

TABLE DES MATIERES

INTRODUCTION.....	7
PREMIERE PARTIE : Terminologie et modèle d'étude	11
I. Historique de la classification des êtres vivants	11
1. Fixisme et créationnisme, héritages de l'Antiquité.....	11
2. Introduction des concepts de transformisme, puis d'évolution	13
3. Une révolution technologique : l'outil moléculaire	15
II. Etude de la diversité du vivant à l'échelle des espèces, et des populations	16
1. Théories sous-jacentes à la reconstruction des phylogénies	16
2. Applications pratiques.....	17
III. <i>Forficula auricularia</i> , organisme modèle pour l'étude de la spéciation	19
1. Présentation générale de l'objet d'étude.....	19
2. Particularité de l'organe caudal, le forceps.....	20
3. Dynamique des populations	22
DEUXIEME PARTIE : Travail personnel.....	25
I. Introduction.....	25
II. Matériels et méthodes	25
1. Collection et identification des individus.....	25
2. Protocole morphogéométrique	27
3. Tests préalables	29
4. Extraction d'ADN, amplification et séquençage de deux gènes mitochondriaux...	30
4.1 Extraction d'ADN	30
4.2 Amplification par PCR de la région 5' de <i>COI</i>	30
5. Analyse des séquences et construction d'arbres phylogénétiques.....	31

III. Résultats	33
1. Résultats préliminaires	33
2. Localisation des deux espèces en Europe	33
3. Phylogénie	35
4. Diversité et structuration des populations	37
5. Flux de gènes entre populations	38
6. Morphogéométrie	40
6.1. Le dépassement de l'aile gauche.....	40
6.2. Le pronotum	43
6.3. Les forceps	46
6.4. Synthèse	47
DISCUSSION	49
CONCLUSION	51
BIBLIOGRAPHIE	53
ANNEXES	57
ANNEXE 1 : Lexique	59
ANNEXE 2 : Liste des contributeurs à l'échantillonnage de perce-oreilles	61

LISTE DES FIGURES

Figure 1 : Rhinocéros de Dürer, gravure de 1515, Allemagne.	14
Figure 2 : L' « homoncule », une des particules sources envisagées pour l'homme, d'après les premières observations de spermatozoïdes au microscope. Gravure de Nicolas Hartsoeker, 1694	15
Figure 3 : Un individu mâle du perce-oreille européen, <i>Forficula auricularia</i>	21
Figure 4 : Comparaison des pinces chez les deux types de mâles et les femelles.	22
Figure 5 : Localisation des structures étudiées en morphogéométrie, sur un individu mâle.	29
Figure 6 : Positionnement des longueurs étudiées, sur le pronotum, pour le test simplifié...	31
Figure 7 : Amorces utilisées pour amplifier nos deux gènes d'intérêt.	33
Figure 8 : Répartition des deux espèces parmi les populations échantillonnées.....	36
Figure 9 : Neighbor Joining de notre échantillon réalisé après 500 bootstraps.	37
Figure 10 : Arbre populationnel de nos échantillons en Neighbor Joining.....	39
Figure 11 : Comparaison des « skylines » entre les deux espèces jumelles A et B.	41
Figure 12 : ACP du dépassement de l'aile gauche sur les deux premiers axes réalisée après superposition Procruste.....	42
Figure 13 : Boîtes à moustaches pour la taille des ailes, comparatives entre les clades.	44
Figure 14 : ACP du pronotum sur les deux premiers axes réalisée après superposition Procruste.	45
Figure 15 : Boîtes à moustaches pour le pronotum, comparatives entre les clades.	47
Figure 16 : ACP des forceps sur les deux premiers axes réalisée après superposition Procruste.	48

LISTE DES TABLEAUX

Tableau 1 : Correspondance des codes, des nombres d'individus séquencés (N) pour les deux marqueurs, <i>COI</i> et <i>COII</i> , et des localisations, pour notre échantillonnage.	28
Tableau 2 : Protocole de PCR utilisé pour nos deux gènes d'intérêt.....	32
Tableau 3 : Comparaison des indices de diversités calculés avec DnaSP sur nos échantillons de grand effectif.....	40
Tableau 4 : Valeurs de Fst calculées avec DnaSP.....	41
Tableau 5 : Synthèse des résultats statistiques.....	49

ABBREVIATIONS

ACP	Analyse en Composantes Principales
ANOVA	<i>Analysis of Variance</i> / Analyse de la variance
ADN	Acide désoxyribonucléique
BLAST	<i>Basic Local Alignment Search Tool</i>
COI	Cytochrome oxydase 1
COII	Cytochrome oxydase 2
ESS	<i>Evolutionarily Stable Strategy.</i>
MANOVA	<i>Multivariate Analysis of Variance</i> / Analyse de la variance multivariée
NCBI	<i>National Center for Biotechnology Information</i>
NJ	<i>Neighbor joining</i>
PCR	<i>Polymerase Chain Reaction</i> / Réaction de polymérisation en chaîne
TMRCA	<i>Time for the Most Recent Common Ancestor</i> / Âge du plus récent ancêtre commun
tRNA leu	ARN de transfert de la leucine

INTRODUCTION

« The so-called species problem can be reduced to a simple choice between two alternatives: are species reality of nature or are they simply theoretical constructs of the human mind ? »
(MAYR, 1982)

(Le problème dit des espèces peut se voir réduit à un simple choix entre deux possibilités : les espèces sont-elles une réalité de la nature, ou simplement des constructions intellectuelles de l'esprit humain?)

Cette question d'Ernst Mayr (1904-2005), biologiste évolutionniste majeur du XX^{ème} siècle, et plus généralement les interrogations sur la création d'êtres divers, aux « fonctions » variées, hantent l'esprit humain depuis que l'homme s'est mis en tête de différencier les êtres vivants entre eux, et est donc probablement presque aussi ancienne que l'homme lui-même. Les grecs anciens, dont Aristote (384-322 av. J. C), le père de la zoologie, ont voulu expliquer la formation des organes et la réalisation d'un être par la notion de finalisme formateur (les organes existent dans le but de remplir des buts préconçus, la jambe car il faut marcher, l'œil car il faut voir, etc.). Pourtant, cela n'empêche pas Aristote d'envisager une théorie alternative, proche de ce qu'écrira bien des siècles plus tard le père de l'Evolution, Darwin : “Ainsi donc, toutes les fois que les choses se produisent accidentellement comme elles se seraient produites en ayant un but, elles subsistent et se conservent, parce qu'elles ont pris spontanément la condition convenable ; mais celles où il en est autrement périssent ou ont péri [...] ». Il rejette pourtant cette alternative quelques lignes plus loin : « Mais il est bien impossible que les choses se passent comme on le prétend [...]. Donc il y a un pourquoi, une fin à toutes les choses qui existent ou se produisent dans la nature.” (ARISTOTE, env. 350 av. J.C.).

La phylogénie, science permettant l'étude des relations entre les espèces, ou plus précisément entre les taxons (groupes dont la composition a été identifiée, donc reconnus dans une classification), se fonde sur l'étude des caractères, ou attributs observables d'un objet biologique, espèce ou gène, pour lesquels on définit différents états. Pendant des siècles, les phylogénistes ont étudié uniquement les caractères morphologiques. Or l'identification des spécimens biologiques sur la base de caractères morphologiques, par clés d'identification taxonomiques, est un processus complexe nécessitant l'intervention de spécialistes. De plus, cela requiert dans la plupart des cas des spécimens intacts, adultes, mâles et femelles (puisque'il n'est pas rare qu'un seul des genres porte les caractères nécessaires à l'identification). Depuis les années 70 et le développement, entre autres, des techniques de séquençage, la phylogénie a pris de plus en plus en compte l'outil moléculaire. Cet outil, utilisé conjointement à l'analyse morphologique, permet de représenter graphiquement les relations de parenté entre les divers taxons de la manière la plus précise.

Lorsque l'on réduit l'échelle d'observation, en matière de phylogénie, et que l'on veut préciser la limite qui existe entre individus particulièrement proches, on étudie alors le phénomène de spéciation. Il s'agit du processus par lequel une espèce initiale se scinde progressivement, par des mécanismes de **sélection naturelle**¹ et de **dérive génétique**, en deux espèces-filles isolées génétiquement et ayant en général des niches écologiques et des modes d'existence différents. Ce phénomène reste partiellement inexpliqué et demeure le sujet d'étude de nombreuses équipes de recherche dans les sciences de l'évolution. La thèse la plus couramment avancée est celle de la « spéciation **allopatrique** », selon laquelle des **populations** géographiquement isolées pendant une longue période se différencient passivement par dérive génétique, et deviennent si incompatibles que leur éventuels hybrides sont stériles. Leur morphologie tendrait à diverger au même rythme que la dérive génétique.

Depuis quelques années, le laboratoire EPHE Biologie intégrative des populations, UMR MNHN/CNRS 7205 Origine structure et évolution de la biodiversité, évalue le réalisme de cette thèse en utilisant le modèle biologique constitué par l'insecte *Forficula auricularia* (le perce-oreille européen). En 1998, l'équipe a démontré que sous ce nom étaient confondues deux espèces jumelles, notion établie précédemment par Mayr (MAYR, 1942), isolées génétiquement, reconnaissables par l'ADN, mais morphologiquement identiques. Ceci illustre le fait que la spéciation est principalement une rupture dans la continuité génétique : elle n'est pas obligatoirement corrélée avec des différences morphologiques.

La mise en évidence de zones de contact entre ces deux espèces est d'un grand intérêt pour les études de l'évolution. L'étude d'une zone de contact dans les Pyrénées (GUILLET *et al.*, 2003) a montré la coexistence d'espèces jumelles à moyenne altitude (1200 m) tandis qu'au même point géographique, à des altitudes supérieures et inférieures, une seule espèce a été trouvée (A en altitude, B en basse altitude). De plus, le dernier maximum glaciaire a eu un impact important sur la répartition des différentes espèces des milieux tempérés (HEWITT, 2000), comme ceux où les perce-oreilles vivent en Europe. La plupart des espèces de milieux aujourd'hui tempérés vivaient dans des refuges de la péninsule méditerranéenne. Récemment, l'utilisation de la **phylogéographie** a permis une avancée des connaissances sur l'histoire de plusieurs espèces au cours des cycles glaciaires. Comprendre les réponses des espèces aux changements climatiques anciens est un moyen efficace de prédiction de l'effet de la période actuelle de réchauffement global sur la répartition des espèces (PROVAN et BENETT, 2008). Il est donc instructif à cet égard de se pencher sur la dynamique des populations de *Forficula auricularia* en Europe, et particulièrement lors du dernier maximum glaciaire, 23 000 à 18 000 ans avant notre ère (DAWSON, 1992 ; HEWITT, 2003 ; HEWITT, 2004).

1 Tous les mots en gras renvoient à leurs définitions données en annexe 1 (ANNEXE 1).

D'autres **espèces cryptiques** de perce-oreilles, mises à part les deux formes mises en évidence en 1998, sont-elles présentes en Europe, et comment se répartissent ces deux dernières ? À l'aide d'une caractérisation par deux gènes mitochondriaux, nous avons étudié la distribution de nombreuses populations en Europe, dans l'absolu mais aussi par rapport à la distribution des refuges glaciaires, puis cherché à augmenter la finesse des résultats par des méthodes de génétique des populations. Considérant une adaptation écologique sous-jacente de chacune des espèces, nous voulions de plus examiner la possibilité d'un lien entre distribution, climat et altitude, mais aussi étudier la conservation des paramètres de croissance morphologique des deux espèces, afin de voir si les structures morphologiques étaient propres à chacune, et corrélées à de possibles adaptations locales. Ainsi, dans une première partie, nous présenterons une brève histoire de la classification des êtres vivants, puis nous détaillerons l'organisme modèle étudié : le perce-oreille européen. Dans une seconde partie, nous présenterons notre travail expérimental de morphométrie et de génétique.

PREMIERE PARTIE : Terminologie et modèle d'étude

Afin de poser les jalons de notre réflexion, après avoir passé en revue le déroulement chronologique des réflexions sur la classification en biologie depuis l'Antiquité jusqu'à nos jours, nous définirons les méthodes existantes pour modéliser ces organisations. Pour cela nous nous appuyerons entre autres, tout au long de cette partie, sur l'ouvrage *Peut-on classer le vivant ? Linné et la systématique aujourd'hui*, de Daniel Prat, Aline Raynal-Rocques, Albert Roguenant, compilant les avis de divers intervenants d'un colloque organisé en 2007 pour le tricentenaire de la naissance de Linné (PRAT *et al.*, 2008), ainsi que sur diverses archives compilées par la Bibliothèque Nationale de France (BNF, 2013). Enfin, nous discuterons de la pertinence du choix de l'insecte *Forficula auricularia* comme modèle d'étude.

I. Historique de la classification des êtres vivants

1. Fixisme et créationnisme, héritages de l'Antiquité

La volonté de l'homme de dresser l'inventaire des espèces vivantes a toujours existé. On peut en trouver des prémices dans la Grèce antique, particulièrement avec Aristote. Le Romain Pline l'Ancien (30 - 79), mort par inhalation de fumée pour avoir voulu observer de près l'éruption du Vésuve, avait déjà nommé un grand nombre d'espèces dans son *Histoire Naturelle*. A cette époque, et pendant des siècles, ce type d'études est principalement motivé par la volonté de dominer la nature, de considérer les êtres vivants selon leurs utilisations possibles pour l'homme, qu'elles soient d'ordre thérapeutique, alimentaire ou magique. Aristote avait d'ailleurs mis en place un classement du vivant, ancêtre de nos arbres phylogénétiques, basé sur un tout autre mode de hiérarchisation, la complexification graduelle de l'âme des êtres vivants, dite « échelle des êtres vivants » ou *Scala Naturae*. Dans cette vision, l'homme est considéré comme l'être le plus complexe après les divinités, et est indépendant du règne animal, en ce qu'il est le seul à avoir une âme rationnelle. C'est cette idée déiste qui est reprise par les religions monothéistes. On peut par exemple lire dans la Genèse (BIBLE, 1910) :

« Dieu dit: *Que la terre produise des animaux vivants selon leur espèce, du bétail, des reptiles et des animaux terrestres, selon leur espèce. Et cela fut ainsi.* Genèse 1.24

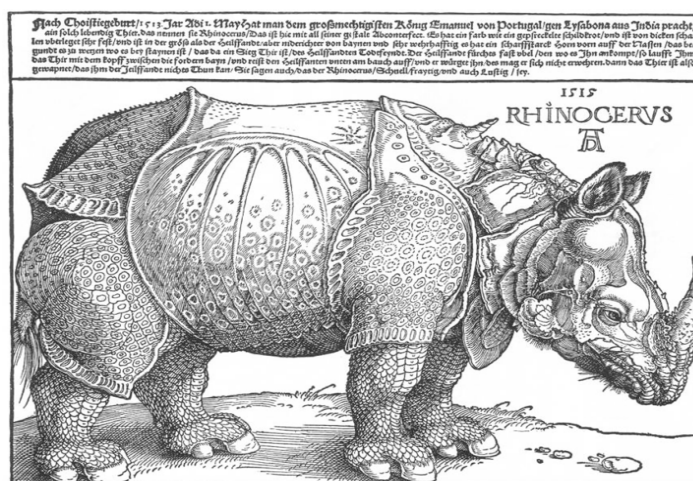
Dieu fit les animaux de la terre selon leur espèce, le bétail selon son espèce, et tous les reptiles de la terre selon leur espèce. Dieu vit que cela était bon. Genèse 1.25

Puis Dieu dit : Faisons l'homme à notre image, selon notre ressemblance, et qu'il domine sur les poissons de la mer, sur les oiseaux du ciel, sur le bétail, sur toute la terre, et sur tous les reptiles qui rampent sur la terre. » Genèse 1.26

Pendant près de quinze siècles, ce sont les classements des Anciens qui vont servir de référence aux biologistes de l'Occident, sans subir presque aucune modification. Au lieu d'observer la nature qui les entoure directement, les divers savants préfèrent se rapporter aux ouvrages de Pline l'Ancien ou du Grec Dioscoride (40 - 90), un des soucis notables de ce retour permanent aux textes antiques étant l'utilisation de notions vétustes quant aux plantes médicinales... De plus, et ce jusqu'au XVI^{ème} siècle, les plantes étaient classées en fonction de leurs propriétés pour l'homme (plante-médecine, plante-poison, etc.) mais aucunement selon leurs relations. Une infinité de noms coexistaient, souvent particulièrement longs et rébarbatifs, sans consensus aucun (LE GUYADER, 2003).

A la Renaissance pourtant va s'amorcer une sortie du carcan des Anciens avec l'augmentation progressive de la part de l'observation dans les sciences de la vie. Les biologistes vont alors se demander si les classifications utilitaires sont bien le reflet d'un « ordre naturel », y répondre par la négative, et subséquemment tenter de modifier ces classifications. Ce tournant s'illustre aussi par l'amélioration notable du réalisme des illustrations des ouvrages de biologie, ainsi que par l'enrichissement global des connaissances grâce aux grandes découvertes des navigateurs. Citons l'exemple du rhinocéros de Dürer (1471-1528), gravure réalisée par l'artiste à Nuremberg uniquement d'après les descriptions venant du Portugal et comportant de ce fait quelques erreurs (Figure 1).

Figure 1 : Rhinocéros de Dürer, gravure de 1515, Allemagne.



Le sultan indien Muzafar II (1511-1526) avait précédemment fait parvenir au roi du Portugal un rhinocéros, animal jamais encore observé en Europe. La gravure est réalisée uniquement d'après descriptions (d'après BM, 2013).

