, plusieurs événements d'introduction ont structuré la diversité génétique des populations de pyrale du buis en France. Les fortes capacités de dispersion active de l'insecte, ainsi que des transports accidentels et des introductions en de multiples points du territoire liés au commerce d'ornement, sont vraisemblablement responsables des patrons observés aussi bien à une échelle régionale que nationale, et peuvent donc rendre compte du caractère fulgurant de l'invasion de cette espèce. Une fois la structure génétique des populations aux échelles française et européenne mieux décrite, il serait intéressant de mettre en œuvre sur un sous-jeu de données d'autres méthodes bayésiennes d'assignations génétiques d'individus et de tests de scénarios d'invasions pour préciser la part de la dispersion passive et le rôle du commerce ornemental.