

Introduction au DNA-Barcoding

Rapport d'étude bibliographique

BOUTELEUX Olivier
Janvier-Juin 2012

Master II "Sciences de l'Insecte"

Sommaire

I)	Introduction.....	1
II)	DNA-Barcoding : Principe.....	2
	• Le Barcode moléculaire.....	1
	• La bibliothèque de référence.....	3
	• Les outils de la Bio-informatique.....	3
	• La délimitation des espèces.....	4
III)	Des applications diverses et variées.....	5
	• L'étude de la biodiversité.....	5
	• La lutte contre les espèces invasives.....	6
	• Une implication à tous les niveaux.....	7

IV)	Le DNA-Barcoding : entre limites et innovations.....	7
	• Le gène COI.....	8
	• La bibliothèque de référence.....	9
	• Analyse, utilisation du DNA-Barcoding.....	9
	• Le NGS : Next Generation Sequencing.....	10
V)	Conclusion.....	11
VI)	Bibliographie.....	12

I) Introduction

La biologie moléculaire est une discipline scientifique qui commença à se développer dès la deuxième moitié du XX^{ème} siècle. L'amorce de cette discipline a notamment été illustrée par la publication de James Watson et Francis Crick (Watson and Crick, 1953) dans laquelle fut présenté pour la première fois, le modèle de la structure de l'ADN. Cette découverte fut en effet considérée comme le pivot à l'établissement de la biologie moléculaire au sein des sciences de la vie (Halloran, 1984). Assurément, la biologie moléculaire a, par la suite, connu de nombreuses avancées et a permis la compréhension de certains mécanismes, notamment la réplication et l'expression des gènes (Lunardi, 2003). Elle est ainsi devenue un outil incontournable et s'est intégrée en fait dans toutes les disciplines s'intéressant au vivant. Aujourd'hui, la biologie moléculaire repose aussi sur des outils enzymatiques qui visent avant tout à manipuler et à caractériser l'ADN. En effet, il va pouvoir être possible de couper, lier, ou encore séquencer l'ADN (Laudenbach *et al.* 1999). Les applications de la biologie moléculaire sont ainsi devenues très diversifiées et ont permis des avancées considérables dans de nombreux domaines comme les sciences médicales, la biologie marine (Narsinh *et al.* 2008), etc. Dans ce contexte, une approche génomique proposée récemment par Paul Hebert (Hebert *et al.* 2003) a fait preuve d'un succès notoire dans les années qui lui ont précédé et est aujourd'hui on ne peut plus d'actualité. Cette technique récente, appelée DNA-Barcoding, a en effet rapidement pris une échelle mondiale au sein de l'univers scientifique, en trouvant notamment sa place dans de nombreux domaines d'application.

Le "DNA-Barcoding" est finalement un outil taxonomique qui utilise un fragment standard de l'ADN pour déterminer à quelle espèce appartient un organisme (Hebert *et al.* 2003). En effet, tout comme les empreintes digitales permettent d'identifier chacun d'entre nous, la variabilité nucléotidique de ce fragment constitue des combinaisons uniques qui permettent d'identifier les espèces. Ce fragment d'ADN séquencé pour chaque organisme permet d'élaborer des données de références, lesquelles constituent un système global d'identification moléculaire des espèces (Hebert *et al.* 2003; Valentini *et al.* 2009; Casiraghi *et al.* 2010).

Cette nouvelle approche moléculaire a été proposée en réponse aux insuffisances de la taxonomie conventionnelle, par laquelle l'identification des espèces peut constituer un véritable défi, notamment face à la perte constante de la biodiversité. En effet, bien qu'établie depuis plus de deux siècles, cette expertise taxonomique, basée notamment sur des critères morphologiques, n'est pas sans présenter un certain nombre de lacunes. Principalement, elle est très coûteuse en temps et requiert des connaissances hautement spécialisées (Costion *et al.* 2011; Raupach *et al.* 2010). Qui plus est, les experts taxonomiques sont très dispersés à travers le monde et leur nombre ne cesse de

décroître chaque année (Tanzler *et al.* 2012). Finalement, cette expertise est très souvent spécialisée à un taxon particulier et ne permet pas une identification exhaustive des espèces (Gaikwak *et al.* 2011).

En 2003, Paul Hebert proposa donc le DNA-Barcoding comme un élément de réponse très prometteur qui viendrait compléter l'accès à l'identification du monde vivant (Hebert *et al.* 2003). Dès lors, cette approche a connu un nombre important d'applications comme la détection d'espèces invasives (Armstrong *et al.* 2005; Cross *et al.* 2010) ou encore la lutte contre le commerce illégal d'espèces menacées (Eaton *et al.*, 2010).

Il sera ainsi détaillé dans ce rapport d'étude bibliographique les réels bénéfices qu'apporte le Barcoding, tout en évoquant aussi quelles en sont les limites. Ceci nous permettra finalement d'interpréter comment est perçue le futur de cette nouvelle discipline.

II) DNA-Barcoding : Principe

L'utilisation de séquences ADN pour caractériser les organismes existait déjà plusieurs années avant la publication de l'article fondateur du DNA-Barcoding proposé par Paul Hébert. En effet, certains travaux parus il y a une voire deux décennies, mettaient déjà en évidence le fait que les variations nucléotidiques présentes au sein de l'ADN représentaient des combinaisons uniques permettant de caractériser les espèces et d'élaborer des phylogénies (Takeyama *et al.* 2000; Gibson *et al.* 1988).