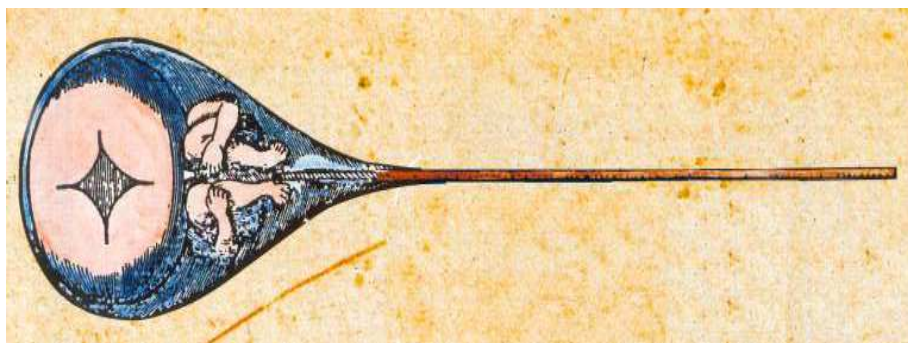
C'est au XVII^{ème} siècle que ce qui ressemble à notre méthode actuelle de classification voit le jour. On retiendra particulièrement l'apport de Carl Von Linné. Né en 1707 en Suède, ce médecin est à l'origine de ce qu'on nomme aujourd'hui la classification linnéenne, ou binominale (donnant deux noms latin, genre puis espèce, à chaque individu), utilisée dans le monde entier. Son ouvrage majeur, le *Systema naturae*, publié en 1735, donne une distribution précise de tous les représentants connus à l'époque des règnes minéral, végétal et animal. Il attribue à chacun une classe, un ordre, un genre et une espèce. C'est encore cette nomenclature normalisée qui a cours aujourd'hui. Malgré les nombreux détracteurs qu'aura

connus Linné (remettant notamment en cause le langage particulièrement technique utilisé, la place de l'homme au sein de sa classification, ainsi qu'une vision de l'ordre qui lui semble inspirée par Dieu lui-même), c'est bien là une révolution dans le monde des sciences naturelles. De nombreux débats ont lieu dans le sillon de cette avancée, et sont particulièrement vifs en France où les philosophes des Lumières discutent abondamment la place de Dieu dans l'établissement d'un ordre naturel. De plus, pour Linné, cette classification, ainsi que la forme des espèces, sont fixées, définitives, loin de la vision faisant aujourd'hui consensus. Malgré cela, la 10^{ème} édition du *Systema naturae* (comprenant la description de 9000 plantes et animaux) de 1758 est acceptée par la communauté scientifique comme le point de départ de toute la nomenclature zoologique, et ce jusqu'à nos jours.

2. Introduction des concepts de transformisme, puis d'évolution

Jusqu'alors, il était admis que les organismes possédaient une forme, immuable au cours des générations. Mais avec l'avènement des Lumières vient également une remise en cause globale de la biologie. Ainsi, une multitude de biologistes, dont Lamarck (1744-1829), vont profondément refonder cette discipline. La morale du siècle n'oppose plus autant de limites à une modification de la vision théologique du monde vivant. C'est ainsi qu'avec sa *Philosophie zoologique*, publiée en 1809, il remet en cause le modèle fixiste, en introduisant le transformisme, c'est-à-dire la variation des espèces. Pourtant, il ne tient pas compte des variations inter-individuelles au sein d'une espèce, les considérant comme négligeables. (LECOINTRE et LE GUYADER, 2001). Pour que la théorie de l'évolution ait pu voir le jour, et mettre à bas la notion de transformisme des espèces, il a tout d'abord fallu revenir sur un certain nombre de grands principes qui sous-tendaient toutes les réflexions sur la biologie, avec en premier lieu l'idée de métamorphose, soit la transformation pure et simple d'un organisme adulte en un autre, d'une autre espèce. Une remise en question très progressive de la génération spontanée s'est également amorcée. Il a fallu de nombreuses années pour qu'elle soit définitivement écartée, en dépit des observations d'Antonie Van Leeuwenhoek (1632-1723) qui, après avoir inventé le microscope, pensa observer ce qu'il nomma « animalcules ». Selon la théorie de la génération spontanée, les êtres se formaient sans passage par ce que l'on nomme aujourd'hui fécondation, mais uniquement par présence d'êtres préformés contenus dans des particules. En 1694, Nicolas Hartsoeker crut observer un homme miniature dans les spermatozoïdes. Il élaborait une théorie selon laquelle l'embryon était préformé dans le spermatozoïde, la théorie « animalculiste » (Figure 2).

Figure 2 : L'« homoncule », une des particules sources envisagées pour l'homme, d'après les premières observations de spermatozoïdes au microscope. Gravure de Nicolas Hartsoeker, 1694



(d'après PHILO, 2011).

Il a également fallu tordre le cou à une idée antique, celle de l'échelle des êtres, la *Scala Naturae*, qui classait depuis toujours les êtres selon leur complexité, selon une gradation linéaire, l'homme étant situé au point le plus haut, et les transitions entre les échelons se faisant d'une manière qui serait vue de nos jours comme particulièrement fantaisiste. Enfin, la prise en compte d'une possibilité de temps géologiques beaucoup plus longs que ceux supposés, ainsi que l'étude par Cuvier des fossiles et de leur significations, a permis de faire le lit des futures théories évolutionnistes (LE GUYADER, 2003).

La notion d'espèce en elle-même semble pourtant toujours délicate à définir. Jusqu'au milieu du XX^{ème} siècle, la plupart des biologistes s'en tiennent à la définition donnée par Cuvier : « *L'espèce comprend les individus qui descendent les uns des autres ou de parents communs, et ceux qui se ressemblent autant qu'ils se ressemblent entre eux* », énoncée selon différentes variantes entre 1797 et 1830. Mais avec l'avènement des théories évolutionnistes à la suite des travaux de Darwin (1809-1882) notamment, il semblait délicat de ne définir une espèce que par des ressemblances, sans prendre en compte la variabilité intra-spécifique, inséparable de la notion même d'espèce. La classification des espèces a alors perdu de sa fixité, pour intégrer la théorie synthétique de l'évolution. Ce nouveau paradigme tint compte des travaux de Darwin sur la sélection naturelle, exposés dans son fameux *Origine des espèces* de 1859, mais aussi de la non-hérédité des caractères acquis de Weismann (1834-1914), ainsi que de l'envol de la génétique, grâce à la redécouverte des travaux de Gregor Mendel (1822-1884), et plus particulièrement des mutations. La modélisation de la génétique des populations, à partir des années 1920, joua également un rôle important dans la finalisation de la théorie synthétique de l'évolution. Il s'agissait alors de comprendre les principes sous-tendant la diversité biologique visible, autrement dit le processus d'apparition d'une espèce à partir d'un individu initial, la séparation irréversible de deux lignées évolutives. Cette séparation est effective à partir du moment où un isolement reproductif est installé, ce qui revient à dire que deux individus d'espèces différentes ne peuvent se reproduire entre eux. Ce processus est nommé cladogenèse, ou spéciation (LE GUYADER, 2003).

Le premier à représenter les relations phylogéniques de l'ensemble des êtres vivants par des arbres sera, quelques années seulement après la publication des travaux de Darwin, Ernst Haeckel (1834-1919). La cladistique à proprement parler, théorie de la classification phylogénétique des clades (un clade, dit aussi groupe monophylétique, étant constitué par un ensemble d'êtres vivants et de leur ancêtre commun, qui partagent tous un nouveau caractère), nommée également taxinomie phylogénétique, a vu ses principes énoncés par l'entomologiste allemand Hennig (1913-1976) à partir des années 1950. Il s'agit d'établir les relations de parenté sur la base des homologues (états évolués des caractères), que l'on considère héritées de l'ancêtre par ses descendants (LECOINTRE et LE GUYADER, 2001).

3. Une révolution technologique : l'outil moléculaire

La théorie synthétique de l'évolution, même si elle est perpétuellement renouvelée et améliorée, n'a plus été fondamentalement remise en cause depuis le début du XX^{ème} siècle. Mais de nos jours, et ce depuis près d'un demi-siècle, ce ne sont plus uniquement les ressemblances morphologiques qui sont prises en compte pour la phylogénie, mais principalement les analogies moléculaires. En effet, en 1965, Zuckerkandl et Pauling sont les premiers à avoir défini comme nouveaux caractères comparables entre les espèces les séquences codant les protéines, partant du constat qu'il existe une grande similarité inter-espèces des enchaînements en acides aminés, et qu'elles ne diffèrent entre elles que par quelques positions (LE GUYADER, 1998). Qu'il s'agisse de méthodes immunologiques, d'analyses de protéines, d'hybridation ADN-ADN, d'utilisation d'enzymes de restriction, ou de séquençage nucléotidique, l'outil moléculaire permet une puissance d'analyse et une précision sans commune mesure avec la description des individus, ou taxonomie. De plus, devant les nombreuses limites de cette dernière, Paul Hebert a proposé en 2003 la mise en place d'un outil systématique, la méthode de barcoding (HEBERT *et al.*, 2003). L'approche moléculaire permet en effet d'utiliser les séquences d'ADN comme des codes-barres pour identifier les espèces. Il s'agit de choisir un marqueur moléculaire d'une taille suffisante pour être amplifié, mais ne dépassant pas une trop grande taille (600 à 700 paires de bases), afin de faciliter l'amplification par PCR et le séquençage. Le marqueur choisi par Paul Hebert, et utilisé depuis, est le gène mitochondrial de la première sous-unité de la cytochrome c oxydase, *COI*. Ce projet de standardisation, intitulé "Barcode of Life", présente néanmoins quelques limites. Par exemple, chez les végétaux, le génome mitochondrial évolue plus lentement que le nucléaire, et chez les champignons, il existe des introns au sein du gène *COI*.

En 2000, on comptait en moyenne 7 000 descriptions par an depuis Linné. En 2012, 1,32 million d'animaux et 270 000 plantes étaient décrits, dont 84% en milieu terrestre. Ce ratio peut être considéré comme assez proche du ratio réel, car il existe moins d'obstacles en milieu marin qui permettraient, par isolation de populations, l'apparition de nouvelles espèces. Il est à noter que 1,5% des espèces décrites sont des vertébrés supérieurs, et presque un tiers des insectes. Pour donner quelques chiffres assez généraux, 5 000 espèces de mammifères, 10 000 d'oiseaux, 8 000 de reptiles non aviens et 7 000 amphibiens ont été décrits. S'ajoutent chaque année environ 50 espèces de mammifères pour 100 de reptiles. Cela peut nous amener à dire que nous ne serions pas dans l'ère des mammifères, comme cela pourrait apparaître au premier abord, mais toujours dans l'ère des reptiles (WILSON, 2002 ; DUFF et LAWSON, 2004 ; MACKAY, 2002). Pourtant, même si le rythme de description se situe aujourd'hui autour de 15000 espèces par an (dont 62% d'insectes), selon Wilson, 83% des espèces seraient encore inconnues (WILSON, 2002). La biodiversité réelle semble être de 8 millions d'espèces. Ainsi, il arrive régulièrement, et ce malgré une anthropisation galopante de l'ensemble des milieux terrestres, qu'il en soit découvert de nouvelles. Citons par exemple le Saola, un gros ongulé du Laos-Vietnam, découvert en 1992 ; le Pin de Wollemi, araucaria découvert dans un parc national australien en 1994 ; ou encore le *Virola kwatae* (espèce d'arbre), découvert en 1997 en Guyane dans une forêt domaniale possédant un centre de recherche. Il arrive même que l'on découvre de nouveaux écosystèmes ! Ainsi en 1977, avec la découverte de la première source hydrothermale, 450 espèces de bactéries comprenant des familles, des types anatomiques mais aussi des adaptations physiologiques nouvelles ont été découvertes (IFREMER, 2013).

II. Etude de la diversité du vivant à l'échelle des espèces, et des populations

1. Théories sous-jacentes à la reconstruction des phylogénies

De très nombreuses méthodes existent pour représenter les relations de parenté entre taxons. De nos jours, sont exploitées comme support presque exclusivement les séquences nucléotidiques. En effet, dans une séquence nucléotidique, chaque nucléotide est considéré comme un caractère. Si l'on étudie une séquence de 500 paires de bases, entre divers individus, la comparaison se réalise de manière bien plus pointue que si l'on comparait une dizaine de caractères observables. Bien sûr, il convient de choisir des gènes présentant peu de variabilité intra-spécifique, et une forte variabilité inter-spécifique, afin de permettre une comparaison aisée entre organismes d'espèces différentes. L'étude de la phylogénie est un vaste domaine et quelle que soit la méthode utilisée, des hypothèses très simplificatrices sur l'évolution biologique des séquences sont nécessaires. Avec les méthodes actuelles, pour reconstruire une phylogénie de bonne qualité, c'est-à-dire la plus représentative d'une réalité biologique, la qualité et le nombre des données provoquent plus de variations au sein d'un arbre qu'un changement de méthode. L'outil informatique permet un traitement de très nombreuses données, ainsi qu'une complexification des algorithmes nécessaires à leur réalisation (LECOINTRE et LE GUYADER, 2001).

Le choix du gène sera particulièrement important dans la réalisation d'un arbre phylogénétique. En effet, par l'utilisation des séquences nucléotidiques, on part du postulat que le ou les gènes choisis évolueront de la même manière que l'organisme qui les porte. Ainsi, lors du phénomène de spéciation, deux populations d'une même espèce en viennent à évoluer indépendamment l'une de l'autre, et donneront *in fine* deux espèces. Le même gène, présent dans les deux espèces, acquerra dans chacune de ces espèces des mutations différentes. On appellera ces gènes « gènes orthologues », et ce sont eux qui nous intéresseront. Néanmoins, il est possible, au sein d'une même espèce, qu'un gène en vienne à se dupliquer, une ou plusieurs fois. Ainsi, chaque répétition du gène en viendra à évoluer séparément, et ces gènes, nommés « gènes paralogues » seront alors un « piège » pour la réalisation de phylogénies moléculaires (DARLU et TASY, 1993 ; LE GUYADER, 2003).

Il existe cependant quelques limites à l'utilisation de l'outil moléculaire pour la reconstruction de phylogénies. Tout d'abord, l'emballement premier à l'arrivée de cet outil a été freiné par le coût important du séquençage, même s'il tend à diminuer avec l'évolution des techniques. De plus, des erreurs dues au choix des protéines ou des gènes sont possibles. Citons l'exemple de Graur en 1991 qui a montré que d'après l'analyse de quinze protéines, la souris serait plus proche des primates que du cochon d'Inde, ce qui contredit toutes les autres données morphologiques, paléontologiques, ou générées par l'étude d'autres protéines (GRAUR, 1991). Le problème résultait de l'utilisation d'un nombre trop faible d'espèces. En effet, l'éloignement de certains taxons de leurs groupes frères peut s'expliquer par un **artéfact d'attraction des longues branches**. Ce phénomène peut nous conduire à définir des relations de parenté fausses. Il faut donc prêter une attention particulière à ce type d'artéfacts lors de l'interprétation des reconstructions phylogénétiques (LE GUYADER, 1998).

2. Applications pratiques

On peut globalement séparer les arbres phylogénétiques en trois types, reposant sur trois approches différentes : phénétique, cladistique, ou probabiliste.

La plus ancienne, la systématique phénétique, date des années 1960, et fait suite aux travaux des généticiens des populations ainsi que, principalement, du biologiste Ernst Mayr. Cette systématique est associée à une école de pensée dite « New Systematics », qui a pour but de réunir la spéciation, donc l'analyse de la divergence entre les espèces à l'aide de l'outil génétique, et la systématique. La systématique phénétique consiste à mesurer les ressemblances entre taxons, principalement ceux éloignés les uns des autres, et à les ramener à des distances (MAYR, 1963).

En parallèle à cette doctrine émerge la phylogénie cladistique, pour laquelle on cherche à placer sur les branches les événements évolutifs ayant eu lieu dans chaque lignée, les états dérivés de caractères homologues, en partant du principe que seuls ces caractères permettent de prouver la parenté entre deux taxons, tandis que les caractères plus anciens, communs à des taxons descendants d'un autre ancêtre, n'apportent pas d'information pertinente. Un organisme et tous ses descendants forment un « groupe monophylétique ». La difficulté consiste à définir les états dérivés (ou plésiomorphes) par rapport aux états ancestraux (ou apomorphes), ces derniers étant ceux apparus moins récemment dans l'histoire, ou au cours du développement. De plus, le cladogramme choisi sera celui présentant le nombre de transformations évolutives le plus restreint : on appelle ce dernier critère la parcimonie. Cela permet entre autres d'éviter les analogies (ressemblances dues uniquement à des convergences adaptatives, par exemple les ailes chez les oiseaux et les chauves-souris, nées d'une même adaptation au vol) et de conserver les homologies (TILLIER, 2000).

Enfin, l'approche probabiliste consiste à choisir un modèle d'évolution, constitué d'un ensemble de paramètres θ sur lesquels on peut formuler différentes hypothèses. On cherche, au sein d'un modèle choisi, à faire tendre les paramètres vers leurs valeurs les plus vraisemblables. Plus on possède de données initiales, plus les valeurs estimées pour les paramètres sont probables. On ne pourra pas néanmoins tirer de conclusions quant à la validité du modèle lui-même (DARLU et TASY, 1993).

Il existe de très nombreux algorithmes permettant d'aussi nombreuses réalisations d'arbres. Il peut être pertinent au cours d'une étude d'en tester plusieurs, et de les comparer. Si les résultats sont proches, cela signifiera que ceux-ci sont probablement exacts. Il est également intéressant de choisir la méthode en fonction d'une question spécifique, ainsi que d'incorporer une possibilité de variation dans les taux de mutations / substitutions, afin de ne pas trop encadrer la réalisation de l'arbre.

Enfin, et ce quelle que soit la méthode choisie, il peut se révéler intéressant d'évaluer statistiquement l'arbre obtenu : on fait alors appel à la méthode de bootstrap, inventée par Bradley Efron en 1979. Il s'agit de perturber la matrice de données utilisée pour la création de l'arbre pour tester la robustesse de chaque nœud, en recréant artificiellement plus d'une centaine de « pseudo-arbres » et en les comparant à chaque fois à l'arbre initial. La fréquence à laquelle on retrouve les sous-parties de l'arbre dans l'ensemble de pseudo-arbres est la valeur

de bootstrap. Une valeur de bootstrap élevée attestera d'un embranchement fiable (EFRON, 1979).

