



Année 2012 – 2013



UNIVERSITE D'ORLEANS  
COLLEGIUM SCIENCES ET TECHNIQUES  
ITP SCIENCES BIOLOGIQUE ET CHIMIE DU VIVANT

---

**Phylogéographie de la processionnaire du pin :**  
*recherche de signatures génétiques de refuges glaciaires  
recouverts par l'Adriatique à l'Holocène*

---

MASTER SCIENCES BIOLOGIES (2<sup>ème</sup> année)

SPECIALITE : BOPE

**Vilajosana David**

Sous la direction de Rousselet Jérôme

Unité zoologique, INRA - Centre Val de Loire, Unité de Recherche de Zoologie Forestière (URZF)  
2163, avenue de la pomme de pin, Ardon, CS 40001, F-45075 Orléans CEDEX 2

**Résumé :** Connaitre la réponse des organismes vivants aux réchauffements climatiques passés présente un intérêt pour prédire les effets des réchauffements globaux contemporains. De nombreuses études phylogéographiques se sont intéressées aux différents patrons de recolonisation post-glaciaire d'espèces terrestres, mais, peu d'études ont traité des variations du niveau marin, qui auraient pu faire disparaître les traces de refuges glaciaires. La reconstitution de l'histoire de certaines espèces pourrait ainsi être brouillée par des contributions insoupçonnées à l'aire actuelle. Au cours de mon stage, j'ai testé l'hypothèse de l'existence de refuges glaciaires majeurs pour la processionnaire du pin au niveau des territoires libérés par la mer Adriatique lors du dernier maximum glaciaire. Le séquençage d'un fragment du gène de la cytochrome c oxidase I a été réalisé sur 155 individus correspondant à 35 échantillons de populations collectés sur tout le pourtour Adriatique. Au niveau des îles et du littoral croates, le marqueur mitochondrial utilisé a mis en évidence des signatures génétiques de populations ayant pu persister dans la région sur plusieurs cycles glaciaires mais ayant probablement subi de fortes fluctuations démographiques. La distribution spatiale de la diversité génétique, et le degré de divergence entre les haplotypes identifiés, sont cohérents avec les résultats attendus sous notre hypothèse.

Mots clés : Balkan, Réchauffements climatiques, Refuge glaciaire, Variation niveau marin

## **ABSTRACT**

Understanding species responses to the Holocene climate warming can help to predict the effects of contemporary climate changes. Numerous phylogeographical studies dealt with post-glacial recolonization patterns but few focused on the impact of the Pleistocene sea level changes. Maybe this event had made disappear old population traces and skewed our interpretation of species history. During this study, I tested the assumption of existence of major refugia for the pine processionary in the territories released by the Adriatic Sea during the last glacial maximum. The sequencing of a fragment of cytochrome C oxidase gene, was carried out on 155 specimen from to 35 population collected around the Adriatic Sea. The spatial distribution of genetic diversity and the degree of divergence between the haplotypes are consistent with the results expected under our assumption. In the Croatian islands and at long the coast, the mitochondrial marker used thus revealed population that have probable persist in this area over several glacial cycles but having probably undergone strong demographic fluctuations.

Keywords: Balkan, Global warming, Ice age refugia, Sea level variation

## **I. INTRODUCTION**

Le changement climatique annoncé pour le siècle à venir par les climatologues pose la question de la façon dont les organismes vivants vont y répondre (expansion ou rétraction d'aires, augmentation de la densité de population ou extinctions, modifications de la phénologie, découplage des interactions entre espèces, ...). Dans ce contexte, les études portant sur la réponse des organismes aux réchauffements climatiques passés, en particulier du dernier maximum glaciaire (-21 000 à -18 000 ans) à aujourd'hui, présentent un nouvel intérêt pour prédire les effets des changements globaux contemporains.

Afin de mettre en lumière ces réponses, il est nécessaire de reconstituer, de façon fiable et exhaustive, les aires de répartition occupées par les espèces dans le passé, afin de les comparer aux aires actuellement occupées. Différentes approches permettent d'obtenir des

données sur l'histoire passée des organismes vivants telles que, des données fossiles, palynologiques (chez les végétaux), génétiques ou bien de modélisation d'enveloppe climatique. Chacune de ces approches présentent des biais d'interprétations différents. En effet, peu d'espèces laissent des traces fossiles ou palynologiques utilisables, et lorsque c'est le cas, les enregistrements sont souvent très partiels. La modélisation d'enveloppe climatique repose sur l'hypothèse du conservatisme de niche qui n'est pas forcément vérifiée.

Le recours aux outils de la génétique peut également permettre d'inférer l'histoire d'un organisme. Cependant, cette approche dépend des empreintes génétiques que les populations contemporaines ont pu conserver dans leurs génomes ainsi que des marqueurs disponibles pour accéder à cette information. Il s'agit ainsi d'interpréter les signatures génétiques laissées par les différentes forces évolutives (sélection, mutation, dérive, migration, ...) susceptibles d'avoir modelé la distribution de la variabilité génétique observable de nos jours (Hewitt 1999 ; Excoffier 2009). Cependant, la structuration de cette variabilité contemporaine est la résultante de nombreux processus, en cours d'intervention ou ayant intervenu durant l'histoire d'une espèce. Ces interventions ont pu opérer à différentes échelles de temps et d'espace. Les signatures génétiques présentes dans ces populations sont, ainsi, plus ou moins utilisables pour retracer leur aire de distribution passée et leur histoire démographique. Enfin, trivialement, il n'est généralement pas possible d'échantillonner et de génotyper des populations disparues. Malgré ces limites, la génétique des populations s'est montrée une approche puissante et féconde, en particulier chez des organismes qui ne laissent aucune trace fossile (comme de nombreux insectes). Elle peut en outre, fournir des données confrontables aux prédictions de modèles bioclimatiques passés permettant ainsi de tester la congruence de ces deux approches différentes. En effet, la confrontation des différentes approches permet de minimiser les biais de chacune.

La phylogéographie est une discipline de la biologie et de la génétique des populations étudiant les principes et processus qui gouvernent la distribution géographique des lignées généalogiques, spécialement celle de niveau intraspécifique (Avice 2000). Elle étudie l'évolution génétique et démographique ayant conduit à la distribution et à la structuration actuelle des populations, sous l'effet, par exemple de la dérive des continents, de l'orogénèse, des modifications de courants océaniques ou bien des oscillations climatiques (Hewitt 1999). Les premières études phylogéographiques se sont surtout intéressées à la réponse d'organismes continentaux rencontrés à des latitudes tempérées (Avice et al. 2000, Taberlet et al. 1998, Hewitt 1999).

Nous savons que durant le Quaternaire (de 2,5 millions d'années à aujourd'hui), le climat de la Terre a répondu à des variations d'ordre astronomique (cycles de Milankovitch) par une alternance de période de glaciation et de réchauffement. En effet les terres émergées de l'hémisphère nord ont connu des phases d'extension et de régression de l'inlandsis (glaciers polaires) dans leurs latitudes élevées, induisant de profondes modifications des habitats dans les zones tempérées actuelles. Les conditions de température et/ou d'habitat (disparition de leur niche écologique) étant devenues défavorables à la survie d'un grand nombre d'espèces, la persistance de leur population n'a été possible que dans la partie sud de l'Europe. Etant des événements récents (à l'échelle des temps géologiques), ces cycles ont laissé de fortes empreintes au niveau intraspécifique. Ils ont eu un impact majeur sur la démographie et la distribution des organismes thermophiles. Les cycles d'expansion/contraction des aires de distribution qui en ont résultés ont fortement contribué à modeler la distribution spatiale de leur variabilité génétique. Malgré la diversité des réponses de chaque organisme (insectes, amphibiens, mammifères et végétaux), un petit nombre de patrons phylogéographiques présentant des caractéristiques partagés (paradigmes phylogéographiques) a pu être défini. La plus générale de ces caractéristiques a été décrite sous la dénomination de « *southern richness and northern purity pattern* » (Hewitt 1999).

Cette signature génétique, commune à de nombreuses espèces tempérées, est que les populations contemporaines rencontrées dans le sud de l'Europe possèdent une grande diversité génétique alors que les populations du nord de l'aire en présentent une plus faible. En effet, lors des glaciations, les populations méridionales se sont maintenues. De formation ancienne, elles ont eu le temps d'accumuler de la diversité génétique par mutation. De tailles importantes et relativement stables démographiquement au cours du temps, elles ont pu la conserver sans la perdre par dérive génétique. Les populations septentrionales, ayant disparu durant la période glaciaire, sont, quant à elles, de formation récente et sont donc généralement moins variables. Le patron de diversité génétique de chaque espèce dépend fortement du processus de dispersion à l'œuvre lors de la recolonisation post-glaciaire, et en particulier des goulots d'étranglement, réduisant la diversité génétique, qui peuvent se produire lors de la fondation de nouvelles populations. Leur structuration dans l'espace dépend également des barrières naturelles rencontrées lors de la recolonisation (Hewitt 2004a, Schmitt 2007).

En Europe, les trois grandes péninsules méridionales (ibérique, italienne et balkanique) ont ainsi été identifiées comme des zones refuges majeures pour les espèces tempérées (Taberlet et al. 1998, Hewitt 1999). Leur contribution à la recolonisation

postglaciaire du nord a souvent été modelée par de grandes chaînes montagneuses comme les Alpes ou les Pyrénées. Ces dernières pouvant jouer un rôle de barrière pour de nombreuses espèces.

Toutefois, les données accumulées ces dernières années montrent que l'histoire de toutes les espèces ne peut pas être réduite à ce seul patron et aux quelques paradigmes initialement énoncés (Hewitt 2000, Schmitt 2007). De nombreux organismes et de nombreux biomes/biota, présentant des réponses très différentes des zones tempérées actuelles, ont été sous-étudiés (Hewitt 2004b). Par exemple, les espèces arctiques/alpines ont connu, à l'inverse des thermophiles, des phases d'expansion lors des périodes froides et de contraction lors des périodes chaudes, parfois accompagnées de processus de migration verticale le long des versants des montagnes (e.g. DeChaine & Martin 2005, Varga & Schmitt 2008, Rousselet et al. 2010). Nous savons également maintenant, que les paléoenvironnements étaient plus complexes que pensés auparavant (dans le temps et dans l'espace en fonction, par exemple, de la proximité des océans et du relief) (e.g. Barton et al 2002, Canestrelli et al. 2006). Des zones refuges cryptiques ont ainsi été mises en évidence à des latitudes plus septentrionales qu'attendues (e.g. Bhagwat et al. 2008, Stewart 2001).