- Le Barcode moléculaire

Comme évoqué précédemment, l'originalité du concept proposé par Paul Hébert a été de proposer un fragment standard de l'ADN, lequel est commun à la plupart des organismes. Le fragment choisi est un gène appartenant au génome mitochondrial. Ce gène (648 pb) code pour la première sous-unité de la Cytochrome-Oxydase (COI), laquelle est une protéine intervenant dans la chaîne respiratoire des mitochondries et est un élément clé dans le métabolisme aérobie. Le choix de ce gène dans l'identification du monde vivant est justifié par un certain nombre d'avantages (Hebert *et al.* 2003a; Hebert *et al.* 2003b):

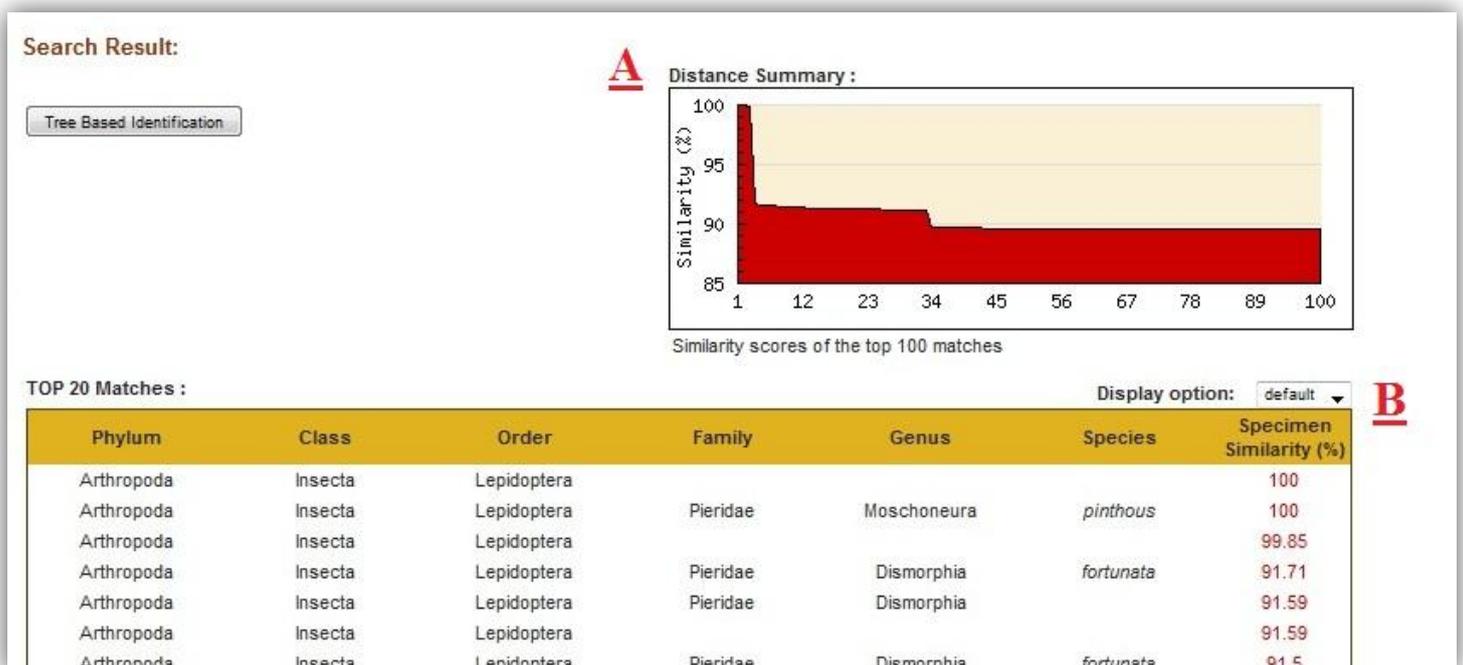
(1) Tout d'abord, ce gène a une vitesse d'évolution relativement élevée et permet une accumulation appréciable de mutations au sein des espèces. De ce fait, par comparaison de séquences ADN, il a été noté que ce gène COI divergeait suffisamment pour permettre de discriminer les espèces, voire même discriminer des populations au sein d'une même espèce (Cox and Hebert 2001; Wares and Cunningham 2001). (2) De plus, les mitochondries faisant partie intégrante de la majorité des types

Figure 1 : Outil d'identification présent sur *BOLD systems*

Pour identifier un individu quelconque via l'outil d'identification, il suffit de soumettre sa séquence en faisant un copier/coller dans un champ prévu à cet effet.

Cet outil nous retourne ensuite un résultat comportant plusieurs éléments : (ci-dessous)

- Il présente notamment un graphique **(A)** comportant en abscisses les 100 premiers individus référencés pour lesquels le pourcentage de similarité avec la séquence soumise est le plus important. En ordonnées est présentée la valeur de ce pourcentage de similarité. Ce graphique nous permet de déterminer si la séquence soumise est déjà plus ou moins représentée parmi la bibliothèque de référence.
- Un second élément **(B)** lui indique les éventuels noms d'espèces associés à chacun de ces individus de référence ainsi que leur pourcentage de similarité avec la séquence soumise. C'est cet élément qui va nous permettre d'attribuer un nom d'espèce à notre individu. Nous pouvons par exemple noter ici que l'espèce *Moschoneura pinthous* est une très bonne candidate car elle présente une similarité du gène COI de 100 % avec la séquence que nous avons soumise.



cellulaires, le gène COI est de ce fait présent en très grande quantité dans les cellules ce qui facilite énormément son séquençage en laboratoire. (3) Enfin, les amorces nécessaires au séquençage de ce gène sont robustes et peuvent être utilisables au sein d'un même grand groupe d'organismes comme les lépidoptères (Hébert et al, 2003a), coléoptères etc.