3. L'apport de la génétique des populations : principes sous-tendant la biodiversité

Après avoir vu les différentes manières de modéliser les relations de phylogénie entre espèces, il peut être intéressant de se pencher sur le stade antérieur à l'apparition d'une nouvelle espèce, c'est-à-dire la population. Les mêmes outils que ceux vus précédemment pourront être utilisés, en considérant cette fois une population, comme unité taxonomique. En effet, dans cette vision, pour passer d'une génération à une autre, ce sont bien les populations qui se reproduisent à l'identique, en reproduisant la même diversité de génotypes, et non les individus. On considère bien dans ce cas la population comme une unité évolutive. Aucun descendant n'est totalement identique à ses parents, mais les fréquences alléliques au sein d'une population sont bien, elles, identiques de génération en génération. Pour Ernst Mayr, important théoricien de la discipline, la population est un pool génique, donc un ensemble évolutif, et il se caractérise par sa diversité génétique à un instant t , et recombine les allèles entre les individus par la reproduction sexuée. Le terme « génétique des populations » n'est donc pas uniquement synonyme de génétique utilisée au niveau de la population comme on en ferait à n'importe quel autre niveau, mais elle identifie que c'est à ce niveau, précisément, que se détermine l'évolution des espèces (HENRI et GOUYON, 2008).

De plus, la génétique des populations est dotée d'outils divers, nous renseignant particulièrement sur les flux de gènes entre populations. Il est alors possible de connaître, à l'aide de tels estimateurs, le taux de différenciation entre deux populations. Il est aussi possible de connaître l'évolution démographique d'une population. Les tests dits de « neutralité sélective », comme le F_s de Fu (FU, 1997) ou le D de Tajima (TAJIMA, 1989), auront pour but de comparer certaines quantités observables (nombre d'allèles de l'échantillon, etc.) à celles que l'on devrait observer en l'absence de sélection. Si la différence est significative, il sera alors possible de s'orienter vers une hypothèse de sélection du gène, ou bien si l'on écarte cette hypothèse, vers une modification démographique de la population, qu'il s'agisse d'une expansion ou d'un goulot d'étranglement (réduction de la **taille efficace** de la population).

III. *Forficula auricularia*, organisme modèle pour l'étude de la spéciation

1. Présentation générale de l'objet d'étude

Originaire d'Europe, d'Asie de l'est et d'Afrique du nord, bien qu'il ait également été introduit en Océanie et en Amérique du nord, le perce-oreille européen, *Forficula auricularia* (Figure 3), est une espèce d'insecte dermaptère (du grec *derma* (peau) et *pteron* (aile)) subsociale omnivore, consommant de nombreuses plantes et animaux. Il présente une grande capacité de dispersion. On sait notamment qu'observé pour la première fois sur le sol américain en 1907 à Seattle, il lui a fallu une vingtaine d'années seulement pour coloniser les états du nord ainsi que le Canada.

Figure 3 : Un individu mâle du perce-oreille européen, *Forficula auricularia*



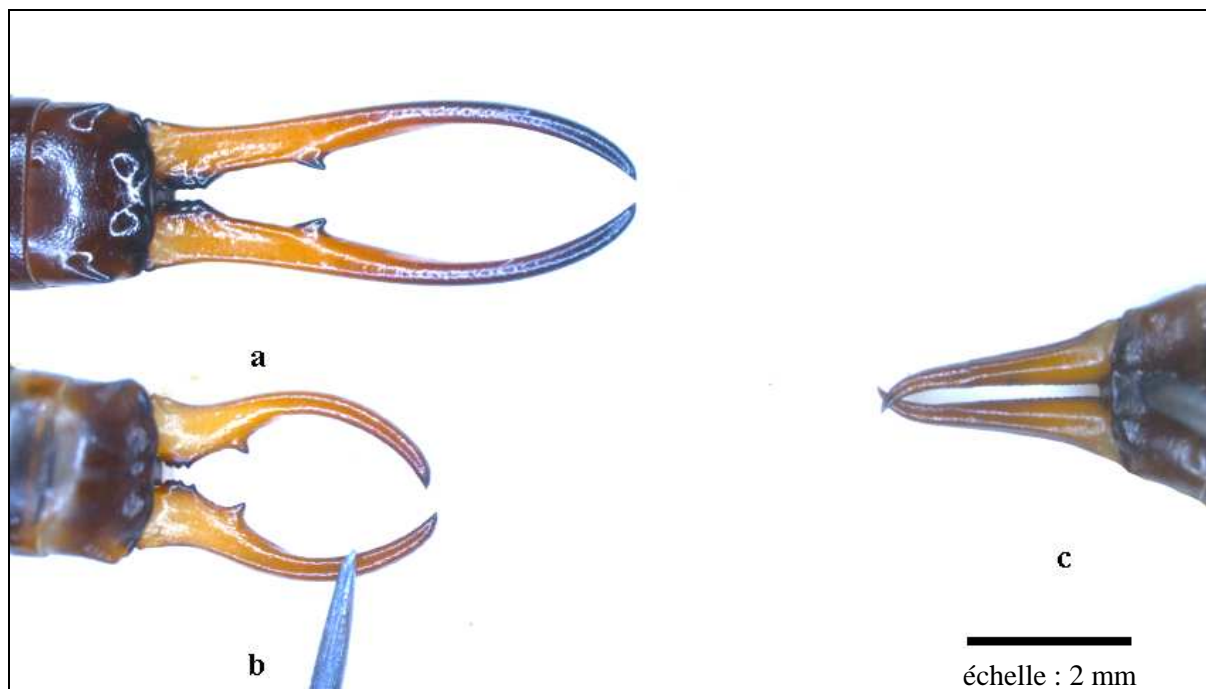
(source ABCGP, 2013)

Le temps de génération de l'espèce est d'un an. L'insecte passe au cours de son développement par quatre stades nymphaux, en se colorant progressivement pour passer de blanc presque translucide à brun foncé à rougeâtre sous sa forme adulte. Les ailes sont cachées sous les élytres, dépassant seulement sur une petite partie de leur surface. Sans poser de problèmes agronomiques majeurs, il est parfois considéré comme un ravageur des cultures (CAPINERA, 2008). Il présente diverses particularités qui justifient son utilisation comme modèle biologique. En effet, la pince du perce-oreille européen, ou forceps, présente un dimorphisme sexuel, mais aussi une bimodalité de taille chez les mâles. De plus, ces insectes présentent, fait original, un instinct maternel. En effet, le perce-oreille européen possède cette particularité, particulièrement rare parmi les insectes (si l'on exclut les insectes sociaux comme les fourmis). Les larves élevées par leur mère montrent une plus grande probabilité de survie que celles qui ne l'ont pas été. Il semble que ce bénéfice soit médié entre autres par l'apport de nourriture de la mère aux petits (KOLLIKER 2007). Néanmoins, il semblerait qu'en cas de restriction alimentaire, une compétition mère-progéniture puisse diminuer nettement l'espérance de survie du couvain, montrant ainsi les limites de l'effet bénéfique de l'instinct maternel dans cette espèce (MEUNIER et KOLLIKER, 2012). On peut également observer que les mères nettoient les oeufs avec leur « bouche », geste qui semble avoir pour but de les protéger des pathogènes (GINGRAS et TOURNEUR, 2001).

2. Particularité de l'organe caudal, le forceps

Forficula auricularia présente une caractéristique rare du point de vue de l'évolution morphologique : la présence, en plus du dimorphisme sexuel, de deux types de mâles, mise en évidence entre la fin du XIX^{ème} siècle et le début du XX^{ème} siècle (TOMKINS et SIMMONS, 1996) (Figure 4). Cette singularité dans le monde animal se serait conservée en tant qu'adaptation utilisée dans la compétition sexuelle (TOMKINS, 1999 ; TOMKINS et BROWN, 2004).

Figure 4 : Comparaison des pinces chez les deux types de mâles et les femelles.



Les pinces, ou forceps, sont différentes chez les mâles et les femelles, mais aussi entre les mâles. On peut voir ici, mises à la même échelle, les trois types de pinces dans une même population. a : forceps de mâle macrolabia, b : forceps de mâle microlabia, c : femelle. (photographies personnelles).

Les premiers entomologistes se sont penchés sur le rôle possible des forceps au cours du XIV^{ème} siècle, sans poser de réelle hypothèse évolutive, si ce n'est que cet organe était possiblement une arme, sélectionnée pour la force qu'elle conférait aux forficules. En 1892, Bateson et Brindley ont évoqué pour la première fois le dimorphisme des forceps de *Forficula auricularia* : il existe en effet au sein d'une même population deux formes de pinces chez les mâles, grande ou petite. Par une analyse comparée à celle des cornes du coléoptère *Xylotrupes*, ils montrèrent que deux formes de forceps étaient possibles pour une taille de corps fixée (BATESON et BRINDLEY, 1892). Des lois mathématiques régissant les variations morphologiques de ce type d'organes (cornes, forceps) ont été proposées au début du XX^{ème} siècle (HUXLEY, 1924). Mais il fut également avancé que les variations portant sur

les forceps n'étaient probablement pas d'origine génétique (DIAKONOV, 1925). Cette notion s'est précisée en 1927, avec Huxley, qui montra et formula mathématiquement la corrélation entre la taille du corps et celle des forceps. De plus, bien que dans presque toutes les classes de tailles corporelles le dimorphisme au niveau de la longueur des forceps semble évident, il existe des classes qui ne présentent pas cette bimodalité : dans ce cas, les corps de grande taille correspondent à de longs forceps, et les corps de petite taille à des forceps court (HUXLEY, 1927). Enfin, il existe une importante corrélation taille du corps/taille des forceps, que le corps soit long ou court. L'hypothèse suivante a alors été posée : ce sont les conditions environnementales qui, influençant la taille corporelle, influencent la taille des forceps. Ces études empiriques touchaient donc du doigt la notion de **plasticité phénotypique**.

Des années plus tard, en 1991, Eberhard et Gutierrez réfutèrent cette hypothèse : selon leurs études, à partir d'un approfondissement purement descriptif (la grande flexibilité du dimorphisme des forceps), ils s'opposèrent à l'hypothèse selon laquelle ce dimorphisme résulterait de contraintes extérieures lors du développement (EBERHARD et GUTIERREZ, 1991). Les termes "brachylabia/macrolabia" ont alors été introduits pour désigner les deux modes de morphologie des mâles perce-oreilles, les mâles brachylabia présentant des forceps plus courts que les macrolabia (TOMKINS et SIMMONS, 1996). Pendant des années, l'idée a perduré que les forceps seraient uniquement des armes. Mais un tout nouveau paradigme est apparu à la fin des années 90, à partir de nouvelles hypothèses inductives et d'observations visant à les conforter : ce serait en fait la sélection sexuelle qui jouerait un rôle majeur sur la taille des forceps, les plus grands conférant un avantage notable à leur porteur. Le choix du partenaire par les femelles serait influencé par la taille des forceps des mâles (TOMKINS et SIMMONS, 1998). De plus, la taille des forceps des mâles semblait hautement conditions-dépendante, les macrolabia apparaissant en nombre plus important lorsque le régime alimentaire était haut en protéines (TOMKINS, 1999). En ce cas, la question se pose de savoir pourquoi de petits forceps continuent à coexister avec les grands ? Ce sont les travaux de Forslund (FORSLUND, 2000) qui apporteront une cohérence entre les deux idées précédentes : ils semblent valider la thèse d'une coévolution conditionnelle. Par des mesures de **fitness**, ou potentiel évolutif, ces travaux montrent que différents régimes de sélection agissent sur les deux modes (brachylabia et macrolabia). Forslund a alors posé l'hypothèse, grâce à la confrontation avec des études précédentes, que le dimorphisme des mâles était une stratégie conditionnelle qui avait évolué sous sélection, les deux modes conférant deux types d'avantages différents. Une étude de 2004 a démontré que le dimorphisme des mâles au sein de l'espèce *Forficula auricularia* répondait à une **ESS**. En effet, dans le cas des perce-oreilles, l'évolution a favorisé au sein du sexe mâle deux stratégies reproductives parallèles, les individus adoptant la tactique capable de générer le meilleur résultat en termes de fitness. Les mâles macrolabia ont plus facilement accès aux femelles, tandis que les brachylabia doivent ruser pour y accéder. A l'inverse, il est moins coûteux énergétiquement de développer de petits forceps. Les auteurs ont montré que le point clé médiant le monomorphisme ou le dimorphisme des mâles était la densité de population (TOMKINS et BROWN, 2004).

Est ensuite venue une réflexion sur la possibilité d'un lien entre la taille des ornements sexuels, aussi utilisés comme arme, et l'immunité (RANTALA *et al.*, 2007). En effet, les forceps de mâles sont utilisés comme arme dans les combats pour l'accès aux femelles, ce qui montre une influence directe de la taille de cet organe sur la fitness, mais il a aussi été suggéré

une influence indirecte, par un choix du partenaire par la femelle sur la base de la taille des forceps. On peut donc supposer qu'il existe un avantage évolutif à posséder des forceps de grande taille, autre que la puissance lors des combats pour l'accès aux femelles. L'étude testait l'hypothèse d'une corrélation entre la taille des forceps et la fonction immunitaire, ainsi qu'une différence de cette immunité entre les sexes. Aucune corrélation avec la taille des forceps n'a alors été trouvée, quel que soit le sexe. Néanmoins, la fonction immunitaire semblait négativement corrélée avec la longueur du corps.

Enfin, une étude récente (MUNOZ et ZINK, 2012) a montré que dans l'espèce de perce-oreille marin *Anisolabis maritima*, les combats entre mâles de taille importante menaient majoritairement à la victoire de l'individu le plus gros, alors que lors de combats entre individus de taille réduite, l'asymétrie des armes constituées par les forceps avantageait son porteur. Les auteurs concluent donc en la possibilité de voir dans l'asymétrie un facteur indirect de supériorité au combat.

3. Dynamique des populations

Lors de diverses études de l'instinct maternel chez les forficules, dans les années 1980, des élevages ont été mis en place. On s'est alors rendu compte que certains des descendants n'étaient pas inter-reproductibles : il s'agissait de deux espèces jumelles, au sens de Mayr (appartiennent à la même espèce deux individus reproductibles et dont la descendance est fertile). Des analyses génétiques ont par la suite cherché à confirmer cette hypothèse. La caractérisation précise date de 1997 : des analyses génétiques prouvèrent ce qui était présumé, qu'il existait bien deux espèces distinctes de *Forficula auricularia*, morphologiquement identiques. L'équipe de Michel Veuille et Thierry Wirth (du laboratoire EPHE Biologie intégrative des populations, UMR MNHN/CNRS 7205 Origine structure et évolution de la biodiversité) a caractérisé précisément les deux espèces, et donné des hypothèses sur leur expansion en Amérique du Nord (WIRTH *et al.*, 1998). En 1999, cette même équipe, après séquençage d'un gène mitochondrial de la cytochrome oxydase II (*COII*) pour six espèces d'insectes dermoptères a remarqué que le taux de bases CG était étonnement élevé dans le genre *Forficula*, par rapport aux autres insectes, particulièrement sur les 3^{ème} bases des codons. Ils montrèrent qu'il y avait eu un changement radical dans la composition des bases au cours de la spéciation de la famille des Forficulidae (WIRTH *et al.*, 1999). Une nouvelle idée vit le jour quelque temps plus tard, lorsque Guillet et ses collègues observèrent un fait nouveau : les deux espèces pouvaient coexister à une même altitude. Les deux espèces étaient donc **sympatriques** (GUILLET *et al.*, 2003).

En expérimentation au laboratoire, il est possible de réaliser des hybrides F1 (première génération) entre les deux espèces, mais avec un très faible taux de succès reproducteur, et il est apparemment impossible d'obtenir des F2 (deuxième génération) (WIRTH *et al.*, 1998). La présence probable d'une **barrière postzygotique** laisse croire que la zone de contact pyrénéenne est probablement une zone de sympatrie entre deux espèces génétiquement différenciées. Cette étude portait sur neuf lignées d'élevage issues de plusieurs régions d'Europe. Elle suggérait des différences de distribution spatiale, et peut-être des différences d'adaptation à un climat plus ou moins froid. Ces deux espèces n'ont pas encore de nom linnéen, leur morphologie étant semblable, ce qui n'a pas encore permis de décrire séparément

les deux espèces. Elles cohabitent en Europe, et particulièrement en France. L'une, nommée provisoirement A, est sémelpare (une ponte par an), excepté dans la zone Méditerranéenne où elle peut boucler deux pontes, et vit principalement en altitude ou en milieu continental. La seconde, nommée provisoirement B, est itéropare (deux pontes par an) et vit majoritairement à basse altitude et dans des habitats océaniques. Un caractère possible d'adaptation de ces espèces au climat pourrait être l'existence d'une ou deux pontes annuelles, ce qui suggère une période de reproduction de durée variable pendant le cycle annuel. Le comportement de soins maternels de ces espèces est une autre singularité notable de l'adaptation de cet organisme au milieu. En effet, les larves élevées par leur mère dans des portées réduites montrent une plus grande probabilité de survie que celles qui ne l'ont pas été (KOLLIKER, 2007). Enfin, une donnée à prendre en compte dans la distribution spatiale des espèces européennes est l'alternance des cycles glaciaires.

DEUXIEME PARTIE : Travail personnel

I. Introduction

Dans une démarche d'étude d'un phénomène de spéciation, nous avons utilisé le modèle biologique constitué par l'insecte *Forficula auricularia* (ou perce-oreille), présentant une scission en deux espèces jumelles, isolées génétiquement, reconnaissables par l'ADN (acide désoxyribonucléique), mais morphologiquement identiques. L'espèce A, une portée par an (deux en Méditerranée), vit majoritairement en altitude ou en milieu continental, alors que l'espèce B a deux portées par an, un habitat de basse altitude ou océanique. Nous avons effectué des analyses conjointes de génétique des populations à l'aide de deux marqueurs mitochondriaux, et de morphologie par étude de contour. Ces travaux ont été effectués dans le laboratoire EPHE Biologie intégrative des populations, UMR MNHN/CNRS 7205 Origine structure et évolution de la biodiversité, sur une période de six mois étalée sur l'année scolaire 2012/2013.