Il existe pourtant un aspect qui n'a probablement pas été suffisamment considéré. En effet, nous tenons généralement pour acquis que la géographie de l'Europe est restée relativement stable tout au long du quaternaire, du fait de l'absence de modification majeure du relief et des bassins versants à cette échelle de temps. Néanmoins, une modification majeure n'est généralement pas prise en considération. Celle-ci concerne le tracé des côtes dû aux variations du niveau de la mer. Ces dernières pourraient avoir effacé les traces de certaines zones refuges et donc brouillé l'histoire de certaines espèces continentales, tout du moins, si ces refuges ont eu une contribution non négligeable à la recolonisation post-glaciaire. En effet, lors du dernier maximum glaciaire, la baisse du niveau marin de 50 à 150 m a pu découvrir, dans certaines régions, de vastes territoires pouvant être colonisés, notamment par des espèces thermophiles pouvant bénéficier de l'effet océanique. Ces territoires, ayant été reconquis par la mer à l'holocène, se pose la question de la possibilité d'en détecter l'existence et la contribution à la recolonisation post-glaciaire. C'est particulièrement le cas pour le nord de l'Adriatique où les territoires libérés ont pu constituer un pont terrestre entre la péninsule italienne et balkanique (Lucchini & Randi 1998, Lambeck et al 2002). Les territoires libérés par la mer dans ces régions montagneuses (Appennins et Alpes dinariques) pouvaient présenter l'avantage d'être à plus basse altitude que les territoires

environnants ainsi que plus proches de la mer et de son influence climatique. Pour des espèces méditerranéennes ou subméditerranéennes susceptibles d'avoir colonisé et prospéré dans ces territoires, les îles et le littoral croates pourraient être occupés par des populations reliques. Médail & Diadéma (2009) ont montré qu'un grand nombre de plantes méditerranéennes ont probablement persisté en plusieurs zones proches du littoral et ont défini 4 zones de hot spots de diversité au niveau des Balkans. Les auteurs ne sont, cependant, pas intéressés aux îles. Celles-ci pourraient présenter des signatures génétiques encore détectables de populations ancestrales ayant colonisé durant l'ère glaciaire un territoire beaucoup plus étendu. C'est cette hypothèse que l'on va tester avec la processionnaire du pin.

En effet, La processionnaire du pin d'hiver, *Thaumetopoea pityocampa* (Lepidoptera: Notodontidae), est un insecte faisant partie d'un complexe d'espèces occupant le pourtour du bassin méditerranéen (Kerdelhué et al. 2009). Ces chenilles, défoliatrices et urticantes, sont capables de s'attaquer à tous les pins et cèdres natifs de son aire de distribution. Toutefois, le pin noir constitue son hôte préféré. Ce pin-hôte se rencontre généralement dans des zones méso à supra-méditerranéennes qui pourraient correspondre à l'optimum climatique de la processionnaire (hiver et été doux) et semble avoir été présent dans la plupart de ses refuges glaciaires déjà identifiés (Rousselet et al. 2010, Simonato et al. 2007).

Au niveau du pourtour de la mer Adriatique, la processionnaire du pin est présente sur le littoral et dans de nombreuses îles sur pin d'Alep et sur pin noir (dont une variété endémique des îles croates, *Pinus nigra nigra dalmatica*). Pour des raisons climatiques, elle n'est pas capable de coloniser, à l'intérieur des terres, les forêts de pins noirs et de pins sylvestres des Alpes dinariques.

La processionnaire du pin est une espèce très fortement structurée génétiquement (Simonato et al. 2007, Kerdelhué et al. 2009, Rousselet et al. 2010). Les femelles présentent un comportement philopatric et ne vivent généralement qu'une seule nuit. Elles ne semblent pas capables de disperser à plus de quelques kilomètres (Robinet et al, 2012). En cas d'existence de refuges glaciaires « adriatiques », on peut donc envisager que des lignées maternelles aient pu être « capturées » dans les îles croates lors de la remontée des eaux. Celles-ci, restées plus ou moins isolées, pourraient alors être détectées, et leur histoire retracée, à l'aide de marqueurs mitochondriaux si le niveau de variabilité du marqueur utilisé est suffisant à l'échelle de la zone d'étude.

Afin de tester cette hypothèse nous analyserons la distribution de la diversité génétique de *Thaumetopoea pityocampa* à partir du séquençage de l'ADN mitochondrial de la cytochrome oxydase I. Les objectifs de l'étude sont de (1) décrire la phylogéographie de l'espèce dans la partie orientale de son aire de répartition (Balkan) afin de compléter ce qui a déjà été fait dans la partie plus occidentale, ainsi que (2) de tester l'hypothèse d'une zone refuge glaciaire « adriatique » qui se trouve actuellement immergée.

Pour ces populations reliques, ayant persisté sur une longue période de temps, on peut s'attendre à trouver des mitotypes présentant une forte divergence les uns par rapport aux autres. Cependant, les fortes fluctuations démographiques qu'auraient connues ces populations au cours des phases de contraction-expansion se traduiraient par : (I) l'existence de nombreux haplotypes manquants lors de la construction d'un réseau d'haplotypes, (II) une diversité génétique faible sous l'effet de la dérive génétique dans les populations italiennes de taille réduite, isolées du continent et (III) une conservation d'une partie non négligeable de la diversité génétique ancestrale, non pas localement, mais sur l'ensemble des îles.

## **II. MATERIEL ET METHODES**

### ***2.1 Echantillonnage***

Parmi les 1020 échantillons collectés sur 81 sites entre 1999 et 2013 dans les différents états du pourtour de la mer Adriatique, nous avons sélectionné 296 individus issus de 35 sites de collecte afin de couvrir l'ensemble de la région et les deux principales essences hôtes présentes (pin noir et pin d'alep) (Tableau I). Au début de l'étude, 41 séquences étaient déjà disponibles (Rousselet et al, 2010) et 255 ont subi le protocole en entier. Pour l'échantillonnage, un effort plus prononcé a été effectué au niveau du littoral et des îles croates afin de tester l'hypothèse du refuge glaciaire adriatique, puisque ces zones seraient susceptibles de correspondre aux zones refuges attendues. De plus, les échantillons ont été sélectionnés, en particulier dans les Balkans, dans les environs de chaque zone de hot spot végétale (décrit par Médail et Diadema 2009). On soulignera néanmoins que, par manque de temps, la côte italienne n'a pas pu être étudiée autant que nous le souhaitions. Les échantillons ont été collectés au stade chenille. Un seul individu par nid a été analysé afin d'éviter de travailler sur des individus apparentés, biaisant l'estimation de la diversité à l'intérieur des populations.

**Tableau I.** Caractéristiques des 35 populations étudiées. Les zone (A=Adriatique Nord-est ; B=Dalmatie septentrionale ; C=Dalmatie central et méridionale ; D=Monténégro ; E=Ligurie, F=Italie ; G=Grèce), les positions par rapport à la mer, les pins hôtes (PN=Pinus nigra ; PA=Pinus halpensis ; PS=Pinus sylvestris), pays, dates de collectes, latitudes, longitudes et nombre d'échantillons séquencés sont représentés.

Code	Régions de collecte	Zone	Position	Pin hôte	Pays	Date	Latitude	Longitude	Nombre
1	Calbarina	F	continent	PN	Italie	1999	45°15'52.28"N	11°43'9.87"E	5
2	Bari	F	continent	PA	Italie	2004	41° 7'7.70"N	16°49'17.44"E	5
3	San Michele	F	continent	PN	Italie	1999	43°30'51.07"N	11°23'59.03"E	5
4	Massinino	F	continent	PS	Italie	1999	44°24'47.59"N	8°26'0.36"E	5
5	Rollo	E	littoral	PA	Italie	1999	43°56'28.91"N	8° 7'29.36"E	5
6	Germagno	F	continent	PN	Italie	1999	45°15'43.21"N	7°28'0.69"E	5
7	Ruines veres	F	continent	PS	Italie	1999	45°38'41.70"N	7°40'36.60"E	8
8	Monfalcone	A	littoral	PN	Italie	2011	45°47'57.44"N	13°34'42.46"E	10
9	Buje	A	littoral	PA	Croatie	2008	45°27'23.80"N	13°37'32.41"E	17
10a	Obrovac	B	continent	PN	Croatie	2011	44°11'45.62"N	15°39'6.37"E	5
10b	Gornji Karin	B	continent	PA	Croatie	2011	44° 7'51.04"N	15°38'6.99"E	3
10c	Benkovac	B	continent	PN	Croatie	2011	44° 4'40.68"N	15°35'49.84"E	10
11	Losinj	A	ile	PA	Croatie	2011	44°34'42.83"N	14°24'32.70"E	20
12	Seget	C	littoral	PA	Croatie	2011	43°30'46.27"N	16°11'39.36"E	9
13a	Pisak	C	littoral	PA	Croatie	2011	43°24'2.22"N	16°52'9.88"E	3
13b	Gornja Brela	C	continent	PN	Croatie	2011	43°24'15.39"N	16°54'18.98"E	4
14	Tucepi	C	littoral	PA	Croatie	2011	43°16'33.38"N	17° 2'49.06"E	10
15	Brac	C	ile	PN	Croatie	2011	43°19'5.19"N	16°35'44.34"E	10
16a	Hvar Ouest	C	ile	PA	Croatie	2011	43° 9'39.39"N	16°29'43.42"E	8
16b	Hvar Centre	C	ile	PA	Croatie	2011	43° 8'22.66"N	16°43'23.29"E	8
16c	Hvar Est	C	ile	PA	Croatie	2011	43° 7'58.53"N	17° 2'42.40"E	9
17a	Korcula Nord	C	ile	PA	Croatie	2011	42°57'3.62"N	16°50'3.83"E	4
17b	Korcula Centre	C	ile	PA	Croatie	2011	42°56'44.32"N	16°58'46.44"E	4
18	Peljesec Nord	C	ile	PN	Croatie	2011	42°57'59.17"N	17°17'3.72"E	6
19a	Igrane	C	littoral	PA	Croatie	2011	43°11'37.19"N	17° 9'27.26"E	2
19b	Drvenik	C	littoral	PA	Croatie	2011	43° 9'11.77"N	17°14'59.02"E	7
19c	Gradac	C	littoral	PA	Croatie	2011	43° 6'48.45"N	17°20'4.12"E	1
20	Dugi	B	ile	PA	Croatie	2011	43°55'59.48"N	15°10'0.11"E	8
21a	Pašman	B	ile	PA	Croatie	2011	44° 0'43.33"N	15°15'39.88"E	4
21b	Ugljan	B	ile	PA	Croatie	2011	44° 8'10.14"N	15° 5'0.60"E	4
22	Zadar	B	littoral	PA	Croatie	2011	44° 8'19.91"N	15°13'38.04"E	9
23	Budva	D	continent	PN	Montenegro	2011	42°18'10.90"N	18°55'0.15"E	6
24	Cemovsko	D	continent	PN	Montenegro	2009	42°27'6.99"N	19°17'54.72"E	5
25	Grece	G	continent	PN	Grece	2003	37°58'8.62"N	22°10'41.46"E	4
26	Tremeti	F	ile	PA	Italie	2013	42° 7'5.47"N	15°29'27.33"E	8