- **La Bibliothèque de Référence**

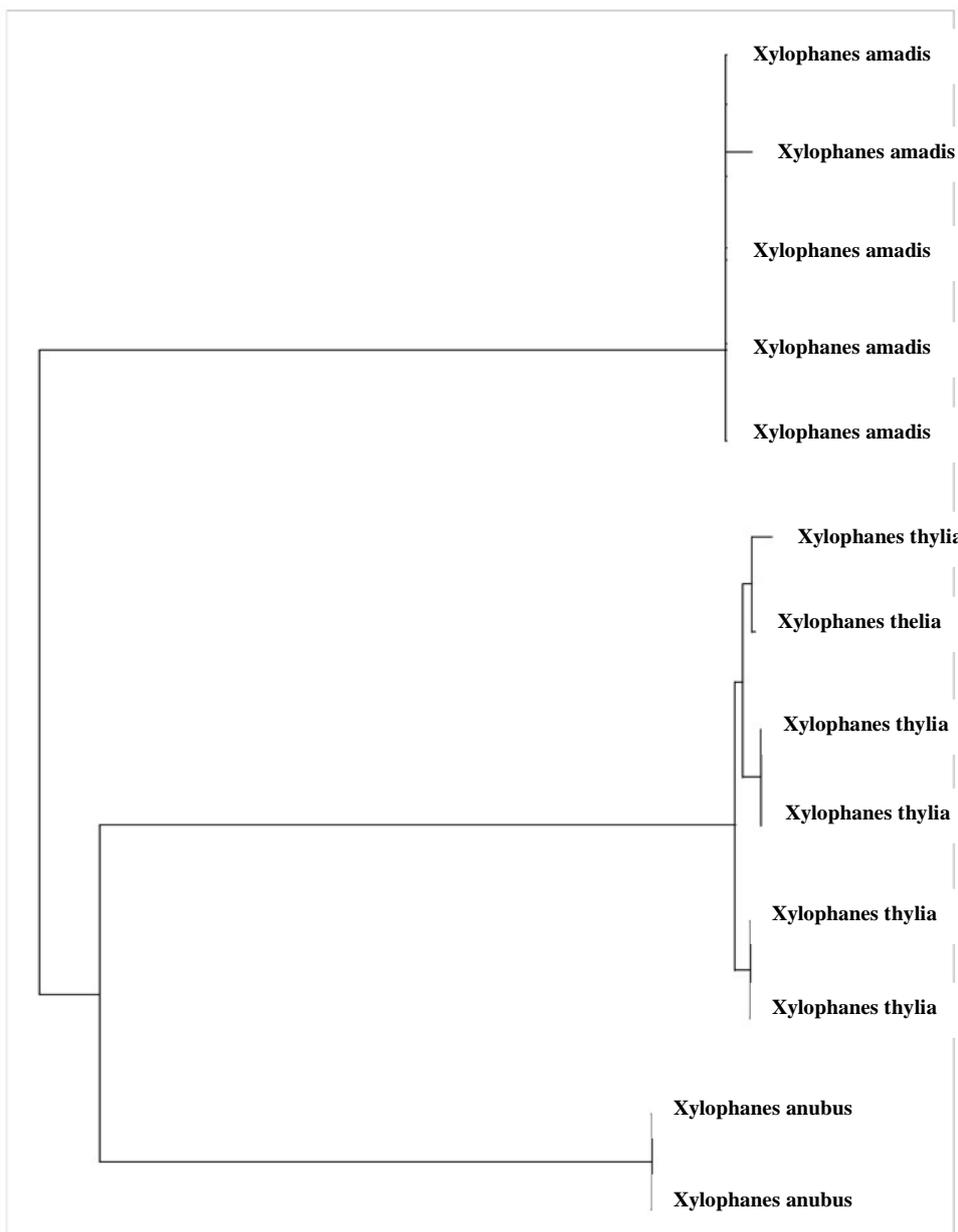
L'objectif principal du Barcoding a été dans un premier temps de séquencer le gène COI pour un nombre très important d'espèces déjà connues, notamment par le biais des muséums et des taxonomistes, pour ensuite regrouper toutes ces séquences et élaborer ainsi une bibliothèque de référence (Ratnasingham and Hebert, 2006). Grâce à cette bibliothèque présente en ligne (www.boldsystems.org), il devient possible pour un utilisateur d'identifier un individu d'une espèce qui lui est inconnue en réalisant le séquençage du gène COI et en le comparant avec les séquences présentes dans cette bibliothèque (**Fig. 1**). Cet outil est très novateur dans le sens où ici aucune expertise taxonomique n'est réellement nécessaire. En effet, la séquence ADN peut permettre à l'utilisateur de trouver un nom d'espèce associé ainsi que de nombreuses autres informations adjointes (photographie, écologie, distribution ...). Comme seule la séquence ADN est nécessaire, il devient alors aisé d'identifier aussi des stades larvaires chez les insectes dont certains étaient jusqu'alors inconnus. Cela a notamment été le cas chez les éphémères (Ball *et al.* 2005), chez les coléoptères (Miller *et al.* 2005), ainsi que chez les lépidoptères (Gossner and Hausmann. 2009). Mieux encore, même des fragments d'individus peuvent être séquencés, ce qui rend réalisable l'identification du contenu du tube digestif de certaines espèces (Riemann *et al.* 2010).

- **Les outils de la bio-informatique**

Dès leur mise en place, les bibliothèques de référence ont montré une grande efficacité. En effet, dans les années 2003 à 2006, 95% des espèces séquencées possédaient une séquence du gène COI bien distincte (Hajibabaei *et al.* 2006; Hebert *et al.* 2003b, 2004a; Ward *et al.* 2005). En parallèle à cela, cette bibliothèque de référence en ligne a mis en place certains outils de la bio-informatique qui permettent d'analyser instantanément les données soumises (Ratnasingham and Hebert, 2007). Parmi ces outils, nous pouvons noter la possibilité de créer des arbres phénétiques basés sur la méthode du Neighbour-joining (Saitou and Nei, 1987). La lecture de ces arbres phylogénétiques, par leur présentation sous forme de clusters, nous permet de distinguer et caractériser les espèces présentes dans un échantillon donné. La configuration type d'un arbre phylogénétique se présente de la façon suivante : les individus appartenant à une même espèce sont

Figure 2: Réalisation d'un arbre phylogénétique

La méthode du Neighbour joining est un des procédés qui permet l'élaboration d'arbres phénétiques à partir d'un échantillon donné d'organismes. Pour agencer ces organismes au sein de l'arbre phénétique, il est nécessaire de mesurer la distance génétique qui les sépare. Cette mesure de la distance génétique se base surtout sur le pourcentage de différence observé entre les séquences ADN (ici le gène COI) de ces organismes, aussi appelés unités taxonomiques opérationnelles (OTU). Plus ce pourcentage de différence est faible, plus la distance génétique qui les sépare sera courte. La connaissance de la distance génétique entre les différents taxons (genre, espèce) et l'utilisation d'algorithmes adaptés permet de regrouper les individus appartenant au même taxon au sein de clusters distincts.