II. Matériels et méthodes

1. Collection et identification des individus

Pour cette étude, nous avons considéré la dispersion de *Forficula auricularia* en Europe selon un échantillonnage large de 43 populations (15 échantillons d'une vingtaine d'individus (ANNEXE 2), et 28 échantillons de quelques individus) (Tableau 1).

Tableau 1 : Correspondance des codes, des nombres d'individus séquencés (N) pour les deux marqueurs, COI et COII, et des localisations, pour notre échantillonnage.

Code	Localisation	N	Code	Localisation	N
BEH	Behorleguy (France)	20	A20	Autoroute A20 Périgord (France)	1
BEL	Quiévrain (Belgique)	6	ALE	Allevard (France)	1
CARD	Cardroc (France)	12	ARC	Bassin d'Arcachon (France)	2
DAN	Holløse, Sjælland (Danemark)	12	ARO	Arouca (Portugal)	1
FER	Mykines (îles Feroe)	17	BAL	Bâle (Suisse)	1
GLI	Hôtel de Glières (France)	23	BAS	Bassaburua (France)	1
GYO	Győr (Hongrie)	4	BRE	Bréhat (France)	2
IRA	Iraty (France)	19	BRI	Briare (France)	1
JDP	Jardin des Plantes, Paris (France)	20	BUN	Bungsberg, Holstein (Allemagne)	1
MIJ1	Mijanes 1800m (France)	15	CBI	Chalets de bise (France)	1
MIJ2	Mijanes 1100m (France)	14	CHI	Chiemsee (Allemagne)	1
PEL	Pelasgia, Attique (Grèce)	10	GAM	Les Gambes (France)	1
PFA	Pfastatt (France)	16	GRO	Groisy (France)	1
ROM	Thermes de Caracalla (Rome)	14	LAR	Lardy (France)	1
SUE2	Standtorp, Öland (Suède)	17	NAP	Capodimonte, Naples (Italie)	1
TU	Istanbul, (Turquie)	16	NOU	Nouâtre (France)	1
			PAV	Pavie (France)	1
			PLA	La Planchette (France)	1
			PRA	Prague (Rep. Tchèque)	1
			RAR	Raraue Rästplatz, Aachen (Allemagne)	3
			STD	St Désirat (France)	1
			SUE1	Gråborg, Öland (Suède)	2

En violet, les échantillons de taille importante, en jaune les échantillons utilisés uniquement à des fins biogéographiques.

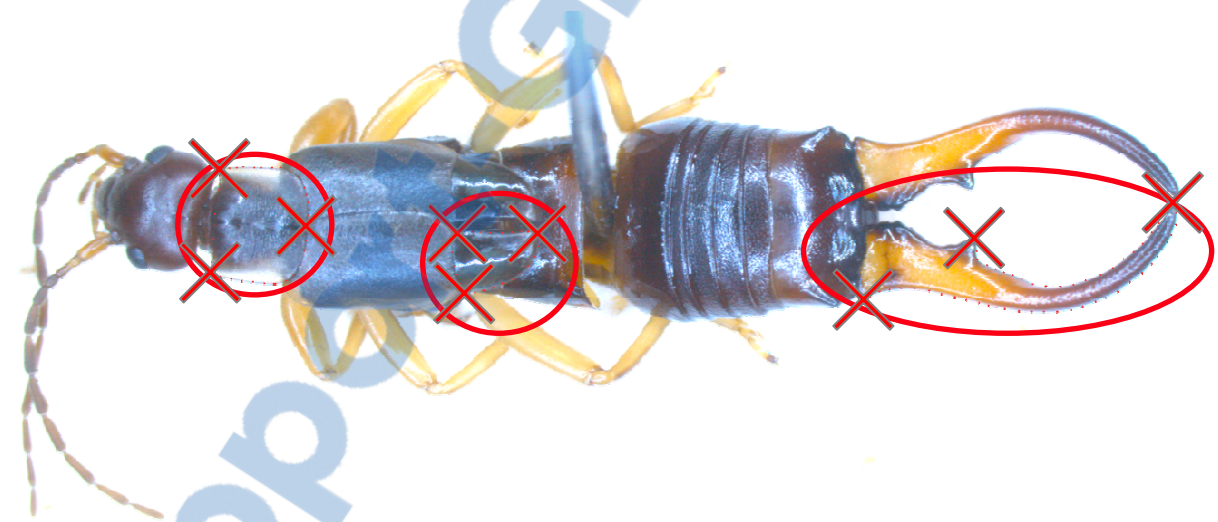
Une attention particulière a été prêté aux régions alpine et pyrénéenne, lieux intéressants qui semblaient, d'après les études antérieures (GUILLET *et al.*, 2003), combiner la double frontière est/ouest et altitudinale entre les deux espèces. L'échantillonnage a été réalisé de manière à pouvoir obtenir des individus des deux espèces décrites. Le nombre d'individus par population correspond d'une part à une volonté de localiser les deux clades, mais aussi de réaliser des analyses précises de génétique des populations, nécessitant un nombre suffisamment élevé d'individus par échantillon. Dans la mesure du possible, nous avons collecté des adultes mâles et femelles, mais aussi des juvéniles (non comptabilisés dans l'étude morphométrique). Les spécimens ont été conservés individuellement dans une solution TE/Eth (Tris EDTA 30% / Ethanol 70%) à -20°C.

2. Protocole morphogéométrique

Pour cette partie de l'étude, nous n'avons travaillé que sur des individus préalablement classés au sein du clade A, B ou C en fonction des critères génétiques (séquençage Sanger et identification des deux principaux clades). Notre échantillonnage comprenait 160 individus répartis entre 12 populations (6 de A, 5 de B et 1 de C) d'une dizaine d'individus chacune. Cette répartition avait pour but de nous permettre d'étudier la différence morphologique entre espèces jumelles, sans se focaliser sur la différence entre localités.

Chaque individu a été photographié selon deux vues dorsales, (individu total et forceps) à l'aide d'un microscope binoculaire (LEICA Microsystems CMS GmbH) puis analysé grâce à la Leica Application Suite (LEICA MICROSYSTEMS, 2003-2009). Une étude indépendante de trois structures (forceps gauche, pronotum, dépassement de l'aile postérieure gauche de l'élytre correspondante), par la méthode des « sliding landmarks » a été réalisée (Figure 5), à l'aide des logiciels tpsDig, tpsUtility et tpsRelw (ROHLF, 2013). Cette méthode correspondait à la mise en place de points anatomiques (« landmarks »), et de point glissants (« sliding landmarks ») sur les structures de tous les insectes.

Figure 5 : Localisation des structures étudiées en morphogéométrie, sur un individu mâle.



Structures anatomiques entourées de rouge, dans l'ordre crânio-caudal : le pronotum, le dépassement de l'aile gauche de l'élytre, le forceps gauche. Chaque structure est étudiée séparément, et son contour est tracé par des points glissants. Les croix rouges représentent les localisations des points anatomiques, qui permettent l'ancrage des points glissants (de 76 à 136 points par structure). (Photographie personnelle).

L'utilisation de *sliding landmarks* permet de décrire des zones d'intérêt biologique qui ne possèdent pas de points de repère anatomique. Cela permet d'étudier les écarts à l'homologie géométrique entre tous les objets. En effet, une même structure peut avoir chez deux individus une forme comparable sans que l'on puisse localiser des points coïncidant sur les deux individus aisément positionnables. Une mise à la même échelle des relevés de formes obtenus, appelée superposition Procruste, permet de comparer les structures des individus.

Les opérations de ces superpositions consistent à réaliser une mise à l'échelle des structures, leurs translations, leurs inversions éventuelles et leurs rotations. Dans un cadre de discrimination, la méthode de minimisation des moindres carrés apparaissait la plus adaptée : la distance entre chaque *semi-landmarks* et le *semi-landmarks* correspondant du consensus a été minimisée. Pour chacun des trente points de mesure d'un individu donné sur une structure donnée, la mesure associée correspond à l'écart avec le point moyen sur le clade considéré (PEREZ *et al.*, 2006). Il nous est parfois apparu une asymétrie des structures étudiées. C'est la raison pour laquelle nous avons mesuré uniquement l'aile gauche et la moitié gauche de la pince, notre étude ne portant pas sur les asymétries.

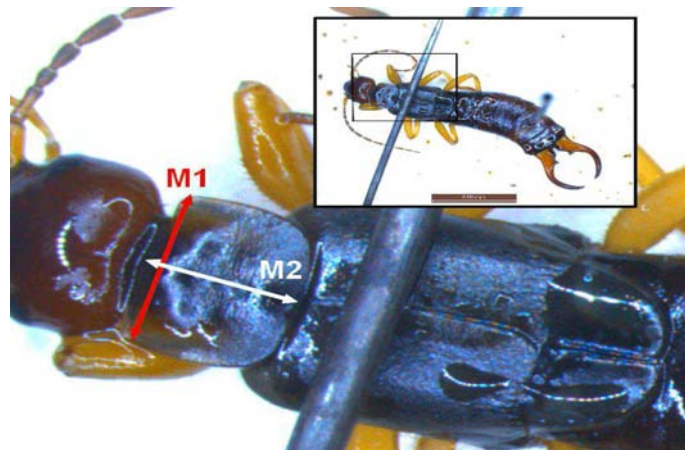
Après superposition Procruste des structures, nous avons effectué des analyses statistiques afin de déterminer si les clades étudiés pouvaient être discriminés par les données morphogéométriques. La forme globale de la structure est composée de la taille d'une part et de la conformation d'autre part. La taille est calculée sous la forme d'une mesure par individu et par structure, grâce au logiciel TPS, et la conformation est représentée par les valeurs numériques des points (trente à quarante) définissant la structure. Trois analyses ont été effectuées pour chaque structure, sous R (R DEVELOPMENT CORE TEAM, 2013) :

- Une analyse en composantes principales (**ACP**), sur l'ensemble des individus dans l'espace à n dimensions correspondant aux n valeurs de mesures de morphométrie effectuée sur chaque individu (n différent selon la structure étudiée). Nous avons observé la répartition de nos individus sur les deux axes principaux, sans a priori de groupe, pour visualiser et quantifier la variabilité en résumant les données en un nombre limité de facteurs principaux ;
- Un calcul de **MANOVA** (ANOVA multivariée), sur les données morphométriques des clades considérés deux à deux, afin de tester l'hypothèse d'homogénéité de leurs configurations morphologiques (cela équivaut au test du T2 de Hotelling, généralisation du test du t de Student à un cas multivarié) ;
- Un test du t de Student sur le centroïde (une seule valeur numérique pour chaque individu, calculée par le logiciel TPS à partir des n mesures qui lui correspondaient), pour tester l'hypothèse d'égalité ou non de la taille de la structure étudiée selon les clades considérés deux à deux.

3. Tests préalables

Afin de ne pas biaiser les résultats par une variabilité trop importante dans la prise de photographies, nous avons pris grand soin de fixer tous les individus de la même façon, à l'aide d'une aiguille plantée dans l'abdomen et d'une autre maintenant la partie crâniale de l'individu le plus à l'horizontale possible. De plus, en coopération avec un stagiaire encadré pendant deux semaines (Thomas Curchod, étudiant en seconde année de l'ENVA), nous avons pu tester l'influence de l'expérimentateur au niveau de la prise de photographies et de la prise de mesure, sur un protocole simplifié (mesure de deux longueurs) et un échantillon réduit (4 populations de 20 individus chacune). Comme ce test devait être réalisé dans un laps de temps assez court, nous avons voulu étudier la variabilité sur la structure qui nous semblait la plus délicate à positionner, le pronotum, dont les mesures sont potentiellement affectées par des problèmes de parallaxe (Figure 6).

Figure 6 : Positionnement des longueurs étudiées, sur le pronotum, pour le test simplifié.



(Photographie et montage de Thomas Curchod).

Les résultats de cette pré-analyse nous ont montré une variabilité significative lors de la prise de mesure sur une photographie par différents expérimentateurs, excluant la possibilité de rajouter des analyses à une étude antérieure sans reprendre entièrement les mesures. Cependant, aucune variabilité significative au seuil de 5% lors de la prise de mesure par un expérimentateur sur des photographies réalisées par différentes personnes n'a pu être détectée. Cette dernière observation validait en première approximation le fait qu'il soit possible de réaliser des mesures sur des photographies prises au cours d'études antérieures, ce qui était le cas pour ce stage. A partir de ces observations, l'effet « expérimentateur ayant pris la photographie » a été pris en compte lors de l'analyse des résultats comme un biais possible, même minime.

De plus, un test de la méthode d'analyse en composantes principales sur des *sliding landmarks* positionnés sur les forceps a été réalisé préalablement à toute mesure, validant la méthode de ségrégation par l'observation d'une plus forte variabilité inter-individus qu'intra-individu (réalisé sur 10 individus photographiés 10 fois et mesurés 10 fois chacun).

4. Extraction d'ADN, amplification et séquençage de deux gènes mitochondriaux

4.1 Extraction d'ADN

L'extraction d'ADN total a été réalisée à l'aide du kit QIAGEN® Blood & Cell Culture DNA Mini Kit (Qiagen, Courtaboeuf, France). Cette extraction comprenait plusieurs étapes. Il s'agissait tout d'abord de disséquer l'individu pour ne conserver que le thorax, et ce afin d'optimiser l'extraction (les autres parties du corps contenant de la chitine à trop forte dose et des graisses), puis de lyser les cellules à l'aide de la protéinase K, laissée au contact du thorax pendant une nuit à 56°C. Ensuite, de fixer l'ADN sur la membrane d'une colonne. Le jour suivant, l'ADN était précipité par utilisation successive de deux tampons de nettoyage suivie d'une re-suspension dans un tampon d'éluion.

La quantification globale de l'ADN extrait a été effectuée par spectrophotométrie. Le ratio de l'absorbance (A) à deux longueurs d'ondes : A260nm/A280nm, permettait une estimation de la pureté de l'ADN. Si ce ratio était inférieur à 1,7, on pouvait considérer l'échantillon comme contaminé par des protéines. Au-delà de 2, de l'ARN contaminait l'échantillon. Entre ces deux valeurs, l'ADN a été considéré comme pur.

Les deux gènes considérés étaient tous deux mitochondriaux. Nous les désignerons dans la suite de ce rapport comme *COI* et *COII*, bien qu'il s'agisse plus précisément d'un fragment de 657 paires de bases chevauchant le gène de la cytochrome oxydase I, et d'un fragment de 493 paires de bases chevauchant les gènes de la cytochrome oxydase I, du tRNA Leu et de la cytochrome oxydase II. Ces fragments ne se chevauchaient pas au niveau de la région *COI*.

4.2 Amplification par PCR de la région 5' de *COI*

La polymérisation en chaîne (PCR), effectuée selon un protocole détaillé (Tableau 2), nous a permis par une série de cycles d'amplifier nos fragments d'ADN. Un cycle correspond à une dénaturation de l'ADN à haute température (94°C) afin de briser les liaisons hydrogènes entre les bases, puis une phase d'hybridation pendant laquelle les amorces se fixent aux deux extrémités du fragment d'intérêt, et enfin une phase d'élongation à 72°C pendant laquelle les brins complémentaires vont être synthétisés grâce à la Taq polymérase DreamTaq (Thermo Electron S.A.S., Courtaboeuf, France) à la suite des amorces (Figure 7).

Tableau 2 : Protocole de PCR utilisé pour nos deux gènes d'intérêt.

	Etape 1	Etape 2			Etape 3			Etape 4
Nombre de cycles	1	5			30			1
Temps	5min	30sec	1min	1min	30s	1min	1min	5min
Température (°C)	94	94	49	72	94	51	72	72

Figure 7 : Amorces utilisées pour amplifier nos deux gènes d'intérêt.
(Eurofins MWG Operon, Ebersberg, Allemagne).

CO1H : 5'-TAAACTTCAGGGTGACCAAAAAATCA-3' ;

CO1L : 5'-GGTCAACAAATCATAAAGATAT'TGG -3' , pour la région *COI* ;

CO2forf4-F : 5'-TGGTTGGGATTCTTTTATTG-3' ;

CO2forf5-R : 5'-TCATCTAAAAGGTACAACAA-3' pour la région *COII*.

5. Analyse des séquences et construction d'arbres phylogénétiques

Le séquençage a été réalisé au Génoscope (Evry, France) pour le gène *COII*, et par la compagnie Eurofins MWG Operon (Ebersberg, Allemagne) pour le gène *COI*. Les logiciels BioNumerics version 7.1 (APPLIED MATHS NV, 2011) et MEGA version 5 (TAMURA *et al.*, 2011) ont été utilisés pour le nettoyage, la vérification et l'alignement des séquences à partir des chromatogrammes.

La construction d'arbres phylogénétiques a été effectuée à l'aide du logiciel MEGA. Il existe divers types d'arbres phylogénétiques pour la représentation des relations de parenté. Nous avons utilisé pour cette étude la méthode de *Neighbor Joining* (NJ), pour laquelle les séquences sont comparées et réunies par paires de manière à minimiser la longueur totale de l'arbre. Plus deux séquences homologues sont différentes, plus la distance évolutive qui les sépare est grande, et donc plus les séquences ont divergé il y a longtemps. A l'inverse, plus deux séquences homologues sont proches, plus la distance évolutive qui les sépare est faible (SAITOU et NEI, 1987). De plus, la méthode de *Bootstrap* permet, en perturbant la matrice de données, de tester la robustesse des nœuds et donc la validité statistiques des clades.

Le logiciel jModelTest (POSADA, 2008) a été utilisé pour choisir le modèle d'évolution le plus vraisemblable en fonction du jeu de données. Le modèle sélectionné était le TPM2mf+G sous les trois modèles d'évaluation (critères d'information d'Akaike et Bayésien (AIC and BIC), ainsi qu'une méthode de théorie de décision).