## ***2.2 Le séquençage du marqueur moléculaire COI. (Premier protocole utilisé)***

Le gène de la sous-unité I de la cytochrome c oxydase (COI), est un gène mitochondrial à hérédité strictement maternelle. Cette particularité nous permet de retracer l'histoire des lignées maternelles, qu'on attend beaucoup plus structurées, du fait de l'existence d'une dispersion sexe biaisée (Salvato et al 2002, Simmonato et al 2007, Rousselet et al 2010). Les mâles avec des capacités de vol d'environ 50 km, contre seulement 5 km pour les femelles, induisent un flux de gènes au niveau nucléaire à plus longues distances. Par contre, les mâles ne sont pas censés transmettre leur mitochondrie (phénomène rarement reporté chez les animaux, en dehors des mollusques).

Nous avons utilisé le même fragment que celui qui a permis de caractériser les principales zones de refuges et de recolonisation post-glaciaire de la partie ouest de l'aire (seconde partie du gène plus variable que la première, correspondant au fragment dit « barcode »)

### ***2.2.1. Extraction de l'ADN***

Les chenilles de *Thaumetopoea pityocampa* ont été disséquées afin de récolter un tube de Malpighi et/ou un morceau de tube digestif (2.5mg). L'extraction d'ADN génomique (ADN mitochondrial + ADN nucléaire) a été effectuée à partir du broyage de ces prélèvements à l'aide du kit « NucleoSpin® Tissue XS » (Macherey-Nagel). Ce kit utilise la protéinase K qui permet la lyse des cellules. L'ADN a ensuite, été isolé grâce à une membrane de gel de silice disposée sur une colonne permettant de filtrer l'ADN par des interactions électrostatiques (d'abord capturé puis relâché et remis en solution dans un éluant). Une fois l'ADN extrait, ce dernier est conservé à -20°C.

### ***2.2.2. Amplification de l'ADN***

L'amplification du fragment du gène COI a été obtenu par *polymerase chain reaction* (PCR). C'est une méthode permettant de réaliser un grand nombre de copies d'un fragment d'ADN ou d'ARN cible en utilisant deux amorces (séquence de nucléotides permettant de débiter les duplications). Dans le cas présent, les amorces utilisées sont les amorces Jerry (5'-CAACATTTATTTTGATTTTTTGG-3') et Pat (5'-TAATATGGCAGATTAGTGCATTGGA-3'), elles permettent d'amplifier un fragment de 878 pb du gène mitochondrial COI (Rousselet et al 2010).

Pour chaque individu extrait, 1ul d'ADN a été mélangé à 24 ul d'une solution (« mix ») contenant une solution tampon, une solution de MgCl<sub>2</sub> (cofacteur de la Taq polymérase), une solution de dNTP (bases azotés permettant la création de nouveau brin d'ADN), les deux types d'amorce (Jerry-Pat), une solution de Betaine (permet de stabilisé l'ADN) et la Taq polymérase. Le protocole PCR est décrit en détail dans l'article Rousselet et al 2010. Il consiste en une amplification de 30 cycles (dénaturation : 95.0°C 0:30 mn, hybridation : 48.0°C 0:45 mn et élongation : 72.0°C 1:00 min).

La qualité des PCR est ensuite vérifiée par migration sur gel d'agarose 1%, puis révélée au BET par lecture sur table UV. Si la qualité du produit de PCR (quantité d'ADN amplifiée et donc intensité des bandes sur le gel) n'était pas assez bonne, une seconde PCR a été effectuée puis regroupée avec la précédente afin d'avoir plus de matériel pour la suite du protocole.

### ***2.2.3. Purification de l'ADN***

Une fois l'amplification des fragments d'intérêt terminée et la qualité suffisante, les solutions ont subi une étape de purification qui a permis de retirer toutes molécules ajoutées non utilisées présentes et de ne garder que l'ADN. Afin d'effectuer cette purification, le kit « PCR clean-up gel extraction » a été utilisé. Pour cela, une colonne, munie d'une membrane, a été utilisé afin de retenir l'ADN. Celle-ci sera ensuite lavée afin d'en retirer tous débris. Puis l'ADN est élué et conservé a -20°C.

Comme pour la PCR, la qualité des solutions purifiées obtenues a été testée à l'aide d'un gel d'agarose 1%.

### ***2.2.4. Réaction de séquence, précipitation et séquençage***

Les échantillons ont ensuite subis une dernière amplification, la réaction de séquence, élaborée grâce à la méthode de Sanger. Chaque échantillon subit deux réactions de séquence (une pour chacune des amorces). Les réactions de séquences ont été effectuées, avec les mêmes amorces que lors de la PCR, à l'aide du Abi Prism® Big Dye® Terminator V.3.1 Cycle Sequencing Kits (Applied Biosystems).

Une fois la réaction de séquence terminée, les échantillons ont été purifiés par précipitation à l'alcool. Cette dernière a permis d'éliminer tous les sels et réactifs non utilisés,

par mélange de la solution avec de l'éthanol, qui ont ensuite été retiré à la pipette. Le culot a été conservé à -20°C jusqu'à utilisation.

Les fragments ont, ensuite, été remis en suspension dans 30µl d'eau, transférés sur une plaque et passées dans le séquenceur Abi Prism® 3500 Genetic Analyser (Applied Biosystems).

### ***2.2.5. Analyses des séquences***

Les séquences ainsi obtenues ont été corrigées à l'aide du logiciel CodonCode Aligner 3.7.1 (CodonCode Corporation, Dedham MA) puis alignées avec le logiciel BioEdit 7.05 (Hall 1999) avant d'être analysées. Ces séquences ont été tronquées (au début et à la fin) afin d'obtenir une partie commune à toutes les séquences, de 802 pb. Ce fragment est identique à celui utilisé lors de l'étude sur la partie ouest de l'air de répartition (Rousselet et al 2010), permettant ainsi la comparaison de la diversité et de la divergence dans les deux parties de l'aire.

### ***2.3 Séquences parasites et nouveaux protocoles testés et utilisés***

Les premières séquences ayant révélé des amplifications parasites conduisant à ne pas pouvoir en exploiter une grande proportion, un second protocole a été utilisé. Les chromatogrammes de nombreuses séquences ont montré à certaines positions des doubles pics (Annexe 1) conduisant à des ambiguïtés rendant la détermination de l'haplotype incertaine (séquencées non conservées).

#### ***2.3.1. Utilisation de nouvelles amorces***

Plusieurs couples d'amorces ont été testés afin d'essayer de résoudre ce problème. Certaines amorces étaient préexistantes comme LCO et d'autres ont été créées afin de permettre d'éliminer les pseudo-gènes des séquences d'intérêts tels que Jerry+65pb (5'-TAACAGATCGTAACTTAAATAC-3') ; Pat+174pb (5'-TCTTAATATTAATATCC-3') et Pat+367pb (5'-CATAACTTCAATATCATTGATGT-3').

Des tests ont été effectués avec différents couples d'amorces : Jerry(standard)-LCO, Jerry+65pb-Pat+174pb et Jerry+65pb-Pat+367pb. Dans l'ensemble du protocole, seules les amorces (LCO, Jerry+65pb, Pat+174pb et Pat+367pb) ont variées par rapport au protocole précédent (Jerry-Pat standard).

### **2.3.2. Utilisation du kit REPLI-g**

Les doubles pics, traduisant la présence de deux séquences en mélange dans le produit de PCR, peuvent être liés soit à une hétéroplasmie soit à des pseudogènes. L'hétéroplasmie est la présence de plus d'un mitotype au sein d'une même cellule ou d'un individu. Cette dernière apparaît par mutation (Goto et al 2011) ou par transmission mitochondriale paternelle (Wolff et al 2013). Elle est considérée comme très rare chez les animaux. Les pseudogènes correspondent, quant à eux, à des copies non codantes de fragments de génome mitochondrial intégrés dans le génome nucléaire. Ils sont fréquents chez certains groupes d'insectes comme les hyménoptères. Afin d'éliminer les amplifications parasites pouvant être liées à des pseudogènes, nous avons testé pour la première fois chez la processionnaire, un kit, « REPLI-g Mitochondrial DNA » (Qiagen), permettant d'amplifier spécifiquement l'ADN mitochondrial. En effet, il permet d'amplifier sélectivement l'ADN circulaire, caractéristique permettant de différencier l'ADN mitochondrial de l'ADN nucléaire, tout deux contenus dans l'extrait d'ADN génomique, à partir de leur origine de répllication. Les échantillons d'ADN ont été dénaturés en présence d'un tampon (REPLI-g mt Reaction Buffer). La dénaturation fut stoppée puis une polymérase fut ajoutée dans les solutions (REPLI-g Midi Polymerase) afin de permettre l'amplification durant huit heures. Cette étape supplémentaire dans le protocole s'insère entre les étapes 2.2.1 et 2.2.2 précédemment décrites.

## **2.4 Analyses des données**

### **2.4.1 Distribution et généalogies des mitotypes.**

Le nombre d'haplotypes et leur fréquence au sein de chaque population ont été calculés avec le logiciel DnaSP version 4.5 (Rozas et al 2003).