Ci-contre :

Prise de vue d'un arbre phylogénétique réalisé à partir d'individus appartenant à trois espèces différentes.

Nous pouvons noter ici que les individus appartenant à la même espèce sont bien regroupés au sein de clusters distincts.

regroupé au sein d'un même cluster, lesquels clusters s'agencent en fonction du taux de divergence du gène COI présent entre les espèces (**Fig. 2**).

Dès lors, combiné avec ces outils de la bio-informatique, le Barcoding créa un engouement prépondérant, et fut qualifié comme étant bien plus qu'un outil d'identification des espèces (Taylor and Harris, 2012; Hebert *et al.* 2003a).

- **La délimitation des espèces**

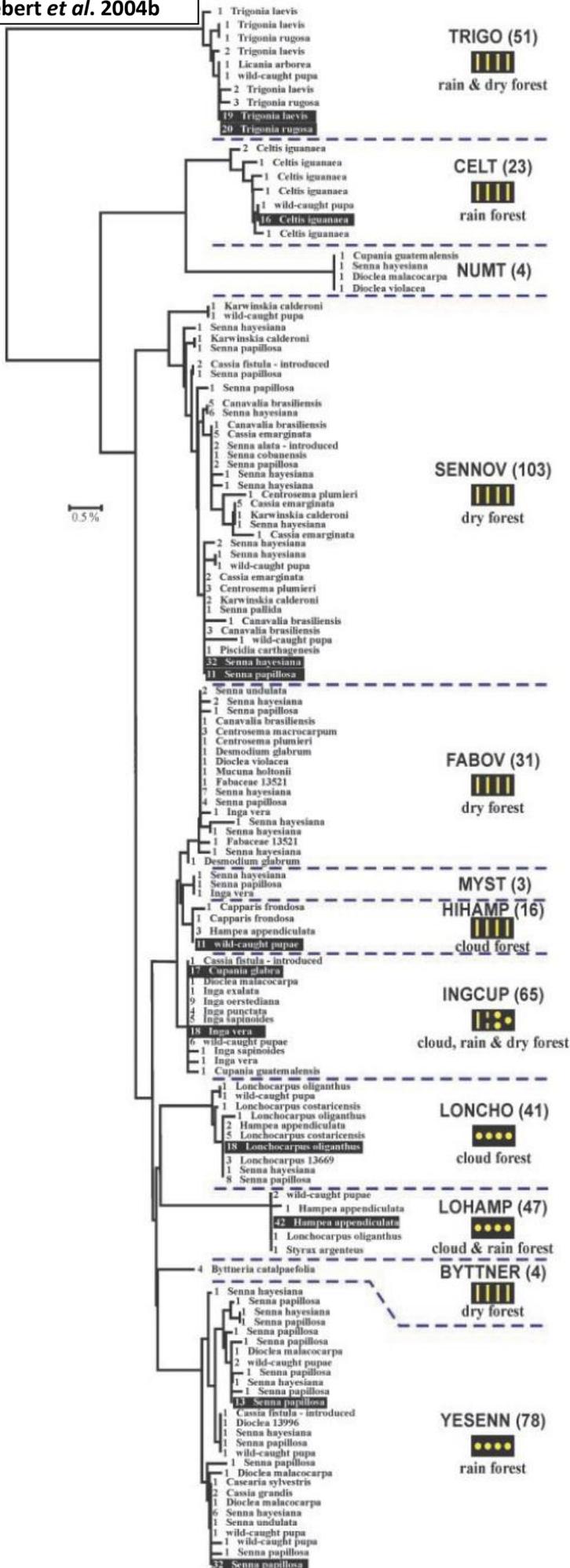
La séparation des espèces au sein d'un arbre phénétique doit se faire selon un critère qui permet avant tout de les différencier. Ce critère, comme évoqué dans la figure 2, réside dans le pourcentage de différences observées au niveau de la séquence nucléotidique du gène COI. Ce pourcentage de différences entre les espèces pour le gène COI peut fortement varier en fonction du groupe d'organismes pris en compte. Par exemple, chez les lépidoptères ce pourcentage de différences est rarement inférieur à 2%. Il fait alors office de *seuil* (ou *threshold*), en deçà duquel les individus concernés sont considérés comme appartenant à la même espèce. Pour que cette méthode soit applicable, il est donc nécessaire que les variations interspécifiques du gène COI soient distinctement plus importantes que les variations intraspécifiques (Meyer and Paulay 2005). D'ailleurs, dans certains cas, d'importantes variations intraspécifiques posent quelques soucis dans la discrimination des espèces (Meier *et al.* 2006).

Ces études de la variation du gène COI au sein et entre les espèces ont permis des découvertes très intéressantes. En effet, parfois les variations intraspécifiques pour ce gène sont telles qu'elles semblent indiquer la présence de plusieurs espèces cryptiques au sein d'un taxon jusqu'alors considéré comme étant une seule espèce distincte. Grâce à des analyses adaptées, les résultats montrent en effet des arbres phénétiques comportant plusieurs clusters bien apparents comme si l'espèce en question était en fait un complexe d'espèces (**Fig. 3**). Ces résultats ont été trouvés dans de nombreux cas, et ont remis en question le statut de certaines espèces (Hebert *et al.* 2004b; Ragupathy *et al.*, 2009; Hausmann *et al.*, 2011) voire de certains écotypes (Bennett *et al.* 2011).