Nous avons enfin utilisé le serveur BLAST (*Basic Local Alignment Search Tool*) sur le site de NCBI (<http://www.ncbi.nlm.nih.gov/>), qui compare une séquence à celles de la base de données. Cette méthode nous a permis d'obtenir le pourcentage de similarité sur la longueur de la comparaison, entre nos échantillons et ceux de la base, la probabilité que cette similarité soit due au hasard, ainsi que le pourcentage de recouvrement des deux séquences. Nous avons donc pu vérifier que nos individus étaient bien des *Forficula auricularia*, mais aussi comparer nos séquences avec celles d'individus éloignés géographiquement, séquencés dans d'autres études.

Enfin, la datation du plus récent ancêtre commun (TMRCA) a été réalisée via les logiciels BEAUty, BEAST et TRACER (DRUMMOND *et al.*, 2012). Ces logiciels limitant la possibilité du modèle d'évolution, nous avons utilisé le modèle HKY + I + G (modèle d'Hasegawa, Kishino et Yano, sites invariants, distribution gamma) (HASEGAWA *et al.*, 1985). Lorsqu'un taux de mutation était nécessaire dans le cadre des estimations de TMRCA

et de taille de populations, nous avons utilisé celui du gène *COI* des insectes, soit $\mu = 3.54.10^{-8}$ (PAPADOPOULOU *et al.*, 2010).

Afin de visualiser l'histoire démographique de notre échantillon, nous avons utilisé les *skylines*. Ces graphiques, réalisés tout comme le TMRCA par les logiciels BEAUty, BEAST et TRACER (DRUMMOND *et al.*, 2012), sont basés sur l'utilisation des **probabilités bayésiennes**. Ils permettent de déterminer les fluctuations de la **taille effective** des deux clades en fonction du temps. Il est possible de tester différents modèles démographiques, ainsi que de fixer ou non une **horloge moléculaire**, et de retenir au final les résultats les plus vraisemblables.

Les divers indices de diversité et de structuration des populations ont été calculés à l'aide du logiciel DnaSP (LIBRADO et ROZAS, 2009). Les populations ont été isolées et comparées deux à deux pour plusieurs tests, de divergence nucléotidique entre les populations, de flux de gènes et de différenciation génétique. Les tests de neutralité, D de Tajima et F_s de Fu, ont été calculés pour les populations présentant un polymorphisme nucléotidique. Pour chaque population de taille importante (plus de dix individus), nous avons évalué les paramètres suivants : le nombre d'**haplotypes** (h), l'estimation de la diversité haplotypique (H_d), le nombre de singletons, le nombre de sites polymorphes (S), l'indice de diversité nucléotidique (Π) correspondant au nombre moyen de substitutions nucléotidiques par site entre deux séquences, le nombre moyen de différences nucléotidiques (k).

Le flux génétique entre les populations, c'est-à-dire la mesure des échanges de gènes possibles entre des populations géographiquement différentes, a été mesuré à l'aide du **Fst** par DnaSP, avec un algorithme basé sur les nucléotides (HUDSON *et al.*, 1992a). De plus, une correction de Bonferroni (correction du seuil de significativité en cas de comparaisons multiples qui compense le fait qu'un test parmi les tests multiples peut apparaître significatif par simple fait du hasard) a été appliquée aux valeurs de significativité du Fst.

Enfin, les graphiques ont été mis au point à l'aide de The Gimp© (KIMBALL, 2013) et Inkscape© (FREE SOFTWARE FOUNDATION, 2013).

III. Résultats

1. Résultats préliminaires

Une partie de nos séquences ainsi que des photographies utilisées proviennent d'une étude antérieure (Master d'Aurore Maillard (2012), effectué sous la direction de Michel Veuille et Thierry Wirth). Cette étude, restreinte à 8 populations, à un gène (*COI*), a montré pour les deux espèces mises en évidence en 1998 que l'une (A) était beaucoup plus polymorphe que l'autre (B). Ce polymorphisme impliquait un TMRCA (âge de l'ancêtre commun le plus récent) supérieur pour l'espèce A que pour l'espèce B. De plus, la variabilité des forceps, montrée par des mesures de longueurs de différentes parties du corps de l'individu, restait considérable chez les mâles, et semblait se comporter de la même façon dans les deux espèces.

2. Localisation des deux espèces en Europe

Contrairement à notre hypothèse de départ, les résultats obtenus n'ont pas montré de répartition est-ouest tranchée des deux espèces, avec les Alpes comme frontière. L'espèce A semblait cependant être plus fréquente en Europe orientale, le clade B n'ayant pas été retrouvé à l'est de Prague (Figure 8).

Figure 8 : Répartition des deux espèces parmi les populations échantillonnées.

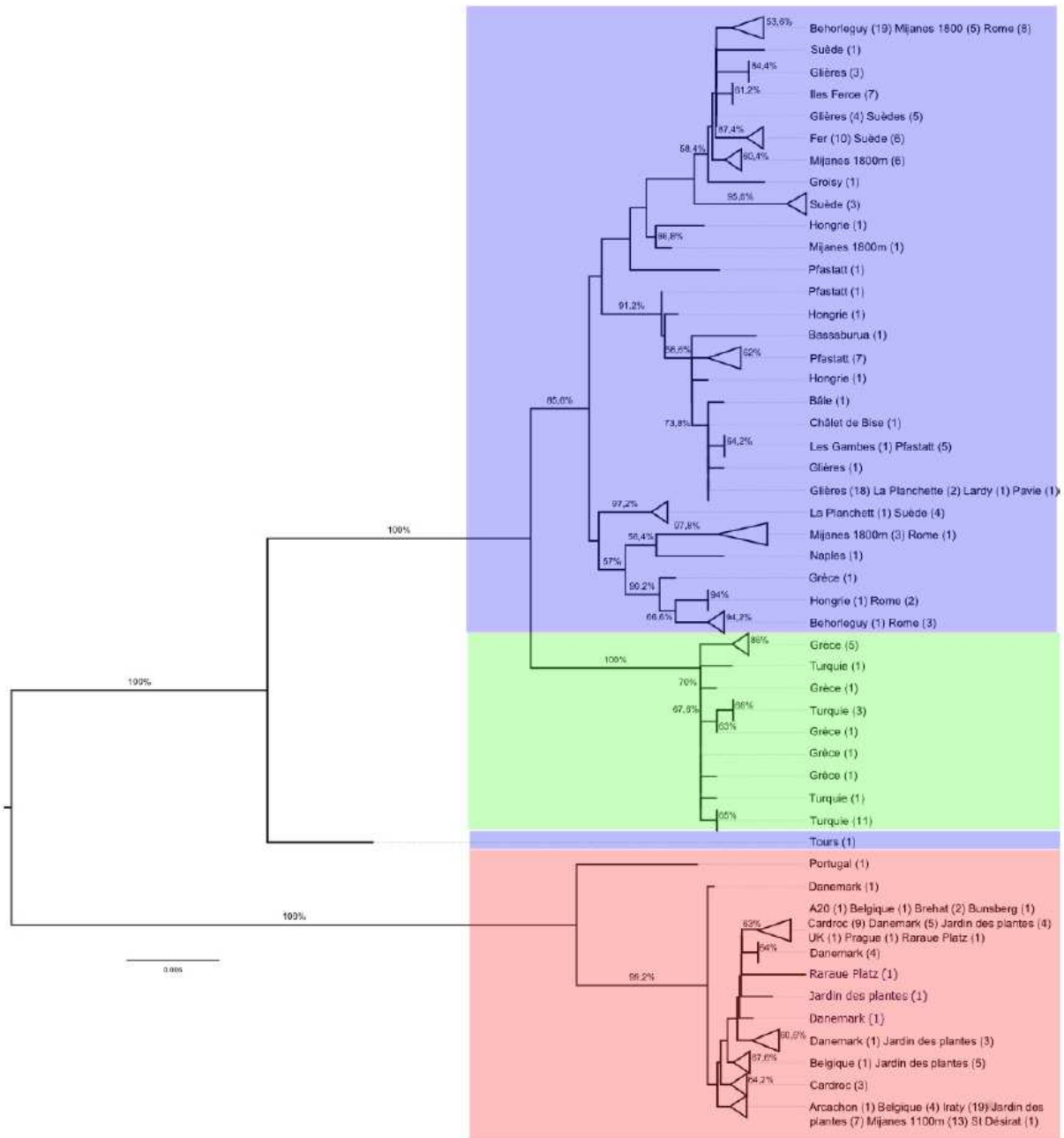


En bleu, l'espèce A, en rouge l'espèce B. Abréviations des localisations : voir tableau 1.

3. Phylogénie

Un arbre en *Neighbor Joining* a été réalisé d'après l'ensemble de nos séquences concaténées pour les deux gènes d'intérêt (Figure 9). Les échantillons de Grèce et de Turquie, contrairement aux autres, étaient particulièrement bien regroupés.

Figure 9 : *Neighbor Joining* de notre échantillon réalisé après 500 bootstraps.



La branche en bleu ségrège l'espèce A, la rouge l'espèce B, et la verte, une nouvelle subdivision. Les noms donnés sont ceux des localités de prélèvement des échantillons, les nombres entre parenthèses représentent le nombre d'individus du même haplotype, les pourcentages représentent les valeurs de bootstrap. Pour une plus grande commodité de lecture, les triangles représentent les regroupements de plusieurs individus. La barre d'échelle représente 0,005 substitution par site.

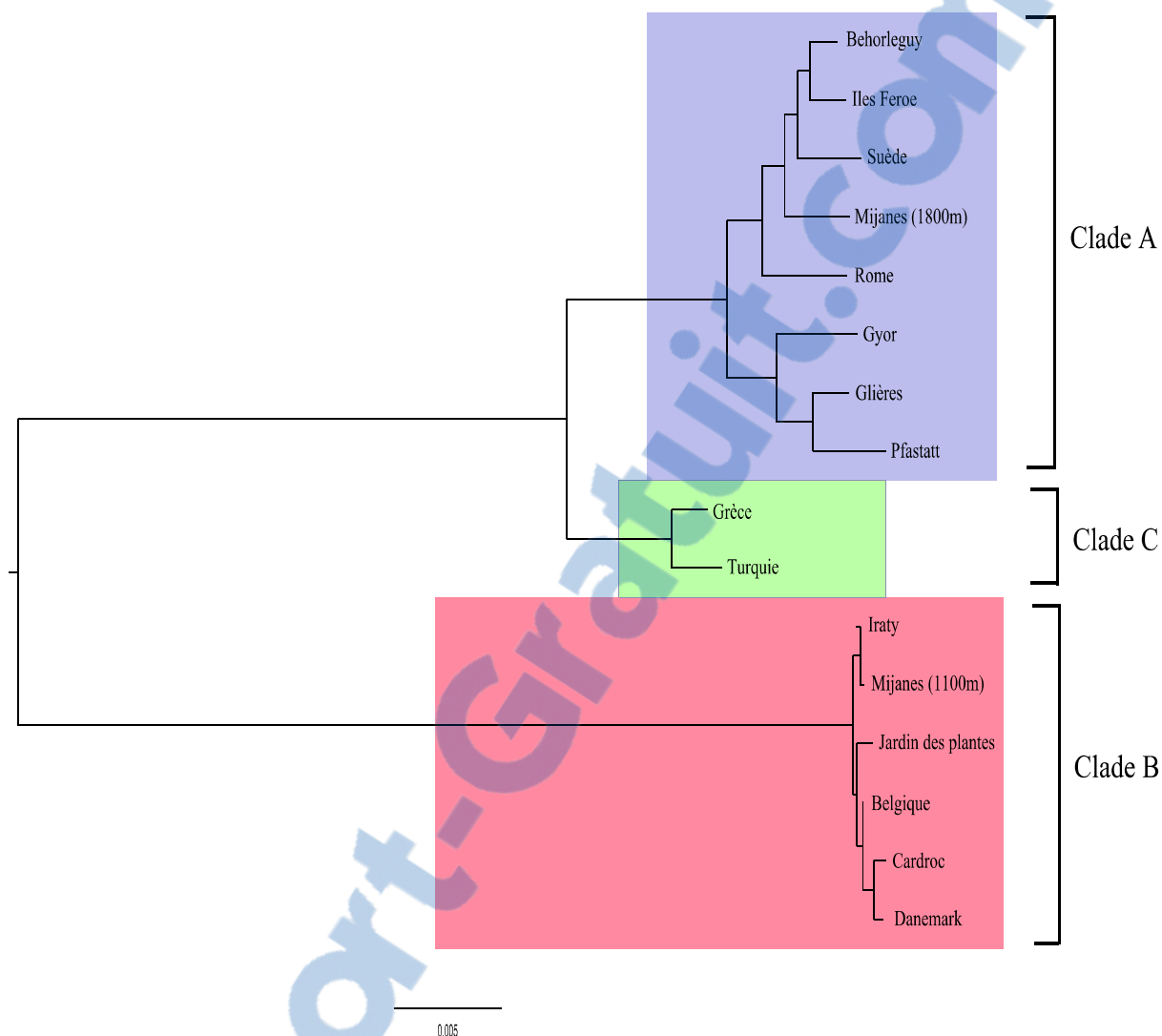
Nous avons observé une très nette ségrégation des deux clades, avec une distance génétique bien supérieure entre les espèces jumelles qu'au sein de chacune d'elles. De plus, les échantillons de Grèce et de Turquie semblaient particulièrement différents du reste des échantillons du clade A, et formaient un sous-clade monophylétique, que nous avons appelé clade C. L'hybridation entre individus du clade A et du clade C n'ayant pas été réalisée, il nous est impossible de désigner ce dernier clade comme une espèce au sens de Mayr (MAYR, 1942). Il est néanmoins possible de constater la proximité génétique importante entre des individus géographiquement éloignés (par exemple, des individus belges et pyrénéens). Ces résultats montraient également que le clade A était génétiquement beaucoup plus variable que le clade B.

De plus, en se basant sur une estimation de la diversité entre les populations (c'est-à-dire, en regroupant à un même nœud tous les individus prélevés à une même localité), nous avons obtenu un arbre simplifié populationnel, qu'il nous a été plus aisé d'interpréter (Figure 10). Nous avons mis en évidence le « clade C », nouveau, constitué d'individus grecs et turcs. Il a été possible de mettre globalement en évidence trois groupes phylogéographiques distincts :

- Clade A :
 - Arc Alpin-Europe centrale (montagnard)
 - Péninsule italienne
 - Pyrénées-Scandinavie (montagnard-nordique)
- Clade B : Bordure maritime occidentale de l'Europe
- Clade C : Péninsule des Balkans

Il est à noter qu'aucun échantillon populationnel (ensemble d'individus collectés au même endroit) n'a été trouvé hétérogène : à l'échelle micro-spatiale, les populations étaient donc non **sympatriques**.

Figure 10 : Arbre populationnel de nos échantillons en Neighbor Joining.



En bleu le clade A, en rouge le B, en vert le nouveau clade mis en évidence, que nous avons nommé clade C. La distance entre les populations est donnée en termes de *p*-distance, ou pourcentage de sites nucléotidiques auxquels deux séquences comparées diffèrent. La barre de légende correspond à 0,005% de substitution par site nucléotidique.

4. Diversité et structuration des populations

Les divers indices calculés entre nos populations dont les effectifs étaient de taille élevée (supérieure à dix individus) nous ont montré une diversité globalement moindre pour l'espèce B comparée à l'espèce A (Tableau 3). En effet, nous avons constaté une plus grande diversité haplotypique chez l'espèce A. Cette diversité se retrouvait également à l'échelle des nucléotides et était particulièrement frappante lorsque l'on comparait le nombre de sites polymorphes entre A et B (valeurs allant de 3 à 25 pour la première et de 0 à 6 pour la seconde). Le clade C, pour sa part, présentait des valeurs intermédiaires, sensiblement plus proches de celles du clade A que du clade B.

Tableau 3 : Comparaison des indices de diversités calculés avec DnaSP sur nos échantillons de grand effectif.

Population	Clade	Taille de l'échantillon	Nombre d'haplotypes h	Diversité haplotypique Hd	Nombre de singletons	Nombre de sites polymorphes S	Diversité nucléotidique Π	Nombre moyen de différences nucléotidiques k
BEH	A	20	4	0,363	18	19	0,00173	1,989
FER	A	17	2	0,515	0	3	0,00135	1,544
GLI	A	23	4	0,502	1	17	0,00541	6,206
GYO	A	4	4	1,000	21	24	0,01090	12,500
MIJ1	A	15	7	0,838	4	24	0,00691	7,924
PFA	A	16	7	0,825	9	15	0,00290	3,333
ROM	A	14	5	0,813	7	26	0,00866	9,934
SUE2	A	17	7	0,868	4	19	0,00574	6,588
TU	A	16	4	0,517	3	6	0,00129	1,483
PEL	A	10	7	0,911	22	25	0,00511	5,867
BEL	B	6	3	0,600	1	2	0,00075	0,867
CARD	B	12	6	0,818	4	6	0,00129	1,485
DAN	B	12	5	0,758	4	5	0,00100	1,152
IRA	B	19	1	0,000	0	0	0,00000	0,000
JDP	B	15	3	0,542	2	3	0,00103	0,679
MIJ2	B	14	3	0,275	2	2	0,00025	0,286

En bleu le clade A, en rouge le B, en vert le C.

Le D de Tajima ne montrait aucune valeur significative au seuil de significativité de 5%, excepté pour la population de Behorleguy, dans les Pyrénées, et pour celle des îles Feroe. Dans le premier cas, la valeur était de -2,377, et dans le second de 2,09. Ces deux valeurs nous orientaient grossièrement vers une expansion de la population de Behorleguy, et une contraction de la population des îles Feroe. Les Fs de Fu n'étaient pas significativement différents.

Tous ces indices convergeaient pour nous faire aboutir à la conclusion que le clade B présentait une structuration génétique plus faible que le A, et semblait donc beaucoup plus récent.