Un réseau d'haplotype a été construit grâce au logiciel TCS1.21 (Clement et al 2000). Ce programme permet d'estimer le réseau d'haplotypes le plus parcimonieux afin de mettre en évidence une structuration hiérarchique des relations de parenté entre ceux-ci.

Les distances génétiques intraspécifiques ont été calculées à l'aide du logiciel MEGA 5.1 basé sur la méthode phénétique. Cette méthode étudie la parenté entre les taxons en s'intéressant à leur degré de similarité et en se basant sur des distances génétiques calculées par paire de taxons. De plus un arbre phylogénétique en *neighbor-joining* a été produit grâce au même programme et *Thaumetopoea wilkinsoni* a été utilisé comme groupe externe.

### ***2.4.2 Diversité et structure des populations***

La variabilité haplotypique (h) de chaque population ainsi que le nombre moyen de différences nucléotidiques des séquences (variabilité nucléotidiques) d'une population (k) ont été calculés à l'aide du logiciel DnaSP version 4.5.

Plusieurs AMOVA ont été effectuées afin de faire apparaître la structure de la zone étudiée à l'aide du logiciel Arlequin version 3.1 (Excoffier et al 2005). Trois groupements différents ont été faits. Un premier groupement pour de tester l'existence d'une structuration géographique (suivant les zones définis dans le tableau I). Un second groupe afin de tester l'existence d'une structuration spatiale entre îles et littoral d'un côté et intérieur des terres de l'autre. Un dernier groupe pour tester l'existence d'un effet hôte.

## **III. RESULTATS**

### ***3.1 Problèmes techniques et résultats méthodologiques***

#### ***3.1.1 Premier protocole utilisé***

Sur les 81 individus séquencés en suivant le premier protocole, seulement 37 séquences ont été retenues. La plus grande partie des séquences non retenues présentaient des ambiguïtés qui n'étaient pas corrigeables.

#### ***3.1.2 Nouvelles amorces***

Le dessin de nouvelles amorces (Jerry+65, Pat+174, Pat+367) et l'utilisation d'amorce préexistante (LCO) n'ont pas permis d'empêcher des amplifications parasites. La présence de doubles pics a persisté.

#### ***3.1.3 Repli-G***

Sur les 174 individus ayant suivi le protocole avec l'étape Repli-G, 77 séquences sont exploitables (sans présence de double pics), 22 séquences ont continué à présenter quelques doubles pics conduisant à des ambiguïtés impossibles à lever ou n'ont simplement pas marché

et 75 séquences se trouvent être contaminées par l'apparition d'autres séquences ne correspondant pas à une séquence du gène COI, tout au moins, pas de la bonne espèce. Dans ce cas de figure, les ambiguïtés ne se limitent pas à quelques sites mais sont présentes, en très grand nombre, sur toute la séquence.

### ***3.1.4 Les séquences parasites***

Les séquences correspondantes à des copies de type COI (hétéroplasmie ou pseudogène) n'étaient pas uniformément réparties dans le jeu de données. En effet, leur fréquence était plus élevée dans des régions correspondant à des zones de diversité mitochondriale avérée ou bien à proximité. (Annexe 2)

## ***3.2 Résultats scientifiques***

### ***3.2.1. Distribution et généalogies des mitotypes***

L'alignement final correspondant aux séquences exploitables (sans ambiguïté) contenait 155 individus. Trente sites polymorphes ont pu être détectés (Annexe 3) et 13 haplotypes ont été identifiés dont 7 jusqu'ici inconnus (HTA, HTB, HTC, HTE, HTF, HTG, HTH) (Tableau II).

On peut constater que le réseau d'haplotype est fortement divergent et complexe (Figure 1A). En effet, plusieurs haplotypes forment des liens complexes (liaison formant des boucles avec d'autres haplotypes). De plus, il y a souvent plus d'un pas mutationnel qui sépare les haplotypes. Le nombre de pas mutationnels autour de l'haplotype majoritaire HT1 va de 1 à 15 pas. Trois haplotypes (HT12, HTC et HTF) sont différents de HT1 par 1 pas mutationnel, quatre (HT2, HTA, HTG et HTH) par 2 pas mutationnels et trois (HT3, HT4 et HTE) par 3 pas mutationnels. Les deux haplotypes les plus divergents de HT1 sont HT17 et HTB, respectivement à 4 et 15 pas mutationnels. De plus, leurs distances avec les autres haplotypes sont de 0,4 à 0,8 % pour HT17 et de 2 à 2,4% pour HTB (Tableau III). Le reste des haplotypes ont une distance entre eux allant de 0.1 à 0.8 %. Une représentation sous forme d'arbre a été effectuée (ANNEXE 4).

**Tableau II.** Haplotypes observés pour chaque population échantillonnée. Le nombre de séquences exploitables (NB seq), le nombre d'haplotypes identifiés (NB HT), le nombre de chaque haplotype observé, la variabilité haplotypique (h) et la variabilité nucléotidiques (k) sont représentés.

Zone de collecte	NB seq	NB HT	Haplotype (nombres)	h	k
Calbarina	5	1	HT1(5)	0	0
Bari	5	1	HT1(5)	0	0
San Michele	4	1	HT1(4)	0	0
Massinino	5	1	HT1(5)	0	0
Rollo	5	3	HT2(2)-HT3(1)-HT4(2)	0,8	3,40048
Germagno	5	1	HT1(5)	0	0
Ruines veres	8	1	HT1(8)	0	0
Monfalcone	5	2	HTA(4)-HT17(1)	0,4	1,604
Buje	4	2	HTA(2)-HT12(1)	0,4	0,401
Obrovac	4	1	HT1(4)	0	0
Gornji Karin	3	1	HT1(3)	0	0
Benkovac	5	2	HT1(4)-HT12(1)	0,4	0,401
Losinj	5	3	HTA-HT17-HTB	0,8	10,60244
Seget	5	1	HTF(5)	0	0
Pisak	3	1	HT1(3)	0	0
Gornja Brela	4	1	HT1(4)	0	0
Tucepi	7	1	HT1(7)	0	0
Brac	8	3	HT1(1)-HTF(6)-HTH(1)	0,464	0,49724
Hvar Ouest	3	1	HTF(3)	0	0
Hvar Centre	2	1	HTF(2)	0	0
Hvar Est	2	1	HTF(2)	0	0
Korcula Nord-Ouest	4	1	HTF(4)	0	0
Korcula Centre	4	1	HTF(4)	0	0
Peljesec Nord	5	2	HTF(4)-HT1(1)	0,4	0,401
Igrane	2	1	HTG(2)	0	0
Drvenik	4	2	HTG(3)-HTF(1)	0,5	0,49724
Gradac	1	1	HTF(1)	0	0
Dugi	7	1	HT1(7)	0	0
Pašman	2	1	HT1(2)	0	0
Ugljan	4	1	HT1(4)	0	0
Zadar	5	1	HT1(5)	0	0
Budva	5	1	HTC(5)	0	0
Cemovsko	5	1	HTE(5)	0	0
Grece	4	1	HT1(4)	0	0
Tremeti	5	1	HT1(5)	0	0

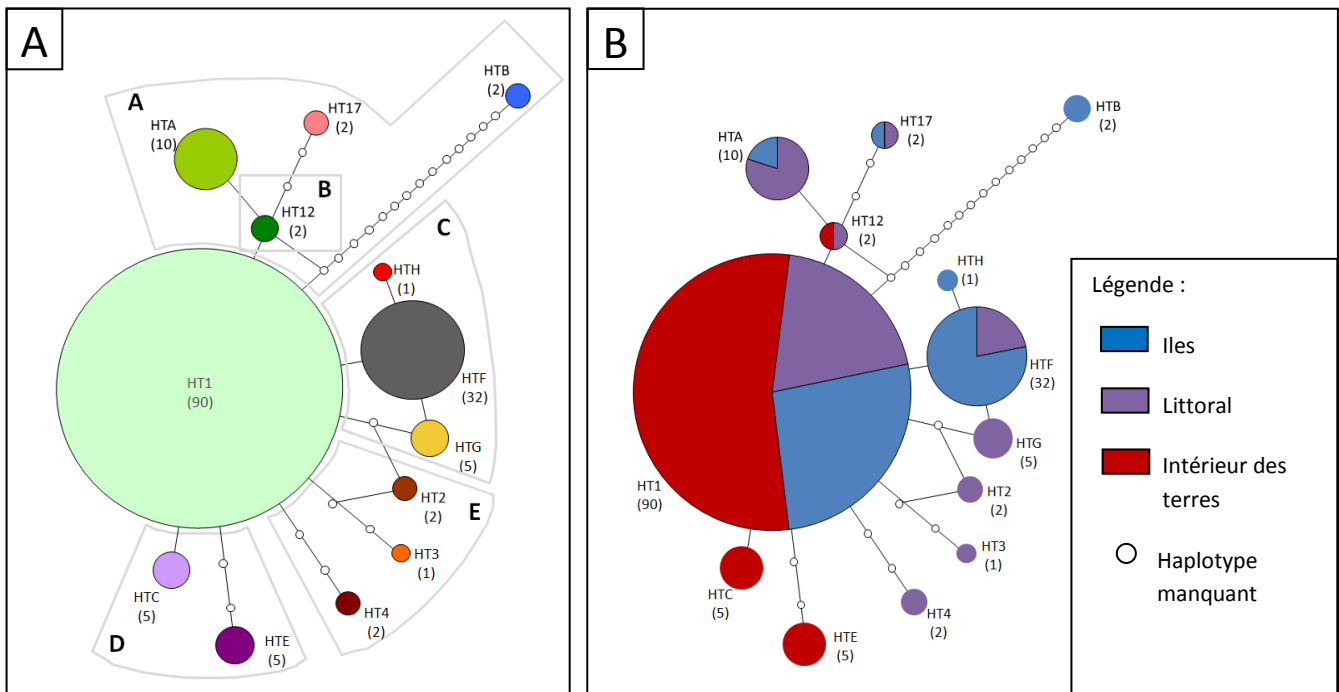
**Tableau III.** Matrice de distance entre les différents haplotypes (0 signifiant que les séquences sont identiques et 1 que les séquences n'ont aucune base en commun). Les valeurs supérieures à 0.01 sont en rouge, celles comprises entre 0.006 et 0.009 sont en noir et les valeurs strictement inférieures à 0.006 sont en gris.