Ainsi, l'utilisation du DNA-Barcoding pour tenter de redéfinir certaines espèces fut largement prise en compte (Blaxter, 2004; Tautz *et al.* 2003). Cet outil fut alors qualifié non seulement comme un outil d'identification mais aussi un outil permettant la délimitation et la découverte potentielle des espèces. Cette autre capacité du DNA-Barcoding fut d'ailleurs mise en avant dès la première publication de Paul Hebert (Hebert *et al.* 2003a), et a permis d'accroître incontestablement son utilisation.

Figure 3 : Mise en évidence d'espèces cryptiques de par l'élaboration d'arbres phylogénétiques

Hebert *et al.* 2004b



Ci-contre :

La figure représente le résultat d'analyses phénétiques réalisées par Paul Hebert (Hebert *et al.* 2004b) sur une espèce de lépidoptères néo-tropicale.

Nous pouvons remarquer ici que cet arbre phénétique est subdivisé en une dizaine de clades, lesquels sont délimités par des pointillés.

La présence de ces différents clusters reflète d'importantes variations génétiques au sein de cette espèce et semble finalement indiquer qu'il s'agit en fait d'un complexe de 10 espèces différentes. Des analyses morphologiques et écologiques complémentaires sont allées dans le sens de cette hypothèse (Hebert *et al.* 2004b).

III) Des applications diverses et variées

Ainsi qualifié comme un outil permettant d'identifier, de découvrir et délimiter les espèces, le DNA-Barcoding s'est implémenté de façon considérable au sein des études scientifiques. En effet, des années 2003 à 2010, pas moins de 411 publications ont mentionné le DNA-Barcoding dans leur titre (Taylor and Harris, 2012).

- **L'étude de la biodiversité**

Comme évoqué en introduction, la principale motivation quant à l'utilisation du DNA-Barcoding fut sans conteste pour combler les lacunes de la taxonomie conventionnelle (Herbert *et al.* 2003a).

La taxonomie conventionnelle est une discipline qui a toujours pris une part importante dans l'étude de la biodiversité, notamment l'étude de la richesse spécifique d'un milieu. En effet, les taxonomistes, de par leur expertise, leurs publications, leurs collectes sur le terrain fournissent une base de données essentielle permettant d'identifier et classifier les espèces (Janzen and Hallwachs, 2011; Moritz and Cicero, 2004). A l'heure où la biodiversité ne cesse de s'appauvrir, et où le temps imparti à son étude doit être le plus court possible, l'arrivée du DNA-Barcoding pour aider à l'identification des espèces fut ainsi prise avec beaucoup de considération (Blaxter 2003; Janzen 2004).

Notamment parce qu'il permet une identification efficace des espèces, le DNA-Barcoding a en effet apporté un certain nombre de bienfaits qui ont permis d'améliorer notablement les techniques de la taxonomie conventionnelle.

Comme cette identification se fait par simple comparaison de séquences ADN, elle est de ce fait beaucoup plus rapide que l'identification traditionnelle des espèces (Costion *et al.* 2011; Gaikwad *et al.* 2011; Hausman *et al.* 2011). En effet, la taxonomie conventionnelle, surtout basée sur l'étude de critères morphologiques et écologiques pouvait parfois nécessiter plusieurs années pour étudier et nommer les espèces (Janzen and Hallwachs. 2011; Raupach *et al.* 2011). Grâce à ce nouveau moyen d'identification des espèces, plus rapide et plus aisé, le nombre de spécimens référencés s'est accru de façon importante (Janzen and Hallwachs. 2011). Dès lors, nomenclature plus rapide et plus aisée s'avère très bénéfique lorsque l'on désire appréhender la biodiversité totale présente dans un milieu et entreprendre ainsi des plans de conservation dans les plus brefs délais (Smith *et al.* 2005).

Aussi, la capacité potentielle du Barcoding à mieux délimiter, voire découvrir les espèces a été très souvent illustrée dans les études sur la biodiversité. En effet, nombreuses sont les études dans lesquelles le DNA-Barcoding est utilisé pour remettre en cause ou apporter des informations

complémentaires à certains groupes taxonomiques déjà établis (Taylor *et al.* 2012). Nous pouvons notamment citer des travaux intéressants chez certaines familles de lépidoptères (Hausmann *et al.* 2011; Janzen *et al.* 2005; Hebert *et al.* 2004b), ou de coléoptères (Raupach *et al.* 2010; Monaghan *et al.* 2005) dans lesquelles furent mises en évidence de potentielles espèces cryptiques.

Cet usage du DNA-Barcoding s'est vu notamment renforcé par le concept de la taxonomie intégrative (Dayrat, (2005); Rubinoff. 2006; DeSalle. 2006; Hebert *et al.* 2004b; Mutanen *et al.* 2012). La taxonomie intégrative est une technique qui vise à mieux décrypter et délimiter les frontières entre les espèces. Il s'agit en fait d'impliquer trois principaux critères : moléculaire (le gène COI, mais parfois d'autres marqueurs), écologique et morphologique propres à chaque espèce. Ce n'est pas un ou deux mais la combinaison de ces trois critères qui est utilisé pour distinguer les différentes espèces. Cette méthode, aussi connue sous le terme de *Three-way correlation*, a été maintes fois qualifiée comme étant le meilleur compromis permettant de caractériser les espèces ou mettre en évidence de nouvelles espèces jusqu'alors difficilement discernables (Janzen *et al.* 2011; Janzen *et al.* 2011; Hebert *et al.* 2004b).

- [La lutte contre les espèces invasives](#)

Les dommages que peuvent causer les espèces invasives sur l'environnement, l'économie et la santé humaine ont par le passé largement été démontrés et sont aujourd'hui on ne peut plus pris en considération (Armstrong and Ball, 2005). Dans ce contexte, l'utilisation du DNA-Barcoding pour contribuer à la lutte contre ces espèces invasives a montré un potentiel très convaincant.