5. Flux de gènes entre populations

Les valeurs de **Fst**, calculées par DnaSP, sont présentées dans le tableau 4. Le cadre orange montre des valeurs très proches de 1, appuyant une grande différenciation génétique, ce qui correspondait à toutes les comparaisons entre les deux clades. Il s'agissait de valeurs toutes supérieures à 0,91802, signalant une différenciation génétique particulièrement importante entre les populations du clade B et celles du clade A. Les valeurs de Fst entre la Grèce et la Turquie d'une part et les autres populations du clade A d'autre part (moindre que celle entre Grèce et Turquie et celles du clade B néanmoins), dans le cadre jaune, signes de distance génétique relativement grande, nous ont confirmé la ségrégation de la Grèce et de la Turquie dans ce que nous avons nommé "clade C". Enfin, les autres valeurs intra-clade étaient plus faibles, montrant une continuité du flux génétique entre ces populations.

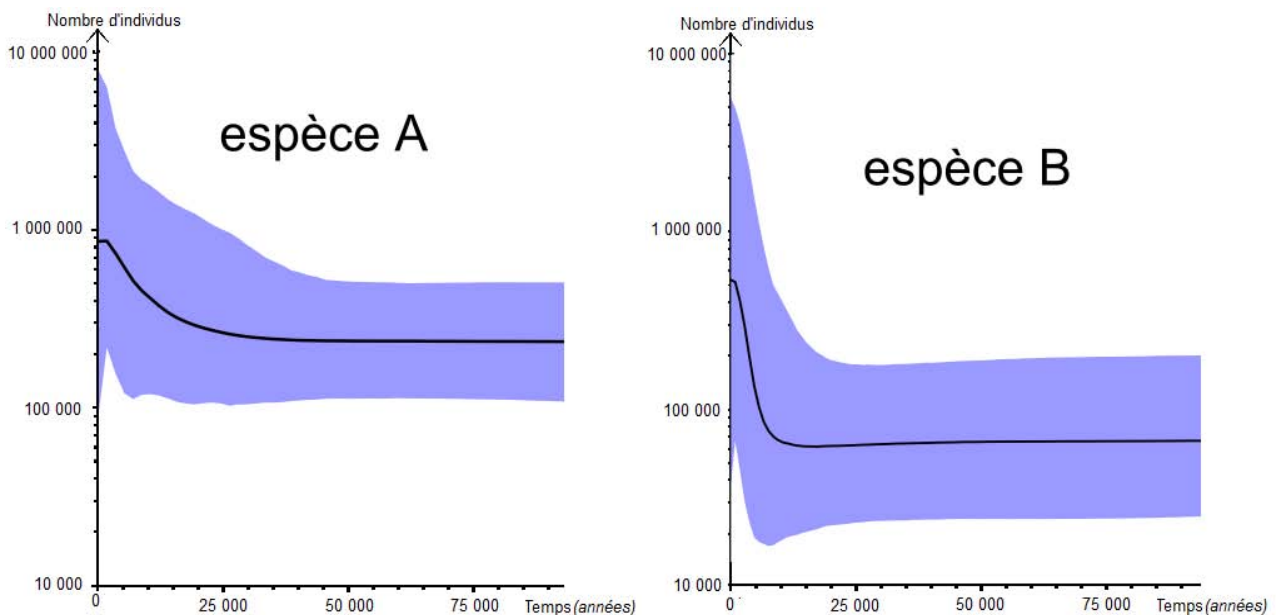
Tableau 4 : Valeurs de *Fst* calculées avec DnaSP.

	GLI	PFA	GYO	MU1	SUE2	ROM	BEH	FER	TU	PEL	BEL	IRA	JDP	MU2	CARD	DAN
GLI																
PFA	0.15618															
GYO	0.02553	0.08873														
MU1	0.41941	0.61602	0.22735													
SUE2	0.43945	0.64094	0.25829	0.09918												
ROM	0.40863	0.58071	0.17565	0.07813	0.16334											
BEH	0.65953	0.82669	0.46136	0.15799	0.29487	0.22141										
FER	0.68344	0.84480	0.49256	0.26661	0.21344	0.36351	0.49350									
TU	0.83678	0.89860	0.68926	0.80279	0.82548	0.75295	0.92840	0.93758								
PEL	0.73646	0.80108	0.57386	0.70193	0.72293	0.64182	0.83334	0.84338	0.20325							
BEL	0.96132	0.97709	0.92484	0.95034	0.95784	0.93850	0.98395	0.98639	0.98572	0.95852						
IRA	0.96600	0.98179	0.92958	0.95515	0.96267	0.94333	0.98880	0.99127	0.99097	0.96379	0.13333					
JDP	0.95436	0.96999	0.91802	0.94324	0.95065	0.93147	0.97659	0.97897	0.97775	0.95072	0.05639	0.32855				
MU2	0.96449	0.98025	0.92808	0.95361	0.96111	0.94179	0.98721	0.98967	0.98925	0.96209	0.10370	0.00000	0.30237			
CARD	0.95847	0.97405	0.92238	0.94754	0.95493	0.93583	0.98073	0.98312	0.98222	0.95536	0.29455	0.59504	0.15253	0.55203		
DAN	0.96022	0.97582	0.92423	0.94939	0.95679	0.93770	0.98258	0.98498	0.98421	0.95731	0.27345	0.61616	0.12915	0.56258	0.12121	

En bleu le clade A, en rouge le B, en vert le C. Le cadre jaune représente les valeurs de *Fst* entre la clade A et le clade C, le cadre orange représente les valeurs de *Fst* entre le clade A et le clade B.

L'analyse des figures de skylines (Figure 11) obtenues à l'aide de BEAST, sur un run de 100 millions de pas, a montré deux figures d'expansion pour les deux espèces, celle de l'espèce A étant plus ancienne (25 000 ans) que celle de l'espèce B (10 000 ans). Le nombre faible d'individus du clade C ne nous a pas permis de le faire entrer dans cette comparaison.

Figure 11 : Comparaison des « skylines » entre les deux espèces jumelles A et B.



La ligne noire représente la médiane du nombre d'individus de l'espèce (en **effectif efficace**) en fonction du temps (le passé étant situé à droite). Le temps de génération considéré est de un par an.

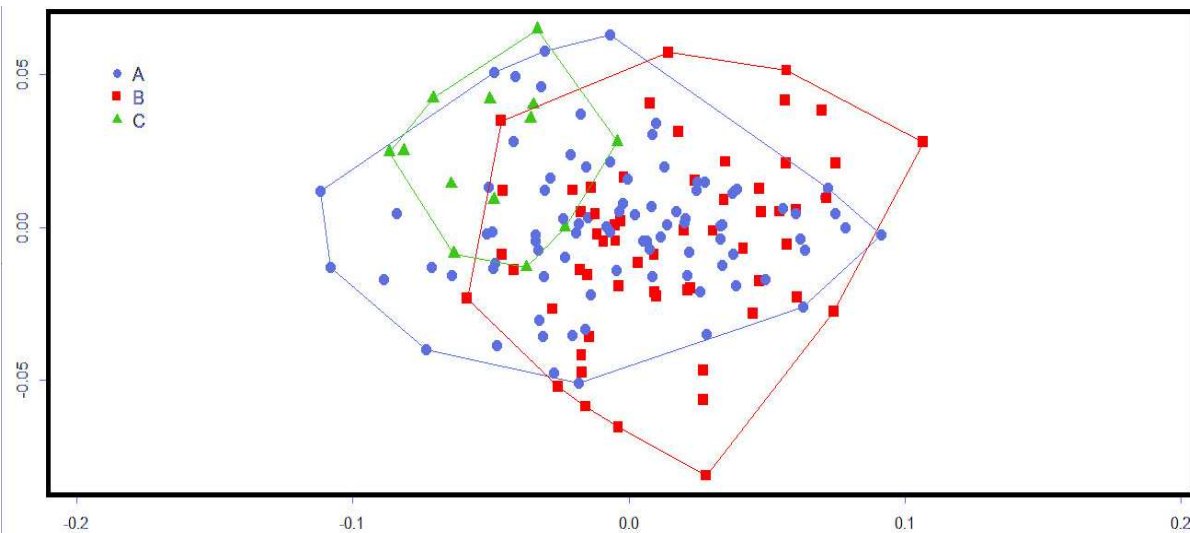
L'expansion de l'espèce B semblait bien plus marquée que celle de l'espèce A, avec un passage en 10 000 ans d'une taille de population de 30 000 à 280 000 individus, soit un taux de croissance de 25 individus par an en moyenne. L'espèce A en revanche, passant de 300 000 à 800 000 individus en 25 000 ans, avait un taux de croissance de 20 individus par an en moyenne. On peut alors globalement conclure que le clade A a vu une augmentation de son **effectif efficace** avec le temps, commencée avec le dernier maximum glaciaire (il y a environ 20 000 ans), et que le clade B a pour sa part vu son effectif efficace augmenter plus tard, après la fonte de l'inlandsis du nord de l'Europe (commencée en – 11.000, optimum climatique en – 6.000). Ces corrélations entre des périodes trouvées et des événements datés nous ont orienté vers la confirmation de l'hypothèse d'horloge biologique que nous avons indiquée.

6. Morphogéométrie

6.1. Le dépassement de l'aile gauche

La première structure comparée entre nos deux espèces était le dépassement de l'aile gauche sous l'élytre. Nous avons tout d'abord effectué une analyse en composantes principales (ACP) de l'ensemble des individus dans l'espace à 76 dimensions correspondant aux 76 valeurs de mesures de morphométrie effectuée sur chaque individu. Nous avons observé la répartition de nos individus sur les deux axes principaux, expliquant respectivement 56% et 21% de la variance (Figure 12).

Figure 12 : ACP du dépassement de l'aile gauche sur les deux premiers axes réalisée après superposition Procruste.



En bleu sont représentés les individus de l'espèce A, en rouge ceux de l'espèce B, en vert ceux de la clade C. Les deux axes principaux expliquent respectivement 56% et 21% de la variance observée. Chaque point représente la projection dans le plan principal d'un individu caractérisé par ses 76 mesures morphométriques.

Les deux axes principaux expliquant ainsi un grande part de la variance (77%), cette méthode de statistique descriptive sans *a priori* montrait une différenciation entre les clades :

- pour les clades A et B, nous avons noté un chevauchement des deux nuages de points, mais aussi de nombreux individus non chevauchant ;

- le clade C apparaissait dans le plan comme sensiblement inclus dans le clade A, (ce qui corroborait l'analyse génétique, (voir deuxième partie, paragraphe III. 4), mais les individus étaient répartis différemment de ceux du clade A (le périmètre de l'ensemble des points du clade C était plus restreint que celui des points du clade A).

Nous avons également effectué une MANOVA sur les données morphométriques des clades considérés deux à deux, afin de tester l'hypothèse d'homogénéité de leurs configurations morphologiques (cela équivaut au test du T2 de Hotelling, généralisation du test du t de Student à un cas multivarié) :

- les morphologies des clades A et B pouvaient être considérées comme différentes avec un risque d'erreur de 4,2% ;

- les morphologies des clades A et C pouvaient être considérées comme différentes avec un risque d'erreur de 6,1% ;

- pour les clades B et C, la comparaison n'était pas possible en raison du trop faible nombre d'individus total au regard du nombre de variables.

Nous avons enfin testé l'hypothèse d'égalité ou non de la taille du dépassement de l'aile selon les clades considérés deux à deux. Pour cela, nous avons effectué le test du t de Student sur le centroïde (une seule valeur numérique pour chaque individu, calculée par le logiciel TPS à partir des 76 mesures qui lui correspondaient) :

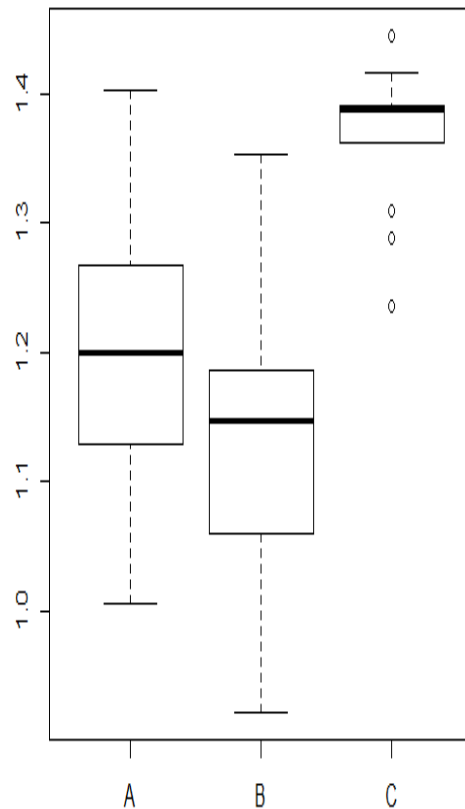
- les tailles du dépassement pour les clades A et B pouvaient être considérées comme différentes avec un risque d'erreur de $9,1 \cdot 10^{-7}$;

- les tailles du dépassement pour les clades A et C pouvaient être considérées comme différentes avec un risque d'erreur de $3,7 \cdot 10^{-9}$;

- les tailles du dépassement pour les clades B et C pouvaient être considérées comme différentes avec un risque d'erreur de $2,1 \cdot 10^{-15}$.

Nous avons donc pu affirmer avec un risque d'erreur négligeable que la taille du dépassement était différente selon le clade : elle est plus élevée dans le clade A que dans le clade B, et plus élevée encore dans le clade C (Figure 13).

Fig.13 : Boîtes à moustaches pour la taille des ailes, comparatives entre les clades.



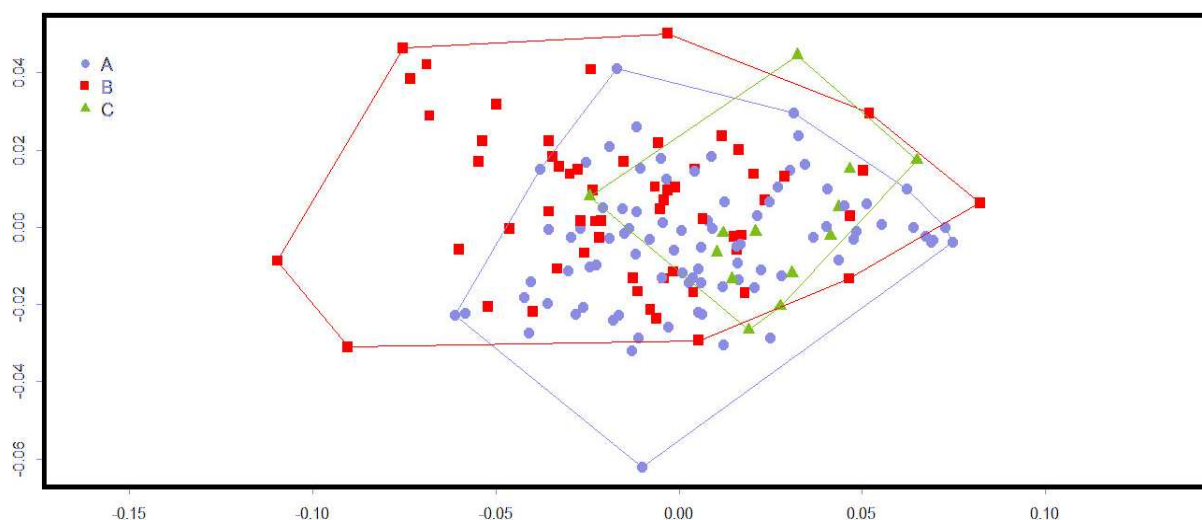
Ce graphique de type « boîte à moustache » représente pour les trois clades la taille des ailes (échelle : logarithme népérien de la taille exprimée en mm). Pour chaque clade, le trait épais central est la médiane, les bords du rectangle figurent respectivement le premier et le troisième quartile, les segments aux extrémités représentent les valeurs extrêmes.

Nous avons conclu de l'ACP, des MANOVA et des tests de Student que le dépassement de l'aile était spécifique à chaque clade.

6.2. Le pronotum

La seconde structure comparée entre nos deux espèces était le pronotum. Après réalisation d'une analyse en composantes principale de l'ensemble des individus dans l'espace à 96 dimensions correspondant aux 96 mesures de morphométrie effectuées sur chaque individu, nous avons observé leur répartition sur les deux axes principaux, expliquant respectivement 62% et 16% de la variance (Figure 14).

Figure 14 : ACP du pronotum sur les deux premiers axes réalisée après superposition Procruste.



En bleu sont représentés les individus de l'espèce A, en rouge ceux de l'espèce B, en vert ceux de l'espèce C. Les deux axes principaux expliquent respectivement 62% et 16% de la variance observée. Chaque point représente la projection dans le plan principal d'un individu caractérisé par ses 96 mesures morphométriques.

Les deux axes principaux expliquant ainsi un grande part de la variance (78%), cette méthode de statistique descriptive sans *a priori* montrait comme une différence entre les clades, analogue à celle qu'avait montrée l'ACP pour le dépassement de l'aile :

- pour les clades A et B, on notait un chevauchement des deux nuages de points, mais aussi de nombreux individus non chevauchant ;

- le clade C apparaissait dans le plan comme sensiblement inclus dans le clade A, (ce qui corroborait l'analyse génétique (voir deuxième partie, paragraphe III. 4), mais les individus étaient répartis différemment de ceux du clade A (le périmètre de l'ensemble des points du clade C était plus restreint que celui des points du clade A).