	HT1	HT2	HT3	HT4	HT12	HT17	HTA	HTA	HTC	HTE	HTG	HTH
HT1												
HT2	0,002											
HT3	0,004	0,004										
HT4	0,004	0,006	0,005									
HT12	0,001	0,004	0,005	0,005								
HT17	0,005	0,008	0,009	0,009	0,004							
HTA	0,002	0,005	0,006	0,006	0,001	0,005						
HTB	0,020	0,023	0,024	0,024	0,019	0,022	0,020					
HTC	0,001	0,004	0,005	0,005	0,002	0,006	0,004	0,022				
HTE	0,004	0,006	0,008	0,008	0,005	0,009	0,006	0,022	0,005			
HTF	0,001	0,004	0,005	0,005	0,002	0,006	0,004	0,022	0,002	0,005		
HTG	0,002	0,002	0,006	0,006	0,004	0,008	0,005	0,023	0,004	0,006	0,001	
HTH	0,002	0,005	0,006	0,006	0,004	0,008	0,005	0,023	0,004	0,006	0,001	0,002

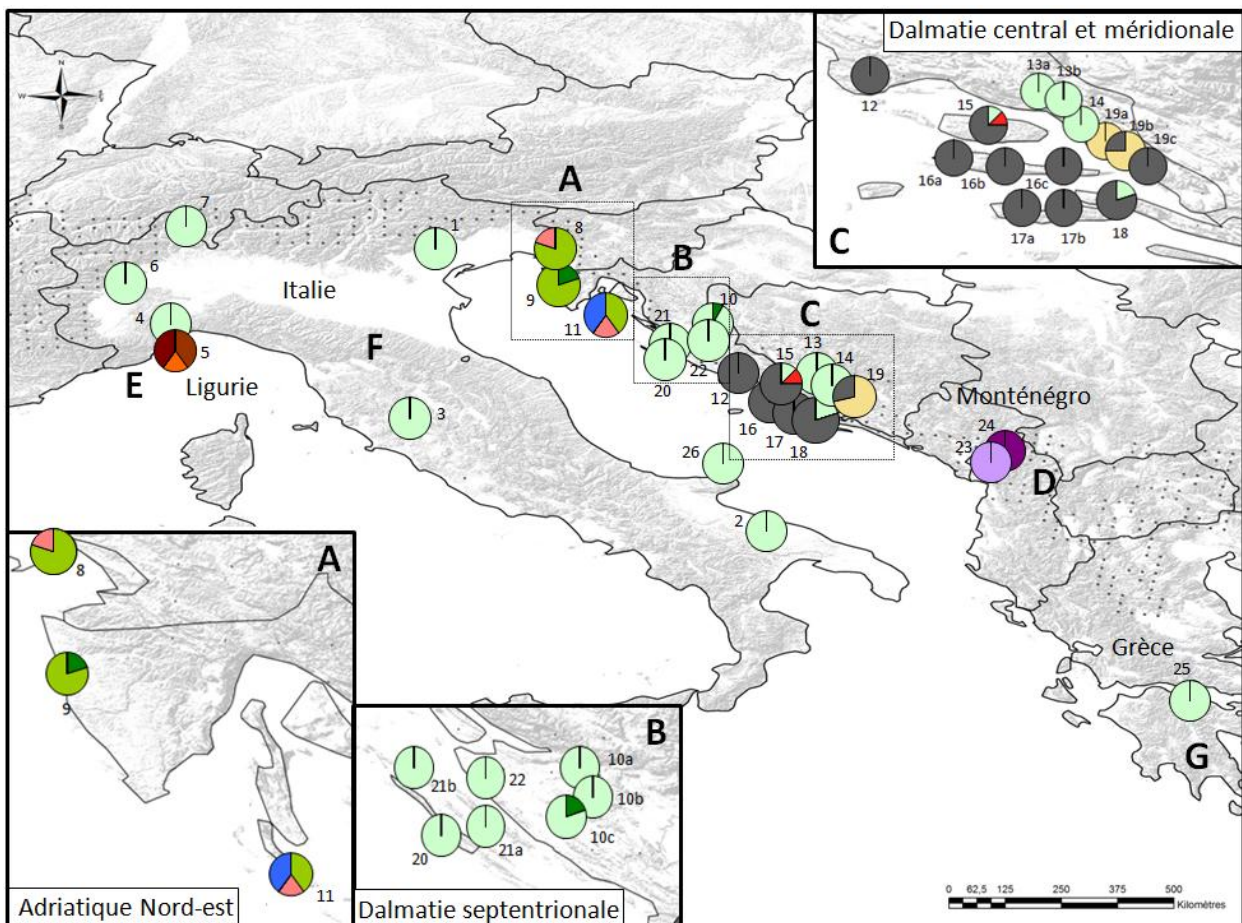
De plus, on observe dans la Figure 1A, un grand nombre d'haplotype manquant. En effet, chacun des pas mutationnels représente un haplotype manquant. Deux haplotypes (HT2 et HTG) sont séparés de l'haplotype majoritaire HT1 par 1 haplotype manquant, 4 (HT3, HT4, HT17 et HTE) sont séparés de HT1 par 2 haplotypes manquants et HTB est séparé de HT1 par 14 haplotype manquant.

Sur le second réseau d'haplotype (Figure 1B), on constate que 2 haplotype (HTB et HTH) ne se trouvent que dans les îles et que 9 haplotype (HT2, HT3, HT4, HT17, HTA, HTB, HTF, HTG et HTH) ne se trouvent qu'au niveau des îles et du littoral. A l'inverse, HTC et HTE ne se trouvent qu'à l'intérieur des terres. HT1 et HT12 se trouvent en proportion presque égale à l'intérieur des terres et au niveau des îles/littoral.





**Figure 1.** Le graphique A représente le réseau d'haplotype pour le gène COI de *Thaumetopoea pityocampa*. Chaque cercle représente un haplotype, la taille des cercles est proportionnelle au nombre d'individus. Les lettres majuscules indiquent dans quelle zone les haplotypes apparaissent (zones énoncées dans le Tableau I et visible sur la Figure 2). Le graphique B représente la distribution de chaque haplotype sur les îles, le littoral ou à l'intérieur des terres.



**Figure 2.** Carte représentant la position de chaque population ainsi que la proportion de chaque haplotype au sein des différentes populations sous forme de diagramme circulaire. Les différentes zones énoncées dans le tableau I y sont indiquées. Les points sur la carte représentent la limite nord de l'aire de répartition de *Thaumetopoea pityocampa*.

Sur la carte de disposition de ces haplotypes (Figure 2) nous pouvons constater que l'haplotype majoritaire HT1, bien qu'il ne se trouve pas dans toutes les populations, se trouve reparti sur l'ensemble de l'aire d'étude de manière très fréquente. Certains autres se trouvent limités à une zone géographique particulière. C'est notamment le cas de HTF, HTG et HTH qui ne se trouvent uniquement qu'en Dalmatie central et méridionale (Zone C). Ou bien le cas de HTA, HTB et HT17 uniquement dans le nord de l'Adriatique (Zone A). HT12 se trouve partagé entre le nord de l'Adriatique et la Dalmatie septentrionale (Zone A et B). Le Montenegro (Zone D) possède, dans une zone très restreinte, deux haplotypes que l'on ne trouve nul par ailleurs (HTC et HTE). On constate de même que les zones géographiques A, C, D et E possède une plus grande diversité haplotypique que les autres zones. Le nord de l'Adriatique regroupe les Haplotypes HTA, HT12, HT17 et HTB. La Dalmatie centrale et méridionale regroupe les haplotype HT1, HTF, HTH et HTG. A l'inverse l'ensemble de l'Italie semble assez monomorphique en n'ayant pour seul haplotype HT1, à l'exception de la zone Ligurie (Zone E) qui possède HT2, HT3 et HT4.

### ***3.2.2. Diversité génétique des populations***

Pour chacune des populations, la diversité génétique (h) et la divergence (k) sont données dans le Tableau II. La diversité génétique varie de 0 à 8 et k est compris entre 0 et 10.6. Dans un certain nombre de populations on constate la présence d'un seul haplogroupe. Les populations de Losinj et de Rollo possèdent la diversité génétique (h) et la divergence (k) la plus élevée. De plus, on constate que la présence de HTB dans une population augmente beaucoup la divergence calculée.

Les AMOVA calculées (Tableau IV) ont permis de montrer, grâce au premier groupement, qu'il y a un effet, expliquant la structure de la zone, non négligeable (18.9 %) lorsque l'on groupe les îles et le littoral contre l'intérieur des terres. De plus, avec le second groupement, on constate que la distinction des zones géographiques explique la moitié de la structure observée et qu'il y a donc une forte structure géographique. A l'inverse, on ne constate aucun effet hôte sur la structure.

**Tableau IV.** Analyse de la variance moléculaire (AMOVA), entre les populations de *Thaumetopoea pityocampa* au niveau des Balkan. Les groupe testés sont, les îles/littoral contre intérieur des terres, les zones géographiques (énoncées dans le Tableau I) et les hôtes (les pin des montagnes (*Pinus nigra* et *sylvestris*) contre les pin maritime (*Pinus halpensis*)).(\*\*) Très significatif ; (\*) Significatif ; (ns) Non significatif.