Le DNA-Barcoding permet notamment la détection d'espèces non-autochtones grâce à une méthode pour le moins aussi efficace que triviale. En effet, dans un premier temps ceci consiste simplement à barcoder et à identifier les individus suspects, notamment grâce à la bibliothèque de références. Il s'agit ensuite d'établir si la localisation géographique des individus en question correspond bien à la répartition géographique de l'espèce pour laquelle ils ont été identifiés.

L'avantage ici est que le diagnostic est relativement rapide, aisé et peut se faire à n'importe quel stade de développement de l'animal notamment chez les insectes (Cross *et al.* 2010), les crustacés, les amphibiens etc.

Pour certaines espèces très proches, ces stades de développement sont très semblables et l'identification des espèces potentiellement invasives dans ce genre de situation s'avère parfois impossible. L'utilisation du DNA-Barcoding permet donc ici de résoudre efficacement le problème. Il a ainsi été possible de détecter certains œufs d'insectes comme appartenant à une espèce potentiellement invasive (Briski *et al.* 2011; Armstrong and Ball. 2005). Cela a aussi été le cas pour de très jeunes stades larvaire (Chown *et al.* 2008).

Cette détection rapide et efficace des espèces potentiellement invasives permettent ainsi de prendre des mesures de lutte dans les plus brefs délais (Cross *et al.* 2010).

Mieux encore, la capacité du DNA-Barcoding à traiter sans souci les individus présents dans les collections anciennes a permis d'apporter une perspective historique dans l'étude des espèces invasives (Cross *et al.* 2010). En effet, l'étude de la variation spatiale et temporelle du gène COI établie sur des individus provenant de vieilles collections a permis de déterminer dans certains cas la date d'introduction d'une espèce invasive ainsi que la vitesse à laquelle cette espèce s'est répandue (Pringle *et al.* 2009; Provan *et al.* 2008; Lees *et al.* 2011; Gleeson *et al.* 2000). Cette contribution du DNA-Barcoding est aujourd'hui considéré comme cruciale, notamment pour développer des stratégies de lutte (Lees *et al.* 2011).

- **Une implication à tous les niveaux**

Outre l'efficacité montrée dans l'aide à la préservation de la biodiversité ainsi que l'aide à la lutte contre les espèces invasives, le DNA-Barcoding fut utilisé pour résoudre une grande variété de problèmes. Nous pouvons citer entre autres la lutte contre le commerce illégal d'espèces menacées (Eaton *et al.* 2010), la lutte contre le commerce de bois abattu illégalement (Lowe *et al.* 2011), le suivi de denrées alimentaires (Rasmussen *et al.* 2009; Barbuto *et al.* 2010), mais aussi la contribution dans la lutte contre certains ravageurs (Greenstone *et al.* 2011) et bien d'autres encore. Aujourd'hui le DNA-Barcoding est même considéré comme une *entreprise mondiale* dont les enjeux ne cessent de se multiplier et se diversifier (Taylor and Harris. 2012).

Cependant, en dépit de toutes ces contributions, l'utilisation du DNA-Barcoding n'est pas sans monter certaines failles, lesquelles sont parfois considérées comme étant réellement problématiques.

IV) Le DNA-Barcoding : entre limites et innovations

Comme nous avons pu le constater jusqu'à présent, le DNA-Barcoding a séduit tant par sa simplicité que par son efficacité. Toutefois, dans le même temps, a été mis en avant de nombreux problèmes que cette méthode de la biologie moléculaire a pu faire émerger (Vences *et al.* 2005; Costion *et al.* 2011). Face à cela, nombreuses ont été les tentatives de procéder différemment en vue de contourner tous ces problèmes (Taylor and Harris, 2012) et assurer finalement un futur au DNA-Barcoding. Il va ainsi être détaillé par la suite quels sont les différents problèmes rencontrés lors de l'utilisation du DNA-Barcoding et, en définitive, comment est perçu l'avenir de cette discipline.

- Le gène COI

Comme nous avons pu l'évoquer au début de ce rapport, le gène COI fut considéré comme un des meilleurs gènes candidat pour aider à l'identification du monde vivant (Hebert *et al.* 2003a). Cependant, bien que ce gène ait en général montré une grande fiabilité au sein du règne animal, son efficacité à identifier les espèces a cependant été moindre chez certains groupes taxonomiques (Vences *et al.* 2005). Alors que certains auteurs considèrent ce gène COI comme le "gène du Barcoding" (Dove *et al.* 2008), d'autres sont en revanche moins enthousiasmés et estiment que ce terme est loin d'être approprié (Taylor and Harris, 2012). En effet, bien que ce gène ait fait face à certains problèmes dans le règne animal, cela est finalement négligeable face aux problèmes rencontrés chez les autres êtres vivants, notamment les mycètes, les végétaux et les protistes.

Pour les végétaux par exemple, ce gène COI s'est avéré inefficace, notamment par ce qu'au sein de ce règne son taux de divergence interspécifique est relativement faible et ne permet pas une identification convenable de la majorité des espèces (Kress and Erickson, 2007). Nombreuses ont été les tentatives pour remédier à ces problèmes, aboutissant parfois à des résultats assez convaincants (Costion *et al.*, 2011). Mais au final, certains estiment que pas moins de 41 fragments d'ADN différents ont été testés (Taylor and Harris, 2012), concluant ainsi que la majorité des études sur les plantes n'ont fait que chercher un marqueur approprié plutôt qu'agréments les bibliothèques de références de barcodes végétaux. Ainsi, les données établies concernant les végétaux sont relativement pauvres, ceci constituant un frein inévitable à leur étude.

Chez d'autres organismes tels que les protistes, des problèmes ont été rencontrés au niveau des amorces permettant le séquençage du gène COI. Dans ce cas aussi, pour remédier à ces problèmes, des séquences autres que celle du gène COI ont été testées. Bien que parmi ces séquences testées, certaines aient été approuvées, elles présentaient cependant une faible divergence interspécifique pouvant s'avérer problématique pour la discrimination des espèces au sein de ce règne (Clarkston and Saunders 2010).