Comme précédemment, nous avons également effectué une MANOVA sur les données morphométriques des clades considérés deux à deux, afin de tester l'hypothèse d'homogénéité de leurs configurations morphologiques :

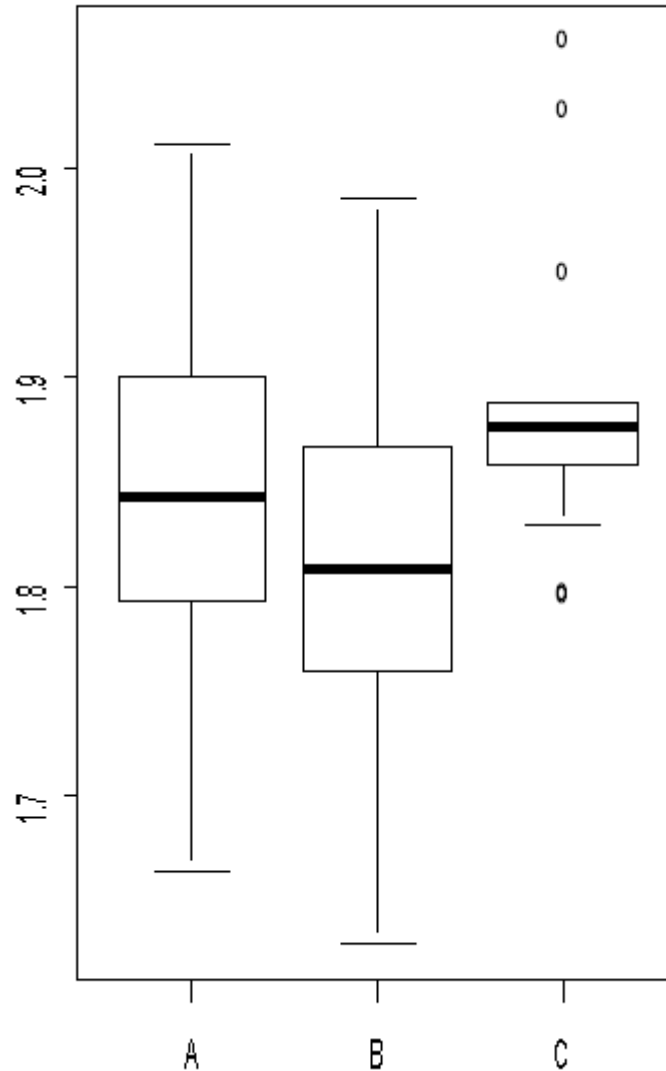
- les morphologies des clades A et B peuvent être considérées comme différentes avec un risque d'erreur de 3,4% ;
- les morphologies des clades A et C ne peuvent être considérées comme différentes avec un faible risque d'erreur ;
- pour les clades B et C, la comparaison n'est pas possible en raison du trop faible nombre d'individus au regard du nombre de variables.

Nous avons enfin testé l'hypothèse d'égalité ou non de la taille du pronotum selon les clades considérés deux à deux ; pour cela, nous avons effectué le test du t de Student sur le centroïde (une seule valeur numérique pour chaque individu, calculée par le logiciel TPS à partir des 96 mesures qui lui correspondent) :

- les tailles du pronotum pour les clades A et B peuvent être considérées comme différentes avec un risque d'erreur de 0,3% ;
- les tailles du pronotum pour les clades A et C peuvent être considérées comme différentes avec un risque d'erreur de 7,8% ;
- les tailles du pronotum pour les clades B et C peuvent être considérées comme différentes avec un risque d'erreur de 0,2% .

Comme pour la taille des ailes, celle du pronotum est plus élevée dans le clade A que dans le clade B, et plus élevée encore dans le clade C. Même si le risque d'erreur relatif à la comparaison entre les clades A et C (7,8%) n'est pas aussi faible que celui relatif aux deux autres comparaisons (de 0,2% à 0,3%), la différenciation aurait peut-être été plus significative avec un échantillon plus grand du clade C (Figure 15).

Fig.15 : Boîtes à moustaches pour le pronotum, comparatives entre les clades.



Ce graphique de type « boîte à moustache » représente pour les trois clades la taille des pronotums (échelle : logarithme népérien de la taille exprimée en mm). Pour chaque clade, le trait épais central est la médiane, les bords du rectangle figurent respectivement le premier et le troisième quartile, les segments aux extrémités représentent les valeurs extrêmes.

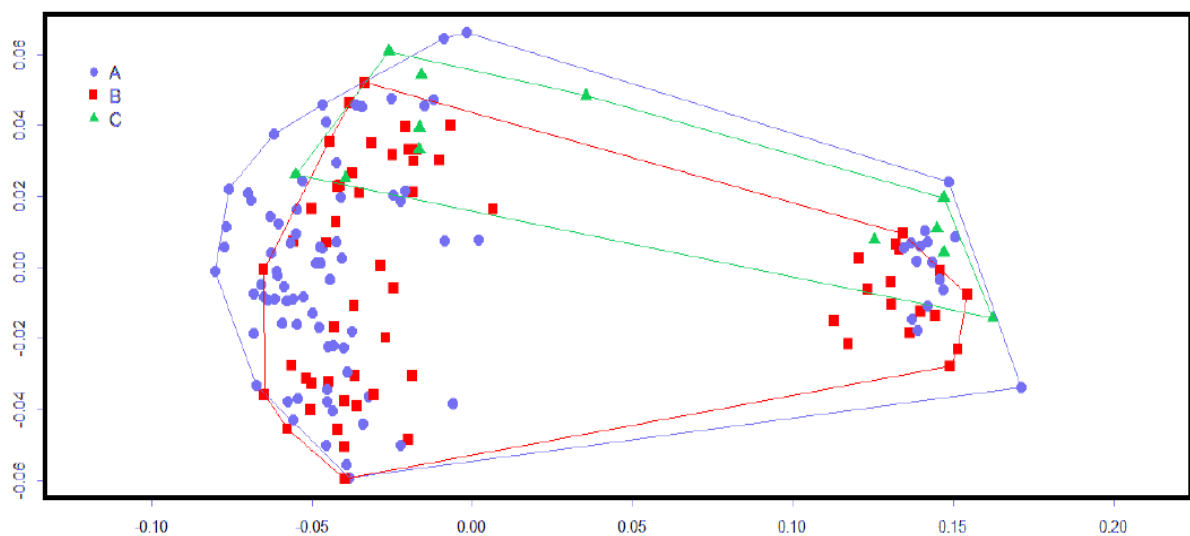
Bien que les différences entre les trois clades semblent un peu moins marquées que pour le dépassement de l'aile, on peut conclure de l'ACP, des MANOVA et des tests de Student que le pronotum est spécifique à chaque clade.

6.3. Les forceps

La dernière structure comparée entre nos deux espèces était le forceps gauche des mâles. L'analyse de cette structure, plus complexe en termes d'évolution que les deux autres, était à réaliser prudemment.

Sur la figure d'ACP (Figure 16), nous avons observé distinctement, au sein de chaque clade, la bimodalité bien connue, figurée par deux zones regroupant des individus proches dans chacun des clades. Pourtant, le recouplement était tel qu'il était malaisé de supposer une différence significative de forme de cette structure entre les trois clades.

Figure 16 : ACP des forceps sur les deux premiers axes réalisée après superposition Procruste.



En bleu sont représentés les individus de l'espèce A, en rouge ceux de l'espèce B, en vert ceux de l'espèce C. Les deux axes principaux expliquent respectivement 84% et 11% de la variance observée. Chaque point représente la projection dans le plan principal d'un individu caractérisé par ses 136 mesures morphométriques.

Pour les clades considérés deux à deux, les MANOVA (pour tester l'hypothèse d'homogénéité des conformations), et les tests de Student (pour tester l'hypothèse d'homogénéité des tailles) n'ont pas donné de résultats statistiquement significatifs (différences non significatives entre les clades).

6.4. Synthèse

Le tableau 5 ci-dessous résume les résultats précédents, obtenus pour les comparaisons des clades deux à deux.

Tableau 5 : Synthèse des résultats statistiques.

	<i>Clades A et B</i>	<i>Clades A et C</i>	<i>Clades B et C</i>	<i>Commentaires</i>
Configuration de l'aile	Différents (risque d'erreur : 4,2%)	Différents (risque d'erreur : 6,1%)	Comparaison non effectuée (trop faible effectif)	- Tailles croissantes : clades B, A et C
Taille de l'aile	Différents (risque d'erreur négligeable)	Différents (risque d'erreur négligeable)	Différents (risque d'erreur négligeable)	- Clade C très spécifique par rapport aux clades A et B
Configuration du pronotum	Différents (risque d'erreur : 3,4%)	Différence non significative	Comparaison non effectuée (trop faible effectif)	- Tailles croissantes : clades B, A et C comme pour l'aile
Taille du pronotum	Différents (risque d'erreur : 0,3%)	Différents (risque d'erreur : 7,8%)	Différents (risque d'erreur : 0,2%)	- Clade C moins spécifique que pour les ailes par rapport aux clades A et B
Configuration des forceps	Différence non significative	Comparaison non effectuée (trop faible effectif)	Comparaison non effectuée (trop faible effectif)	Les forceps ne semblaient pas différents
Tailles des forceps	Différence non significative	Différence non significative	Différence non significative	

DISCUSSION

Au cours de cette étude, nous avons pu tout d'abord confirmer la présence des deux clades A et B très divergents, correspondant aux espèces déjà caractérisées auparavant. Néanmoins, la répartition des deux espèces jumelles obtenue à l'issue de cette étude n'était pas celle à laquelle nous nous attendions.

La propagation des deux espèces dans l'ensemble de l'Europe a pu présenter un caractère artificiel, lié à une forte anthropisation du milieu et à l'augmentation des flux de transports. Ce phénomène récent aurait pu ainsi conduire à un brassage des lignées et à l'effacement partiel d'une structuration plus ancienne.

L'étude réalisée nous a également permis de mettre en évidence l'existence d'un troisième clade, le clade C, présent en Grèce et en Turquie. De plus, la comparaison de nos résultats et d'un échantillon de *F. auricularia* provenant de la base de données (GENBANK, 2013) et originaire d'Iran (comparaison non détaillée dans l'étude), ainsi qu'avec des individus océaniques séquencés en parallèle, a indiqué que la situation était sans doute plus complexe encore et que de nombreux clades divergents pouvaient exister, le long d'un axe ouest-est le long de l'Eurasie.

Notre étude nous a permis, entre autres, d'obtenir la phylogéographie des clades et nous a montré que, comparativement à l'étude de Guillet (GUILLET *et al.*, 2000), nous n'avions pas obtenu, à une altitude, une latitude et une longitude données, la présence des deux espèces A et B. Les populations échantillonnées ne correspondaient donc pas à un mélange, ce qui suggèrait que les espèces étaient allopatriques. Pourtant, à deux points des Pyrénées, nous avons trouvé, à deux altitudes différentes, les deux espèces A et B. Ce qui était particulièrement frappant était la présence à l'est (Mijanes) de l'espèce A plus en altitude que l'espèce B, et à l'ouest (Behorleguy et Iraty), la distribution inverse. L'altitude n'était donc pas le seul facteur influençant la distribution des espèces. Il est probablement nécessaire de prendre également en compte l'ensoleillement, si l'on reste cohérent avec une hypothèse d'adaptation locale.

Lors de l'étude de la diversité et de la structuration des populations, les divers indices calculés entre nos populations dont les effectifs étaient de taille importante (supérieure à dix individus) nous ont montré une diversité globalement moindre pour l'espèce B que pour les deux autres clades. Ces résultats sont à pondérer du fait du faible nombre d'individus dans le clade C pris en compte pour l'analyse, comparativement aux deux autres clades A et B.

Les analyses morphogéométriques réalisées, ont fait apparaître une différence globale entre les trois clades. Elles ont mis à jour le fait que la taille des ailes était plus élevée dans l'espèce A, espèce à une ponte, adaptée originellement à des climats plus rudes, que dans l'espèce B. Ces résultats sont en contradiction avec l'hypothèse d'un faible développement corporel dans des conditions climatiques difficiles. La supériorité de taille des ailes des individus du clade C sur celles des deux autres clades s'expliquait plus facilement par ce genre d'argumentation : il est courant qu'en milieu chaud (Grèce et Turquie pour le clade C) le développement corporel

soit plus important. Cette gradation de taille (B puis A puis C dans l'ordre croissant) se retrouvait également pour le pronotum, ainsi qu'une différence de conformation entre les trois clades.

Il est à noter que cette différenciation morphométrique, montrée ici en première approche, n'a jamais été observée auparavant. Néanmoins, il nous faut modérer nos conclusions, entre autres à cause du faible nombre d'individus du clade C pris en compte dans cette étude, ce clade ayant été découvert pendant l'étude elle-même.

Il serait intéressant, dans une perspective d'approfondissement de ce travail, de collecter davantage d'individus du dernier clade mis en évidence, le clade C, et de différencier au cours des analyses statistiques les zones de prélèvement, afin de pouvoir faire la part du poids du clade, comparativement à celui du biotope de l'individu, dans la taille et la conformation des structures. Cette démarche nous permettrait de séparer les différences de formes liées uniquement à l'environnement sur une génération (température, hygrométrie, etc.) de celles propre à un clade, et donc transmises de génération en génération. Discriminer chaque facteur (en réalisant une comparaison statistique par facteur et non par clade) puis les corrélérer pourrait aboutir à une meilleure compréhension de la dynamique des populations de *Forficula auricularia* en Europe, et en extrapolant, de multiples autres espèces européennes.

En ce qui concerne la datation réalisée quant à l'expansion des espèces, nous avons pu voir pour l'espèce A, une colonisation de tout l'espace européen, différencié en plusieurs zones. De plus, l'expansion récente de l'espèce B, faiblement différenciée génétiquement, limitée à la bordure atlantique de l'Europe, à plus faible effectif efficace que l'espèce A, nous orientait vers une colonisation récente depuis le sud au postglaciaire, il y a 10 000 ans. L'hypothèse d'une différenciation de l'espèce A en plusieurs refuges pourrait expliquer l'apparente « expansion » contre-intuitive, puisqu'elle date du dernier maximum glaciaire. Il s'agirait d'un événement de structuration en plusieurs isolats, qui augmenterait l'effectif efficace d'une population structurée au fur et à mesure de la divergence. En effet, bien qu'il existe une forte corrélation entre les périodes d'expansion trouvées et des événements connus (dernier maximum glaciaire entre autres) qui nous oriente vers la confirmation de l'hypothèse d'horloge biologique que nous avons indiquée, il nous faut néanmoins faire attention au fait que l'espèce A est fortement hétérogène, ce qui peut induire des datations en lien avec la structuration de l'espèce en îlots, non représentative de la réalité de l'espèce.

Cette étude nous a montré, sur un large échantillon, la répartition des deux espèces jumelles en Europe. Celle-ci n'étant pas corrélée de manière directe à l'altitude ou à la position géographique, il serait désormais intéressant de se questionner sur les mécanismes autres qui auraient pu aboutir à la spéciation observée. Il pourrait à l'avenir être judicieux de confronter les données des marqueurs mitochondriaux avec un marqueur nucléaire, *H3K9*, dont le laboratoire EPHE Biologie intégrative des populations possède déjà quelques séquences. En effet, une délétion dans la séquence de ce locus, préalablement observée, permettrait (grâce à des techniques rapides de PCR puis migration sur gel d'agarose) de détecter d'éventuelles zones d'hybridation, et d'affiner par là encore notre vision de la spéciation de cette espèce.

Enfin, certaines zones européennes sont encore peu connues en termes de forficules. Une augmentation des prélèvements, à l'avenir, vers les Balkans ou le Portugal, pourrait encore apporter un éclairage nouveau sur l'histoire évolutive de cet insecte.

CONCLUSION

La classification des êtres vivants a toujours été un défi pour l'homme. Nous avons passé en revue, de l'Antiquité à nos jours, l'évolution des techniques la sous-tendant, et plus généralement des principes sur lesquels s'est appuyée cette science. Ces principes ont grandement évolué, partant de notions théologiques et fixistes pour glisser progressivement vers les théories prenant en compte l'évolution, en passant par un changement d'échelle d'observation grâce aux avancées de la génétique des populations.

Sur le thème précis de la classification d'espèces jumelles, l'étude réalisée nous a permis d'aboutir à de nombreux résultats. Tout d'abord, nous avons eu confirmation des deux clades A et B très divergents, correspondant aux espèces préalablement caractérisées. Ensuite, nous avons pu mettre en évidence le fait que les populations échantillonnées ne correspondaient pas à un mélange, ce qui suggérait que les espèces étaient allopatriques. L'espèce B était faiblement différenciée génétiquement, limitée à la bordure atlantique de l'Europe, de faible effectif efficace, ce qui laissait envisager une colonisation récente de cette espèce depuis le sud, au postglaciaire. L'espèce A avait colonisé tout l'espace européen, et était différenciée en plusieurs espaces. L'hypothèse d'une différenciation de l'espèce A en plusieurs refuges pourrait expliquer l'apparente « expansion » contre-intuitive, puisqu'elle date du dernier maximum glaciaire.

Le regard rétrospectif sur le passé lointain du dernier maximum glaciaire et sur une période plus récente d'évolution nous a permis d'approfondir certains aspects de la phylogénie, de la différenciation et de la spéciation d'une espèce dont les contours restent encore imprécis, puisque outre les deux clades préalablement identifiés et provisoirement baptisés A et B, nous avons proposé la possibilité d'existence d'un troisième clade C. Dans le contexte actuel de diminution drastique de la biodiversité et de changements climatiques majeurs, où il est particulièrement important de se pencher sur la structuration des populations, sur l'apparition et la disparition d'espèces, les leçons ainsi tirées du passé peuvent contribuer à mieux comprendre les événements en cours ou à venir.