Structure	Source de la variation	Variance (%)	P-value
Regroupement	Entre les groupes	18,90	0,00978(*)
Îles/Littoral contre intérieur des terres	Entre les populations d'un groupe	60,56	0,00000(**)
	Dans les populations	20,54	0,00000(**)
Regroupement par zone géographique	Entre les groupes	50,22	0,00000(**)
	Entre les populations d'un groupe	29,55	0,00000(**)
	Dans les populations	20,23	0,00000(**)
Regroupement par pin hôte	Entre les groupes	-2,28	0,59531(ns)
	Entre les populations d'un groupe	79,13	0,00000(**)
	Dans les populations	23,15	0,00000(**)

#### **IV. DISCUSSION**

##### ***4.1 COI : un marqueur suffisamment variable pour notre étude... mais avec trop de copies***

Avec 13 haplotypes au sein de notre zone d'étude, la variabilité et le degré de divergence du marqueur utilisé se sont avérés être suffisants pour détecter des centres de différenciation de la processionnaire du pin sur le pourtour de la mer Adriatique. Par contre, la présence de multiples copies du gène nous a amené à ne pas prendre en considération un grand nombre de séquences, pouvant conduire à des biais d'observation dans la distribution de la diversité génétique. Dans l'ensemble, la diversité a probablement été sous-estimée dû aux séquences ambiguës éliminées qui auraient pu révéler des haplotypes rares non observées dans les séquences pures restantes. Toutefois, les problèmes rencontrés appellent une vérification de la nature réellement mitochondriale de tous les haplotypes mis en évidence, en particulier du plus divergent (HTB). Dans ce cas, la diversité mitochondriale aurait pu être surestimée en considérant, comme nouveaux haplotypes, des copies non mitochondriales. Le nombre de pas mutationnels entre HTB et l'haplotype majoritaire HT1 est très important et bien supérieur à celui observé pour les autres haplotypes de la région (de 1 à 4 pas mutationnels). Ceci pourrait être la signature d'une copie non codante intégrée dans le génome nucléaire. Toutefois, la séquence traduite en acides aminés ne contient aucun codon

stop (Annexe 3), ce qui est compatible avec une séquence fonctionnelle et une origine mitochondriale. Enfin, les 2 individus porteurs de cet haplotype dans le jeu de données n'ont posé aucun problème lors de la correction des séquences. Il paraît peu vraisemblable que ces séquences aient pu être amplifiées préférentiellement s'il s'agissait de séquences nucléaires parasites. Quand cet haplotype est apparu sous forme de pic minoritaire ou équivalent au milieu d'une autre séquence, il était généralement distribué dans des populations proches de l'île de Losinj où il a été détecté. Ceci pourrait laisser penser que HTB puisse exister sous forme de copie mitochondriale originale dans certaines populations mais aussi sous forme de copie transportée et transmise par les mâles dans des populations environnantes. Il en était de même pour la plupart des séquences en mélange quel que soit l'haplotype. Pour que des doubles pics soient présents, il fallait généralement observer les deux séquences sous forme pures dans les environs.

Pour valider tous les HT mitochondriaux mis en évidence et déterminer leur nature et leur origine, il faudra réaliser un clonage des produits de PCR des séquences contenant des sites ambigus afin d'identifier la copie mitochondriale de la copie nucléaire. Ensuite, en supposant qu'il s'agisse de copies nucléaires, il faudra mettre en œuvre un protocole de lyse alcaline (Girish et al 2013) pour isoler les mitochondries et identifier les séquences réellement mitochondriales. Un séquençage des mêmes individus avec et sans repliG permettra de valider l'intérêt et déterminer les limites de cette technique beaucoup plus simple à mettre en œuvre que la lyse alcaline (nécessitant du matériel biologique frais).

## ***4.2 Identification des zones refuges***

### ***4.2.1. Balkan : une zone refuge majeur de la processionnaire.***

Sur l'ensemble de la zone d'étude, aucun effet hôte sur la structure génétique n'a été mis en évidence (Tableau IV). Cependant, l'existence d'une forte structure géographique a été révélée par l'AMOVA prenant en compte sept groupes géographiques de A à G (Tableau IV et Figure 2).

En Italie, seule une population (ligure zone E) très localisée a montré de la diversité haplotypique (Figure 2), avec la présence de 3 haplotypes endémiques relativement divergents, de 2 à 3 pas mutationnels (Figure 1A) et des valeurs de h de 0.8 et de k de 3.4

(Tableau II). Sinon, le reste de l'Italie semble plutôt monomorphique. Toutefois, l'échantillonnage dans cette région, en particulier sur la façade adriatique, mériterait d'être renforcé.

A l'inverse, la côte adriatique balkanique, de la frontière italo-slovène au Monténégro, présente une très forte diversité génétique. Le nombre d'haplotypes et le degré de divergence se trouvent être d'un ordre de grandeur comparable à celui observé sur l'ensemble des principales zones refuges identifiées, dans la partie ouest de l'aire, le long de la cote ibérique méditerranéenne (de la chaîne bétique aux Pyrénées).

La France correspond à une zone de recolonisation post-glaciaire partagée entre deux lignées respectivement à l'ouest et à l'est (Rousselet et al 2010). L'étude de 2010 n'avait pas permis d'identifier clairement l'origine de la lignée est. Il apparaît maintenant très vraisemblable que son origine et son centre principal de différenciation se situent au niveau des Balkans. Ce patron phylogéographique n'est pas un cas isolé puisqu'il a déjà été observé chez de nombreuses espèces. (Taberlet & Fumagalli 1998, Hewitt 1999, Schmitt 2007).

#### **4.2.2. Structuration de la zone refuge : « refugia within refugia »**

Cette zone refuge globale n'est pas homogène et présente une forte structure spatiale permettant de distinguer plusieurs zones de différenciation, dont principalement le nord-est de l'Adriatique (A), la Dalmatie (B, C) et le Monténégro (D). Il est intéressant de noter que pour chacune des 4 zones de hot spot de diversité de plantes méditerranéennes, présentes dans les Balkans et décrites dans l'article de Médail et Diadéma 2009 (Annexe 5), on observe une diversité plus forte que dans le reste de l'aire (peu marqué pour la zone B à très marqué pour les zones A, C et D). L'existence de plusieurs zones refuges au sein d'une zone refuge globale est bien connue pour la péninsule ibérique et a été décrite sous le patron « refugia within refugia » (Gomez & Lunt 2006).

L'existence de zones de diversité chez un grand nombre d'espèces thermophiles, en particulier végétales, indique que le climat glaciaire dans cette région est probablement resté suffisamment favorable pour une espèce comme la processionnaire. Par ailleurs, alors qu'il est plus difficile de se prononcer sur la présence ou non de pin d'alep faute de donnée, le pin noir, qui persiste dans cette région aujourd'hui, a probablement toujours été présent. Comme pour la péninsule ibérique (Carrion et al 2000, Robledo-Arnuncio et al 2005 Afral-Rafii & Dodd 2007), les pins de montagne comme le pin noir et le pin sylvestre ont probablement

connu des phases d'expansion démographique et d'extension de leur aire de distribution lors de certaines phases de la période glaciaire (avec une descente en latitude et en altitude). Lors du réchauffement de l'holocène, la régression de ses pins a été compensée par un développement des populations de pins thermophiles comme le pin d'alep. Il y a donc toujours eu persistance d'hôtes de la processionnaire dans cette région. Il est important de noter qu'il existe des reliques de populations de pins noirs dans de nombreuses îles croates (voire de sylvestres sur l'île de KrK au nord de l'Adriatique) (Annexe 5). C'est notamment le cas sur les îles de Hvar et de Brac où est situé une variété endémique de pin noir. Cette région dominée par l'HTF pourrait donc correspondre à une des zones refuges majeures de l'insecte pendant les glaciations et à un des centres majeurs de différenciation au sein des Balkans. Pour la zone nord-est de l'Adriatique, autre zone majeure de différenciation, on peut envisager qu'il y ait eu une rétraction vers le nord ainsi qu'en altitude de l'aire de répartition des pins noirs et sylvestres avec le réchauffement de l'holocène. L'insecte dispose aujourd'hui principalement de pin d'alep mais il existe des populations reliques persistantes de pin de montagne au niveau d'altitude plus élevée. Globalement les îles et le littoral présentent une forte diversité haplotypique et on y rencontre la majorité des haplotypes les plus divergents. En effet, l'AMOVA nous montre un effet significatif pour le regroupement île/littoral contre intérieur des terres. Seul le Monténégro, plus au sud, semble avoir pu offrir des zones de persistance plus à l'intérieur des terres.

#### ***4.2.3 L'ancienne plaine Adriatique : un refuge recouvert par les eaux ?***

Le degré de différenciation observé entre îles et continent n'est clairement pas compatible avec un isolement datant de l'holocène et dû à la remontée des eaux car on estime, que COI a un taux d'évolution chez les insectes de 1,5% par Million d'année (Farrell 2001). Même si les datations moléculaires sans possibilité de calibrage sont sujettes à caution, on peut voir que les distances génétiques (allant de 0.1% à 2.4%) entre des haplotypes îliens et continentaux ne sont pas compatibles avec une divergence récente. Les distances forte seraient même plutôt compatibles avec une persistance de population dans cette région, sur un à plusieurs cycles glaciaires. On peut donc, à l'inverse, envisager qu'il a existé au niveau de l'Adriatique de vastes zones favorables à la Processionnaire du pin aujourd'hui recouvertes par la mer qui a fragmenté l'aire. La superficie des territoires disponibles pour la processionnaire dans cette région auraient fortement variée au cours du temps. Cependant, une partie des territoires disponibles aurait toujours subsisté. En effet, au vu de leur

endémisme, les haplotypes partagés entre îles, sont plus vraisemblablement le fruit d'une fragmentation de l'aire du à l'envahissement des vallées par les eaux, que d'une dispersion récente d'île en île. A l'inverse de ces haplotypes très localisés, HT1 est peu ou pas présent dans les îles. Ceci indiquant que cet haplotype, qui a colonisé toute l'Europe, n'a pas été capable de coloniser les îles, sans doute dû au fait que les femelles sont incapables de franchir la mer. En effet, HT1 n'a pas réussi à se retrouver sur les îles. De plus, HTF n'a pas réussi à coloniser le continent pourtant aussi proche que les îles entre elles. Exception faite au niveau de Seget probablement dû au fait qu'une suite de petites îles forme un « pont » permettant à quelques femelles de passer des îles au continent. Cette zone ne semble pas avoir énormément contribué à la recolonisation post-glaciaire. En effet, HTF étant très localisé dans cette région, on peut supposer que les individus de cette zone faisaient parties d'une même population qui s'est fragmentée en îles et littoraux devenant ainsi des populations reliques.

Le patron de diversité et de divergence observé mais aussi le nombre d'haplotypes manquants dans le réseau sont compatibles avec les prédictions formulées dans le cadre de l'hypothèse d'une zone refuge Adriatique recouvert par la mer. Même si l'insecte a persisté malgré les variations du climat, les variations du niveau de la mer ont probablement engendré des variations drastiques du niveau des populations et donc une perte de diversité génétique. Le patron de diversité et le degré de divergence observés seraient le fruit de la persistance de l'insecte dans cette région malgré les variations climatiques. Mais, la structure du réseau et le nombre d'haplotypes manquants seraient l'empreinte des fortes fluctuations démographiques subi au cours des cycles d'expansions/contractions de l'aire.