Nous pouvons aussi constater la même chose chez les mycètes. En effet, le gène COI s'est aussi avéré inapproprié pour l'étude de ce règne (Eberhardt 2010) et finalement, aucun gène ne s'est pour l'instant avéré réellement efficace.

Bien que l'utilisation du gène standard COI soit un concept majeur du DNA-Barcoding, nous avons pu constater ici que l'étude de certains taxons par le biais de ce gène n'était finalement pas envisageable. La résolution de ce problème s'avère cependant indispensable dès lors que l'on désire barcoder et étudier toutes les formes de vie présentes dans un milieu (Taylor and Harris, 2012). Pour autant, la méthodologie du DNA-Barcoding a connu bien d'autres problèmes.

- **La bibliothèque de référence**

Comme nous avons pu le mettre en évidence jusqu'ici, la bibliothèque de référence a montré une efficacité notoire dans l'aide à l'identification d'espèces préalablement décrites. Cependant, parmi certains taxons, de nombreux individus séquencés n'ont pu être identifiés via la bibliothèque de référence ou semblent appartenir à une espèce jusqu'alors non décrite (Seifert 2009). Selon certains auteurs, ce nombre croissant d'individus non-identifiés renforcerait finalement le besoin d'une expertise taxonomique qualifiée, ce qui au final, serait contraire à l'un des objectifs principaux du DNA-Barcoding (Packer *et al.* 2009).

- **Analyse, utilisation du DNA-Barcoding**

Dans certains cas, la manière d'analyser les données séquencées chez certains auteurs a été sujette à de nombreuses critiques. Notamment, certaines études ont fait remarquer que le seuil de divergence ou "threshold", permettant de discriminer les espèces semblait avoir été établi arbitrairement (Prendini, 2005; Vogler, 2006). Or, il est très important de bien déterminer ce seuil au sein du taxon que l'on étudie, en particulier afin de discriminer correctement les différentes espèces parmi ce groupe d'organismes. De ce fait, certains efforts ont récemment été réalisés pour tenter de remédier à ce problème (Puillandre *et al.* 2011).

Un autre point souvent remis en question à propos du DNA-Barcoding concerne ses différents usages. Nous avons pu voir jusqu'à présent que le DNA-Barcoding était considéré comme étant non seulement un outil permettant d'identifier les organismes, mais aussi un outil permettant de découvrir et délimiter les espèces. Si la capacité du DNA-Barcoding à identifier les organismes s'est révélée souvent très efficace, son utilisation pour délimiter et découvrir potentiellement les espèces a en revanche été de très nombreuses fois controversée (DeSalle, 2006). Il a notamment été mis en avant le fait que certaines espèces avaient été décrites uniquement sur les bases du DNA-Barcoding, lequel n'étaient ni corroborées par des données morphologiques, ni par des données écologiques (DeSalle, 2006). En l'absence de congruences au niveau morphologique et écologique, le statut de ces nouvelles espèces est alors souvent considéré comme hypothétique (Goldstein and DeSalle, 2011) et requièrent des études supplémentaires, ceci ne contribuant finalement que très partiellement à l'avancée des connaissances sur le taxon en question. L'utilisation des données moléculaires seules a cependant été justifiée pour aider à une découverte plus rapide des espèces. En effet, les hypothétiques espèces découvertes à partir de ces données moléculaires seraient prises en compte et fourniraient des informations potentiellement utiles pour des études futures (Kress and Erickson 2008).

Ces controverses ont alors fait naître deux concepts différents : le *DNA-Barcoding* dont le principe de base consiste en l'aide à l'identification des espèces, et la *DNA-Taxonomy* qui apporterait une contribution dans la délimitation et la découverte des espèces (DeSalle 2007).

Malgré ces problèmes rencontrés, le DNA-Barcoding reste malgré tout un outil encore très utilisé de nos jours. Cet outil, dont les limites ont pu être mises en avant, est finalement aujourd'hui en constante optimisation. La naissance de nouvelles techniques de la biologie moléculaire pourraient permettre par la suite au DNA-Barcoding d'avoir un succès encore plus grand dans l'accès à l'identification du monde vivant.

- **Le NGS : Next Generation Sequencing**

Certains auteurs estiment que le futur du DNA-Barcoding pourrait notamment se voir influencé par le NGS (Taylor and Harris, 2012). Le NGS, comme son nom l'indique, résulte des toutes dernières techniques de séquençages du génome lesquelles sont bien plus rapides et efficaces que les techniques de séquençage actuelles (Shendure and Hanlee, 2008; Valentini et al. 2009). Bien qu'il soit encore relativement coûteux, le NGS permettrait notamment de réaliser le séquençage du gène COI et ce, pour la totalité des organismes présents dans certains micro-habitats (Bergeron et al, 2007). Il peut ainsi notamment être estimé la biodiversité en micro-organismes présents dans les échantillons de sol, ou dans les échantillons d'eau (Bik et al, 2012).

Le NGS permettrait aussi la réalisation du génome entier des organismes dans un temps imparti relativement court (Valentini et al. 2009). Ceci montrerait aussi un très grand intérêt, notamment par le fait que l'étude du génome entier serait davantage représentative des variations entre les espèces et permettrait ainsi de les discriminer avec plus de certitude (Humphries and Winker 2011). Cependant, la réalisation de génomes entiers générerait un nombre considérable de données à analyser et interpréter. Les outils de la bio-informatiques dont nous disposons actuellement ne seraient d'ailleurs certainement pas en mesure de pouvoir les analyser.