BIBLIOGRAPHIE

- ABCGP (2013). Site internet d'ABCGP [en-ligne].
[<http://www.abcgp.ca/fr/bibliotheque-parasites.asp?categorie=7>] (consulté le 18 septembre 2013).
- APPLIED MATHS NV (2011). BioNumerics version 7.1 [logiciel]. En vente chez l'éditeur : Applied Maths NV, Keistraat 120, 9830 Sint-Martens-Latem, Belgique.
- ARISTOTE (env. 350 av. J.C.). Physique, tome II, livre II : De la Nature, Chap. VIII. Traduction française : Barthélémy Saint-Hilaire, 1862, librairie philosophique de Ladrangé, 496 p.
- BIBLE (1910). Bible, version de Louis Second. Site internet Info-Bible.com, [en-ligne].
[<http://www.info-bible.org/lsg/01.Genese.html>] (consulté le 5 septembre 2013).
- BM (2013). Site internet du British Museum, Londres, [en-ligne].
[http://www.britishmuseum.org/explore/highlights/highlight_objects/pd/a/albrecht_d%C3%BCrers_rhinoceros.aspx] (consulté le 19 août 2013).
- BNF (2013). Site internet de la Bibliothèque Nationale de France, Paris. Dossiers pédagogiques de la BNF, [en-ligne].
[<http://classes.bnf.fr/dossism/index.htm>] (consulté le 20 août 2013).
- BATESON W., BRINDLEY H. H. (1892). On some cases of variations in secondary sexual characters statistically examined. *Proc. Zool. Soc. Lond.*, 585-594.
- CAPINERA J. L. (2008). Encyclopedia of entomology 2nd edition, volume 2. Kluwer Academic Publishers, Dordrecht/Boston/London, 4346 p.
- DARLU P., TASSY P. (1993). Reconstruction phylogénétique, concepts et méthodes. Masson, France, 245 p.
- DAWSON A. G. (1992). Ice Age earth: late Quaternary geology and climate, Routledge physical environmental series, New York, 293 p.
- DIAKONOV D. M. (1925). Experimental and Biometrical Investigations on Dimorphic Variability of Forficula. *Journal of Genetics.*, **15** (2), 201-232.
- DRUMMOND A. J., SUCHARD M. A., XIE D. et RAMBANT A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.*, **29**(8), 1969-73. [logiciels] (logiciels libres, téléchargeable sur http://beast.bio.ed.ac.uk/Main_Page).

- DUFF A. et LAWSON A. (2004). *Mammals of the World: A Checklist*. A & C Black, London, UK, 312 p.
- EBERHARD W.G. et GUTIERREZ A. A. (1991). Male dimorphism in beetles and earwigs and the question of developmental constraints. *Evolution*, **45** (1), 18-28.
- EFRON B. (1979). Bootstrap methods: another look on the jackknife. *The Annals of Statistics*, **7** (1), 1-26.
- FORSLUND P. (2000). Male-male competition and large size mating advantage in European earwigs, *Forficula auricularia*. *Anim Behav.*, **59** (4), 753-762.
- FREE SOFTWARE FOUNDATION, 2013. Logiciel Inkscape© version 0.48. [logiciel] (logiciel libre, téléchargeable sur <http://inkscape.org/?lang=fr>).
- FU Y. X. et LI W. H. (1993). Statistical tests of neutrality of mutations. *Genetics*, **133** (3), 693-709.
- FU Y. X. (1997). Statistical tests of *neutrality* of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915-925.
- GENBANK (2013). Site internet du National Center for Biotechnology Information, [en ligne] [<http://www.ncbi.nlm.nih.gov/nucleotide/361050964>].
- GINGRAS J. et TOURNEUR J-C. (2001). Timing of adult mortality, oviposition, and hatching during the underground phase of *Forficula auricularia* (Dermaptera: Forficulidae). *The Canadian Entomologist*, **133** (2), 269-278.
- GRAUR D., DURET L., GOUY M. (1991). Phylogenetic position of the order Lagomorpha (rabbits, hares and allies). *Nature*, **379**, 333-335.
- GUILLET S., GUILLER A., DEUNFF J. et VANCASSEL M. (2000) Analysis of a contact zone in the *Forficula auricularia* L. (Dermaptera : Forficulidae) species complex in the Pyrenean Mountains. *The Genetical Society of Great Britain, Heredity*, **85**, 444-449.
- HASEGAWA M., KISHINO H., YANO T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.*, **22** (2), 160-74.
- HEBERT P. D., RATNASINGHAM S., DEWAAD J.R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. B (Suppl.)*, **270**, 96-99.
- HENRI J. P. et GOUYON P.H. (2008). Précis de génétique des populations : Cours, exercices et problèmes résolus. Dunod, Paris, 186 p.
- HEWITT G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907-913.

HEWITT G. M. (2003). Ice ages: their impact on species distributions and evolution. *Evolution on Planet Earth*, 339-361. Academic Press.

HEWITT G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **359**, 183-195.

HUDSON R. R., SLATKIN M., MADDISON W. P. (1992a). Estimation of levels of gene flow from DNA sequence data. *Genetics*, **132**, 583-589. equation 3.

HUDSON R. R., BOOS D. D., KAPLAN N. L. (1992b). A statistical test for detecting population subdivision. *Mol. Biol. Evol.*, **9**, 138-151.

HUXLEY J. S. (1924). Constant differential growth-ratios and their significance. *Nature*, **114**, 895-896.

HUXLEY J. S. (1927). Discontinuous variation and heterogony in Forficula. *Journal of Genetics*, **17** (3), 309-327.

IFREMER (2013). Site internet d'Ifremer [en-ligne].
[<http://wwz.ifremer.fr/Decouvrir-les-oceans/Se-documenter/Sciences-de-la-Vie-de-la-Terre/La-vie-autour-des-sources-hydrothermales>] (consulté le 18 septembre 2013).

KIMBALL S., MATTIS P. et l'équipe de développement de GIMP (1995-2013). The Gimp©, version 2.8.4. [logiciel]. (logiciel libre, téléchargeable sur <http://www.gimp.org/downloads/>).

KOLLIKER M. (2007). Benefits and costs of earwig (*Forficula auricularia*) family life. *Behavioral Ecology and Sociobiology*, **61**, 1489-1497.

LEICA MICROSYSTEMS (2003-2009). Leica Application Suite [logiciel]. Leica microsystems company (logiciel payant, <http://www.leica-microsystems.com/products/microscope-software/life-sciences/las-easy-and-efficient/details/product/leica-application-suite/>).

LECOINTRE G., LE GUYADER H. (2001). Classification phylogénétique du vivant, 3ème édition, Belin, France, 544 p.

LE GUYADER H. (1998). L'Evolution. Belin, France, 192 p.

LE GUYADER H. (2003). Classification et évolution. Le Pommier, France, 124 p.

LIBRADO P. et ROZAS J. (2009). DnaSP v.5 : A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451-1452. [logiciel] (logiciel libre, téléchargeable sur <http://www.ub.edu/DnaSP/>)

MACKAY R. (2002). The Atlas of Endangered Species, Earthscan Publications, London, RU, 128 p.

- MAYR E. (1942). Systematics and The origin of species from the viewpoint of a zoologist, Columbia University Press, États-Unis, 348 p.
- MAYR E. (1963). Animal species and their evolution, Harvard University Press, États-Unis, 814 p.
- MAYR E. (1982). The growth of biological thought, Harvard University Press, États-Unis, 997 p.
- MEUNIER J. et KOLLIKER M. (2012). When it is costly to have a caring mother: food limitation erases the benefits of parental care in earwigs, *Biol. Lett.* **8**(4), 547-550.
- MUNOZ N. E. et ZINK A. G. (2012). Asymmetric forceps increase fighting success among males of similar size in the maritime earwig. *Ethology*. **118**(10), 943-954.
- PAPADOPOULOU A., ANASTASIOU I. et VOGLER A. P. (2010). Revisiting the Insect Mitochondrial Molecular Clock : The Mid-Aegean Trench Calibration, *Mol. Biol. Evol.* **27**(7), 1659-1672.
- PEREZ S. I., BERNAL V. et GONZALES P. N. (2006). Differences between sliding semi-landmark method in geometric morphometrics, with an application to human craniofacial and dental variation. *J. Anat.*, **20**, 769-784.
- PHILO (2011). Site internet philo-ECS-saint-exupéry, [en-ligne]. [<http://philo-ec.blogspot.ca/2011/04/ec-2-micrographie-et-imagination.html>] (consulté le 18 septembre 2013).
- POSADA D. (2008). Logiciel jModeltest version 0.1.1. In press. GUINDON S. et GASCUELV O. (2003) jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution*. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology* **52**, 696-704. [logiciel] (logiciel libre, téléchargeable sur <http://www.molecularevolution.org/software/phylogenetics/jmodeltest>).
- PRAT D., RAYNAL-ROCQUES A. et ROGUENANT A. (2008). Peut-on classer le vivant ? Linné et la systématique aujourd'hui. Belin, France, 438 p.
- PROVAN J., BENNETT K. D. (2008). Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution* **23**(10), 564-571.
- SAITOU N. et NEI M. (1987). The neighborjoining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* **4**(4), 406-425.
- TAJIMA F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**(3), 585-95.
- TAMURA K., PETERSON D., PETERSON N., STECHER G., NEI M. et KUMAR S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Likelihood, Distance, and

Parsimony methods [logiciel]. Molecular Biology and Evolution. (logiciel libre, téléchargeable sur <http://www.megasoftware.net/>).

TILLIER S. (2000). Systématique, ordonner la diversité du vivant, Tec&doc, France, 257 p.

TOMKINS J. L. et SIMMONS L. W. (1996). Dimorphisms and fluctuating asymmetry in the forceps of male earwigs. *Journal of Evolutionary Biology* **9**(6), 753-770.

TOMKINS J. L. et SIMMONS L. W. (1998). Female choice and manipulation of forceps size and symmetry in the earwig *Forficula auricularia* L. *Animal Behaviour*. **56**, 347-356.

TOMKINS J. L. (1999). Environmental and genetic determinants of the male forceps length dimorphism in the European earwig *Forficula auricularia* L. *Behav. Ecol. Sociobiol.*, **47**, 1-8.

TOMKINS J. L., BROWN G. S. (2004). Population density drives the local evolution of a threshold dimorphism, *Nature*, **43**, 1099-1103.

RANTALA M. J., ROFF D. A. et RANTALA L. M. (2007). Forceps size and immune function in the earwig *Forficula auricularia*. *Biological Journal of the Linnean Society*, **90**, 509-516.

R DEVELOPMENT CORE TEAM (2013) R: A language and environment for statistical computing, version 3.0.1 [logiciel] (logiciel libre, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>).

ROHLF F. J. (2013). Logiciels TPS : tpsDig version 2.17, tpsUtility version 1.55, tpsRelw version 1.49, [logiciels] (logiciels gratuits, droits réservés, URL respectifs <http://life.bio.sunysb.edu/morph/soft-dataacq.html> ; <http://life.bio.sunysb.edu/morph/soft-utility.html> ; <http://life.bio.sunysb.edu/morph/soft-tps.html>, Ecology&Evolution, SUNY, Stony Brook).

WATTERSON G. A. (1975). On the number of segregation sites. *Theor. Popul. Biol.* **7**, 256-276.

WILSON E. O. (2002). The Future of Life, Knopf, USA, 256 p.

WIRTH T., LE GUELLEC R., VANCASSEL M. et VEUILLE M. (1998). Molecular and reproductive characterization of sibling species in the European earwig (*Forficula auricularia*). *Evolution*, **52**(1), 260-265.

WIRTH T., LE GUELLEC R. et VEUILLE M. (1999). Directional Substitution and Evolution of Nucleotide Content in the Cytochrome Oxydase II Gene in Earwigs (Dermapteran Insects). *Mol. Biol. Evol.* **16**(12), 1645-1653.

ANNEXE 1 : Lexique

ACP	L'Analyse en Composantes Principales consiste à transformer des variables liées entre elles (corrélées) en nouvelles variables décorréelées les unes des autres, pour simplifier l'information. Les nouvelles variables décorréelées sont appelées « composantes » ou « axes » principaux, et peuvent être visualisées sur un graphique.
Spéciation allopatrique	Mécanisme de spéciation par isolement géographique de deux populations d'une même espèce, aboutissant à un isolement reproductif. Le flux génétique est coupé, chaque population évoluant alors séparément.
Artéfact d'attraction des longues branches	D'après la théorie dite de l'horloge moléculaire , certaines espèces possèdent un taux d'évolution moléculaire significativement plus élevé que la majorité des autres espèces. Les séquences de ces espèces ont alors tendance à se regrouper, indépendamment de toute relation de parenté. Ces regroupements, liés à un grand nombre de positions, auront des valeurs de <i>bootstrap</i> élevées. La pression de sélection qui joue sur un gène diffère selon les espèces, mais aussi au cours du temps. Il faut prêter une attention particulière à rentrer ce type d'information dans les algorithmes de reconstruction phylogénétique.
Barrière postzygotique	Type de barrières biologiques limitant la reproduction inter-espèce, existant après la fécondation. Ce phénomène serait dû à une incompatibilité des génomes, provoquant une stérilité ou une inviabilité des hybrides.
Dérive génétique	Effet du hasard pendant l'échantillonnage des gamètes, qui abouti à une fréquence des allèles différents entre l'échantillon à la génération n et celui à la génération n+1.
Effectif efficace	Puisque tous les individus d'une population ne participent pas à la reproduction et au brassage des gamètes, on appelle N_e l'effectif efficace, soit le nombre d'individus de la population qui suffiraient au brassage génétique observé dans la population réelle.
Espèces cryptiques	Espèces au sens biologique (isolement reproductif, différences génétiques) non différenciables morphologiquement.
ESS	Acronyme anglais de Stratégie Evolutivement Stable. Une stratégie est une ESS s'il s'agit d'un optimum, c'est-à-dire que lorsqu'elle est adoptée par l'ensemble de la population, aucune stratégie déviante ne peut s'implanter dans cette population.

Fitness	Mesure de l'adaptation d'un individu à son environnement.
Fst	Indice de différenciation entre deux populations, calculé grâce à la déviation des fréquences génétiques dans ces populations par rapport à celles que l'on observerait si le brassage génétique s'effectuait de manière continue. Des valeurs proches de 1 indiquent un très faible échange génétique.
Haplotype	Ensemble d'allèles de loci liés situé sur un même chromosome et transmis en groupe.
Horloge moléculaire	Théorie selon laquelle les séquences accumulent des mutations à un rythme dicté par l'intensité de la pression de sélection à laquelle elles sont soumises. En phylogénie, on peut utiliser des séquences à taux de mutation plus ou moins élevé selon qu'on veuille dater des événements plus ou moins récents.
Phylogéographie	Etude conjointe des lignées généalogiques et de leur distribution géographique, en lien avec des événements type dérive des continents, glaciations, etc.
Plasticité phénotypique	Capacité d'un organisme à exprimer divers phénotypes avec un génotype fixé, en fonction de son environnement.
Population	Ensemble d'individus de la même espèce qui occupent un même biotope, c'est-à-dire littéralement un même « <i>lieu de vie</i> ». En génétique des populations, on dira aussi qu'il s'agit d'un ensemble d'individus qui montrent une unité de reproduction, tous les individus d'une population ayant la même probabilité de se croiser entre eux, en se reproduisant moins avec les populations voisines.
Probabilités bayésiennes	Méthodes probabilistes s'appuyant sur le théorème de Bayes, où l'on déduit la probabilité d'événements en prenant en compte la probabilité d'autres événements préalablement observés.
Spéciation sympatrique	Mécanisme de spéciation sans isolement géographique, au sein d'une population ou dans une aire partagée. Les descendants sont alors en contact direct avec l'espèce ancestrale.
Taille efficace	Voir « effectif efficace »

ANNEXE 2 : Liste des contributeurs à l'échantillonnage de perce-oreilles

J. Arabi, A. Balard, A. Bouiges, T. Deuve, S. Duffield, A. Gilabert, H. Pitet, S. Quarrell,
M. Veuille, A. Whibley.

Un grand merci à toutes les personnes qui ont permis de fournir le laboratoire en forficules.

PHYLOGÉNIE, DIFFÉRENCIATION ET SPÉCIATION, EXEMPLE DU PERCE-OREILLE EUROPÉEN (*Forficula auricularia*)

NOM et Prénom : BALARD Alice

Résumé

Dans une première partie, nous avons effectué une synthèse bibliographique de l'histoire et des méthodes de classification des espèces. Nous avons dégagé les éléments théoriques et pratiques permettant d'appréhender les modifications de la répartition géographique dans le temps, des espèces, en particulier en lien avec les évolutions climatiques. Dans une deuxième partie, nous avons approfondi cette problématique par une recherche personnelle effectuée au laboratoire de l'École Pratique des Hautes Études (Biologie intégrative des populations) et portant sur la répartition spatio-temporelle des espèces jumelles de *Forficula auricularia*, en fonction de facteurs géographiques (altitude, latitude) et climatiques (influence du dernier âge glaciaire). Pour cela, nous avons conduit conjointement des analyses de génétique, portant sur deux marqueurs mitochondriaux, et de morphométrie, qui nous ont permis de mettre en évidence des différences entre les deux espèces et de faire l'hypothèse de l'existence d'une troisième population.

Mots clés

PHYLOGENIE, PHYLOGEOGRAPHIE, SPECIATION, GENETIQUE DES POPULATIONS, MORPHOMETRIE, CLADE, ESPECES JUMELLES, CLASSIFICATION, PERCE-OREILLE, FORFICULE, FORFICULA AURICULARIA.

Jury :

Président : Pr ...Professeur à la Faculté de Médecine de CRÉTEIL

Directeur : Dr. Marie Abitbol, Maître de conférences à l'ENVA

Assesseur : Pr. Philippe Bossé, Professeur à l'ENVA

PHYLOGENY, DIFFERENTIATION AND SPECIATION, EXAMPLE OF THE EUROPEAN EARWIG (*Forficula auricularia*)

SURNAME : BALARD

Given name : Alice

Summary

First, we have focused on methods of species classification. We have identified the theoretical and practical elements that explain the changes in species distribution over time, particularly in relation to climate change. We further examined in depth a specific aspect of this issue by performing a personal research at the École Pratique des Hautes Études laboratory (Integrative Population Biology), focusing on the spatio-temporal distribution of sibling species of *Forficula auricularia* according to geographical factors (altitude, latitude) and climate (influence of the last ice age). For that purpose, we have performed genetic analyzes of two mitochondrial markers and morphometric analyzes. These analyses allowed us to identify differences among the two species and to postulate the existence of a third population.

Keywords :

PHYLOGENY, PHYLOGEOGRAPHY, SPECIATION, POPULATIONS GENETICS, MORPHOMETRY, CLADE, SIBLING SPECIES, CLASSIFICATION, EARWIG, FORFICULE, FORFICULA AURICULARIA.

Jury :

President : Pr. Professor at the CRÉTEIL Medical Faculty

Director : Dr. Marie Abitbol, Associate Professor at ENVA

Assessor : Pr. Philippe Bossé, Professor at ENVA