## **V. CONCLUSIONS ET PERSPECTIVES**

Nos analyses ont montré que le réseau d'haplotype de cette zone se trouvait être complexe, de nombreux haplotypes se trouvent être manquants et les haplotypes présents sont souvent séparés par plusieurs pas mutationnels. De plus certaines zones géographiques se trouvent être plus riches en haplotype que d'autres. Ces éléments représentent des arguments favorables à l'hypothèse d'une zone refuge adriatique recouverte par les eaux durant l'holocène.

Il faudra cependant, effectuer de nouvelles collectes d'échantillons dans toutes les îles de la région afin d'avoir plus d'informations sur les haplotypes présents dans les îles et leurs

liens entre elles. Des échantillonnages plus à l'intérieur du continent devraient être faits afin de mieux délimiter cette zone refuge.

Il faudrait de plus, pour certains haplotypes, comme HTB et HTH, faire plus de tests tels que le clonage ou la lyse alcaline afin de définir si ce sont de vrais haplotypes ou seulement des pseudogènes.

L'origine précise de l'haplotype HT1 reste incertaine. Il est très fréquent sur le littoral et les îles de la Dalmatie septentrionale (Zone B) mais cette zone a été très fortement perturbée par les activités anthropiques. Ces activités nous empêchent de savoir si cette zone correspond au lieu d'origine de HT1 ou si cette fréquence est seulement dû à l'activité humaine. Il serait intéressant de faire plus d'échantillonnage sur place afin de voir s'il existe d'autre haplotype dans cette zone ou si elle se trouve être purement colonisée par HT1. Il faudrait, de plus, faire des recherches sur le type d'activité humaine et vérifier les importations de pin dans cette zone qui pourraient être la cause de ce monomorphisme local.

### **Remerciements.**

A l'issue de mon stage, je tiens à remercier toutes les personnes qui ont contribué de près ou de loin à l'écriture de ce mémoire. Je tiens, tout d'abord, à remercier Alain ROQUE ainsi que François LIEUTIER qui m'ont permis d'effectuer ce stage de master 2. Je tiens à remercier Jérôme ROUSSELET, qui, en tant qu'encadrant, s'est toujours montré à l'écoute et très disponible tout au long de la réalisation de ce mémoire, ainsi que pour l'inspiration, l'aide et le temps qu'il a bien voulu me consacrer. Un grand merci à Emmanuelle MAGNOUX, Béatrice COURTIAL et à Claudine COURTIN pour leur apprentissage de la technique de laboratoire ainsi que pour leur aide lors des nombreux problèmes que j'ai pu rencontrer tout au long de ce stage. Je tiens, de plus, à remercier l'ensemble de l'équipe zoologique pour leur accueil et leur aide tout au long du stage. Enfin, je tiens également à remercier les post docs, thésards et les autres stagiaires, avec qui j'ai pu partager la salle de travail, pour leur accueil et leur gentillesse qui m'ont permis d'effectuer ce stage dans une excellente ambiance.



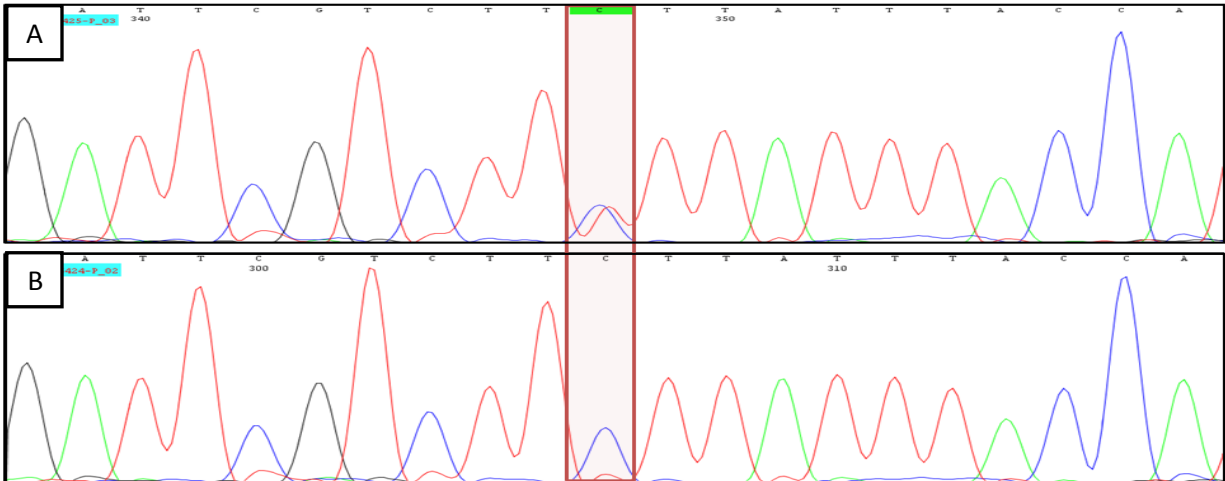
## **Références bibliographiques :**

- Afzal-Rafii, Z. and R.S. Dodd.** 2007. Chloroplast DNA supports a hypothesis of glacial refugia over postglacial recolonization in disjunct populations of black pine (*Pinus nigra*) in western Europe. *Molecular Ecology*. **16**:723–736.
- Avise J.C.** 2000. *The history and formation of species, Harvard*. 447 pages
- Barton E. and D. Pollard.** 2002. High-Resolution Climate Simulations of Oxygene Isotope Stage 3 in Europe. *Quaternary research*. **58**:296-309
- Bhagwat S.A. and K.J. Willis.** 2008 Species persistence in northerly glacial refugia of Europe: a matter of chance or biogeographical traits? *Journal of Biogeography*. **35**:464-482.
- Canestrelli D., A. Cimmaruta, V. Costantini and G. Nascetti.** 2006. Genetic diversity and phylogeography of the Apennine yellow-bellied taod *Bombina pachypus*, with implications for conservation. *Molecular Ecology*. **15**:3741-3754.
- Carrion, J.S., C. Navarro, J. Navarro and M. Munuera.** 2000. The distribution of cluster pine (*Pinus pinaster*) in Spain as derived from palaeoecological data: relationships with phytosociological classification. The Holocene. **10**:243–252.
- Clement M., D. Posada and K.A. Crandall.** 2000. TCS: a computer program to estimate gene genealogies, *Molecular Ecology*. **9**:1657–1659.
- DeChaine E.G. and A.P. Martin.** 2005. Historical biogeography of two alpine butterflies in the Rocky Mountains: broad-scale concordance and local-scale discordance, *Journal of Biogeography*. **32**:1943-1956
- Excoffier L.G, X. Laval and S. Schneider.** 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*. **1**:47–50.
- Excoffier L.G., M. Foll and R. Petit.** 2009. Genetic consequences of range expansions. *Annual Review of Ecology, Evolution and Systematics*. **40**:481-501
- Farrell B.D.** 2001. Evolutionary assembly of the milkweed fauna : cytochrome oxidase I and age of *Tetraopes* beetles. *Molecular Phylogenetics and Evolution*. **18**:467-478
- Girish P.S., S. Haunshi, S. Vaithyanathan, R. Rajitha and C. Ramakrishna.** 2013. A rapid method for authentication of Buffalo (*Bubalus bubalis*) meat by Alkaline Lysis method of DNA extraction and species specific polymerase chain reaction. *Journal of food science and technology-mysore*. **50**:141-146
- Gomez A and D.H. Lunt.** 2006. Refugia within refugia : patterns of phylogeographic concordance in the Iberian Peninsula. *Phylogeography of southern European refugia – evolutionary perspectives on the origins and conservation of European biodiversity*. Eds. S. Weiss and N. Ferrand. Kluwer. Academic Publishers, Dordrecht, pp 155–188.
- Goto H., B. Dickins, E. Afgan, I.M. Paul, J. Taylor, K.D. Makova and A. Nekrutenko.** 2011. Dynamics of mitochondrial heteroplasmy in tree families investigated via a repeatable re-sequencing study. *Genome Biology*. **12**:6
- Hall T.A.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, *Nucleic Acids Symposium Series*. **41**:95–98.
- Hewitt G.M.** 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*. **68**:86-112
- Hewitt G.M.** 2000. The genetic legacy of the Quaternary ice ages. *Nature*. **405**:907-913

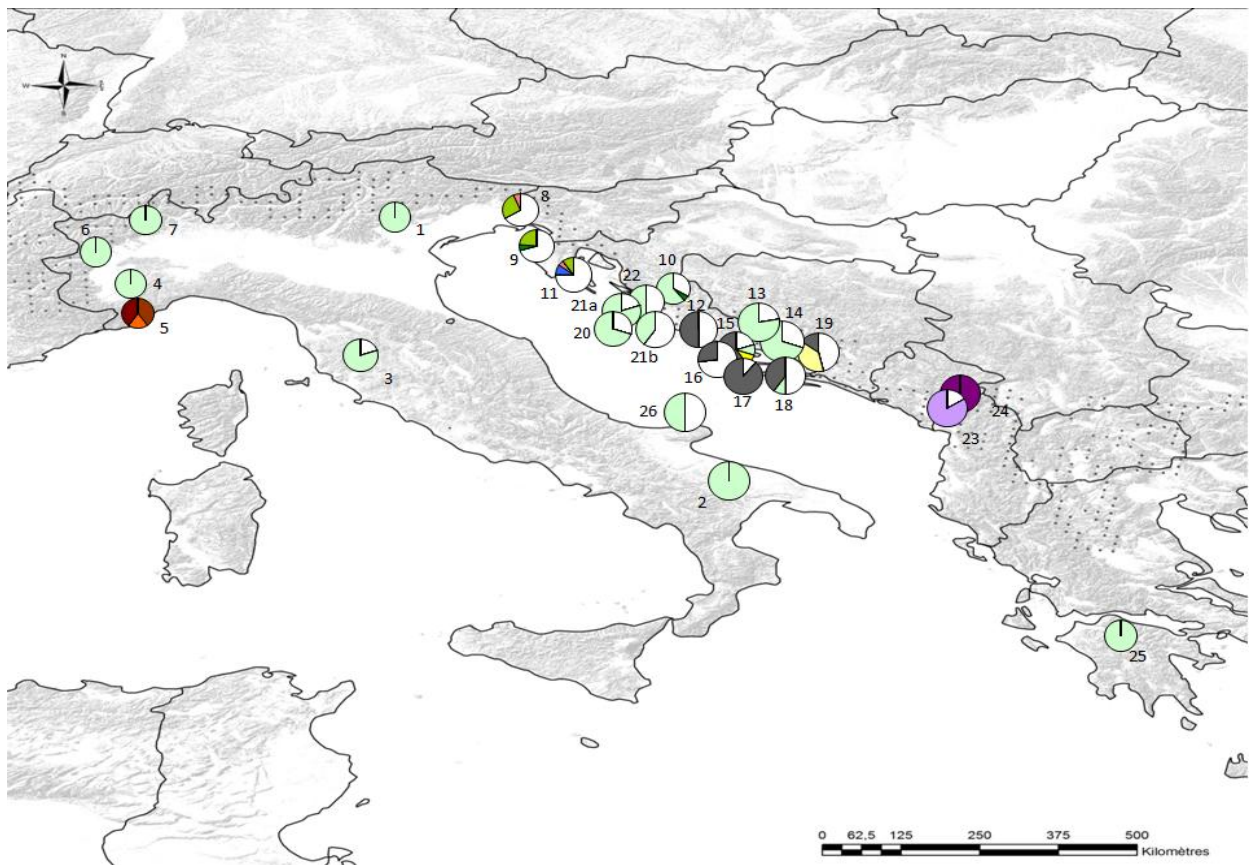
- Hewitt G.M.** 2004a. Genetic consequences of climatic xxx in the Quaternary. *Royal society of London*. **359**:183-195
- Hewitt G.M.** 2004b. The structure of biodiversity-insight from molecular phylogeography. *Frontiers in zoology*. **1**:4
- Kerdelhué C., L. Zane, M. Simonato, P. Salvato, J. Rousselet, A. Roques and A. Battisti.** 2009. Quaternary history and contemporary patterns in a currently expanding species C. *Evolutionary Biology*. **9**:220
- Lambeck K., T.M. Esat and E.K. Potter.** 2002. “Links between climate and sea levels for the past three million years”. *Nature*. **419**:199-206
- Lucchini V. and E. Randi.** 1998. “Mitochondrial DNA sequence variation and phylogeographical of rock partridge (*Alectoris graeca*) populations” *Heredity*. **81**:528-536
- Médail F. and K. Diadema.** 2009. Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography*. **36**:1333–1345.
- Robinet C., C.E. Imbert, J. Rousselet, D. Sauvard, J. Garcia, F. Goussard and A. Roques.** 2012.. Human-mediated long-distance jumps of pine processionary moth in Europe. *Biol Invasions*. **14**:1557-1569
- Robledo-Arnuncio, J.J., C. Collada, R. Ali´a, and L. Gil.** 2005. Genetic structure of montane isolates of *Pinus sylvestris* L. in a Mediterranean refugial area. *Journal of Biogeography*. **32**:595–605.
- Rozas J, J., C. Sanchez-DelBarrio, X. Messeguer and R. Rozas.** 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*. **19**:2496–2497.
- Rousselet J., R. Zhao, D. Arga, M. Simonato, A. Battisti, A. Roques and C. Kerdelhué.** 2010. The role of topography in structuring the demographic history of the pine processionary moth, *Thaumetopoea pityocampa* (Lepidoptera: Notodontidae). *Journal of Biogeography*. **37**:1478-1490
- Salvato P., A. Battisti, S. Concato, L. Masutti, T. Patarnello and L. Zane.** 2002. Genetic differentiation in the winter pine processionary moth ( *Thaumetopoea pityocampa-wilkinsoni* complex), inferred by AFLP and mitochondrial DNA markers, *Molecular Ecology*. **11**:2435-2444.
- Schmitt T.** 2007. Molecular biogeography of Europe: Pleistocene cycle and postglacial trends. *Frontiers in Zoology*. **4**:11
- Simonato M., Z. Mendel, C. Kerdelhué, J. Rousselet, E. Magnoux, P. Salvato, A. Roques, A. Battisti and L. Zane.** 2007. “Phylogeography of the pine processionary moth *Thaumetopoea wilkinsoni* in the Near East”. *Molecular Ecology*. **16**:2273-2283
- Stewart J.R., A.M. Lister.** 2001. Cryptic northern refugia and the origins of the moder biota. *TRENDS in Ecology & Evolution*. **11**:608-613
- Taberlet P., L. Fumagalli, A.G. Wust-saucy and J.F. Cosson.** 1998. Comparative phylogeography and postglacial colonization routes in Europe. **7**:453-464
- Varga Z. and T. Schnitt.** 2008. Type of orcal and oreotundral disjunctions in the western Palearctic. *Biological Journal of Linnean Society*. **92**:415-430
- Wolff J.N., M. Nafisinia, P. Sutovsky and J.W.O. Ballard** (2013) Paternal transmission of mitochondrial DNA as an intedral part of mitochondrial inheritance in metapopulations of *Drosophila simulans*. *Heredity*. **110** :57-62

## ANNEXES

**Annexe 1.** Comparaison d'électrophorégrammes de séquences d'ADN mitochondrial. Electrophorégrammes de l'échantillon 5425 avec présence de pseudogènes (A) et de l'échantillon 5424 sans pseudogènes (B).



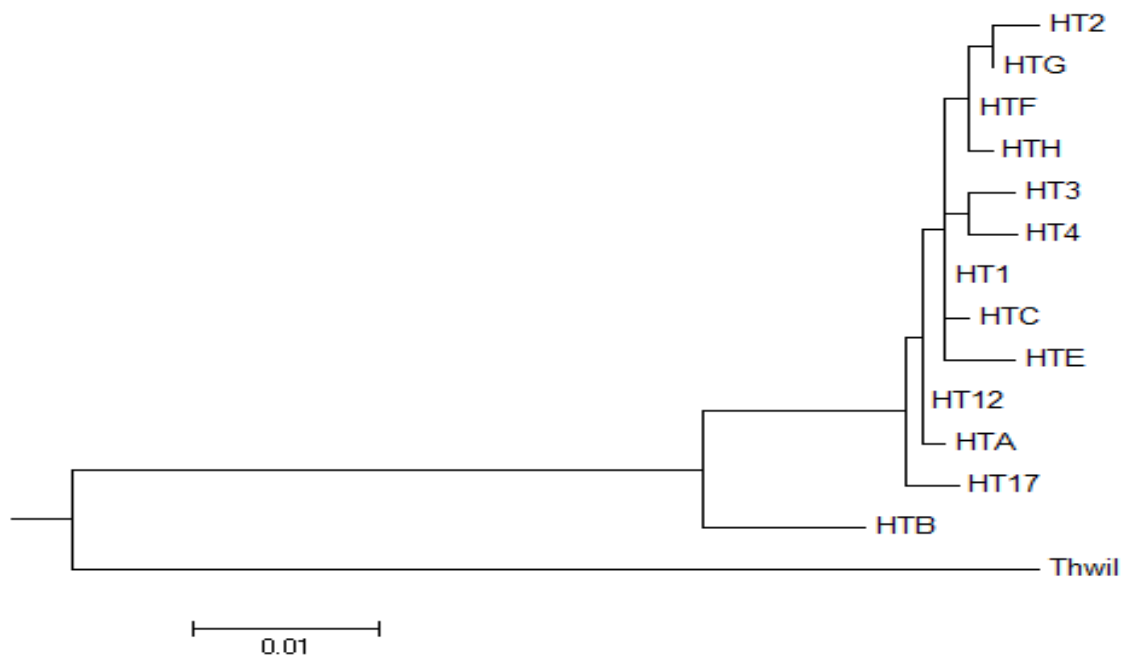
**Annexe 2.** Carte représentant la position de chaque population, la proportion de chaque haplotype, ainsi que la proportion de séquence non exploitable (en blanc), au sein des différentes populations sous forme de diagramme circulaire. Le nombre de séquences tentées pour chaque population se trouve dans le tableau I. Les points sur la carte représentent la limite nord de l'aire de répartition de *Thaumetopoea pityocampa*.



**Annexe 3.** Sites polymorphiques sur 13 haplotypes trouvés dans les alignements des séquences de COI chez *Thaumetopoea pityiocampa*. Sur le fragment de 802 pb, 30 sites de polymorphismes ont été mis en évidence. La seconde partie du tableau présente les changements observés sur les acides aminés.

	Position nucléotidiques																															Position acides aminés				
	15	43	81	84	105	136	156	162	174	192	256	306	318	321	372	417	441	480	489	570	582	636	652	654	721	723	724	736	744	783		15	86	218	241	246
HT1	C	T	A	A	A	T	A	C	A	A	G	C	T	G	T	A	A	C	C	A	T	A	T	T	T	C	C	C	C	C	HT1	S	A	F	S	P
HT2	.	C	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	HT2	P	T	.	.	.
HT3	.	C	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	HT3	P	.	.	P	.
HT4	.	.	.	.	.	.	G	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	HT4	.	.	L	.	.
HT12	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	HT12	.	.	.	.	.
HT17	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	G	.	.	.	.	.	.	.	.	T	.	.	HT17	.	.	.	.	S
HTA	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	HTA	.	.	.	.	.
HTB	T	.	G	.	G	.	.	.	G	.	.	T	C	A	.	G	.	T	T	.	C	.	.	.	.	A	T	G	T	T	HTB	.	.	.	.	A
HTC	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	HTC	.	.	.	.	.
HTE	.	.	G	.	.	.	.	.	.	C	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	HTE	.	.	.	.	.
HTF	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	HTF	.	.	.	.	.
HTG	.	.	.	G	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	HTG	.	T	.	.	.
HTH	.	.	.	G	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	HTH	.	.	.	.	.

**Annexe 4.** Arbre en *Neighbor-joining* représentant la distance génétique séparant deux haplotypes. Thwil ( *Thaumetopoea wilkinsoni*) est utilisé ici en tant que groupe externe.



**Annexe 5.** Carte représentant les zones de hot spot décrites par Médail et Diadema en 2009, les zones de pins reliques ainsi que l'emplacement de prélèvement des échantillons.