V) Conclusion

Dès l'année 2003, cette nouvelle approche moléculaire qu'est le DNA-Barcoding a permis d'améliorer nettement l'accès à l'identification du monde vivant, en particulier au sein du règne animal. Notamment parce que protéger la biodiversité repose avant tout sur l'identification et la connaissance des espèces, le DNA-Barcoding fut accueilli et appliqué avec un enthousiasme certain. En effet, auparavant l'étude de la biodiversité présente dans certains milieux nécessitait énormément de temps et des connaissances hautement spécialisées. La connaissance des zones les plus riches en biodiversité, où s'épanouissent les forêts ombrophiles, était d'ailleurs évoquée d'un point de vue assez pessimiste, tant il reste à découvrir parmi ces régions du globe. Le DNA-Barcoding a finalement été vue ici comme étant une des solutions permettant d'améliorer considérablement l'étude de ces milieux, aussi complexes soient-ils.

Cependant, l'enthousiasme important qu'a connu le DNA-Barcoding lui a peut être finalement porté préjudice, en particulier par le fait que peu de temps ait été consacré pour tenter d'améliorer les différents problèmes que cette nouvelle approche a pu rencontrer. Comme nous avons pu le voir, ces problèmes aussi bien analytiques que méthodologiques ont donné lieu à la suggestion de solutions alternative chez certains auteurs (Puillandre et al. 2011; Costion et al, 2011; DeSalle et al. 2005). Aujourd'hui malgré l'urgence des études à réaliser, notamment les études visant à enrayer l'érosion de la biodiversité, certains auteurs suggèrent qu'il serait davantage judicieux de trouver en priorité des solutions à tous ces problèmes rencontrés (Taylor and Harris, 2012). Malgré tout, le DNA-Barcoding continu à avoir un franc succès et semble s'implémenter de plus en plus au sein des études scientifiques. La naissance des nouvelles techniques de séquençage telle que le NGS, ne font que rendre le DNA-Barcoding d'autant plus passionnant.

VI) Bibliographie

1. Armstrong KF, Ball SL (2005) DNA barcodes for biosecurity: invasive species identification. *Phil Trans R Soc B* 360: 1813–1823.
2. Ball SL, Hebert PDN, Burian SK, Webb JM (2005) Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. *Journal of the North American Benthological Society*, 24, 508–524
3. Barbuto M, Galimberti A, Ferri E et al. (2010) DNA barcoding reveals fraudulent substitutions in shark seafood products: the Italian case of “palombo” (*Mustelus* spp.). *Food Research International*, 43, 376–381.
4. Bennett, Kyle F; Reed, Andrew; Lutz, Richard A. (2011). DNA barcoding reveals *Brachidontes* (Bivalvia: Mytilidae) from two ecologically distinct intertidal habitats on Long Key, Florida Keys, are cryptic species, not ecotypes. *NAUTILUS*. Vol. : 125 iss : 2 pg : 63 -71
5. Bergeron A, Belcaid M, Steward GF, Poisson G (2007) Divide and Conquer: Enriching Environmental Sequencing Data. *PLoS ONE* 2(9): e830. doi:10.1371/journal.pone.0000830
6. Bik, H. M., Sung, W., De Ley, P., Baldwin, J. G., Sharma, J., Rocha-Olivares, A. and Thomas, W. K. (2012), Metagenetic community analysis of microbial eukaryotes illuminates biogeographic patterns in deep-sea and shallow water sediments. *Molecular Ecology*, 21: 1048–1059. doi: 10.1111/j.1365-294X.2011.05297.x
7. Blaxter M (2003) Molecular systematics – counting angels with DNA. *Nature*, 421, 122–124.
8. Blaxter, M. L. (2004) The promise of a DNA taxonomy. *Phil. Trans. R. Soc. B* 359, 669–679. (doi:10.1098/rstb.2003.1447.)
9. Boykin Laura M., Armstrong Karen F., Kubatko Laura and De Barro Paul. (2012) Species Delimitation and Global Biosecurity. *Evolutionary Bioinformatics* : 8, 1–37
10. Briski, E; Cristescu, ME; Bailey, SA; MacIsaac, HJ. (2011) Use of DNA barcoding to detect invertebrate invasive species from diapausing eggs. *Biol Invasions* 13:1325–1340 (DOI 10.1007/s10530-010-9892-7)
11. Chown, S., Sinclair, B. & van Vuuren, B. (2008) DNA barcoding and the documentation of alien species establishment on sub - Antarctic Marion Island. *Polar Biology*, 31, 651 – 655.
12. Clarkston BE, Saunders GW (2010) A comparison of two DNA barcode markers for species discrimination in the red algal family Kallymeniaceae (Gigartinales, Florideophyceae), with a description of *Euthora timburtonii* sp nov. *Botany-Botanique*, 88, 119–131.
13. Costion C, Ford A, Cross H, Crayn D, Harrington M, et al. (2011) Plant DNA Barcodes Can Accurately Estimate Species Richness in Poorly Known Floras. *PLoS ONE* 6(11): e26841. (doi:10.1371/journal.pone.0026841)

14. Cox, A. J. & Hebert, P. D. N. 2001 Colonization, extinction and phylogeographic patterning in a freshwater crustacean. *Mol. Ecol.* 10, 371–386.
15. Cross HB, Lowe AJ, Gurgel CF. (2010) DNA barcoding of Invasive Species. In Richardson D, ed. *Fifty years of invasion ecology: The legacy of Charles Elton* (Ed. Richardson D.), Chapter 22. Blackwells.
16. Dayrat, B. (2005), Towards integrative taxonomy. *Biological Journal of the Linnean Society*, 85: 407–415. doi: 10.1111/j.1095-8312.2005.00503.x
17. DeSalle R. (2006) Species Discovery versus Species Identification in DNA Barcoding Efforts: Response to Rubinoff. *Conservation Biology* Volume 20, No. 5, 1545–1547 Society for Conservation Biology (DOI: 10.1111/j.1523-1739.2006.00543.x)
18. DeSalle R (2007) Phenetic and DNA taxonomy; a comment on Waugh. *Bioessays*, 29, 1289–1290.
19. Dove CJ, Rotzel NC, Heacker M, Weigt LA (2008) Using DNA barcodes to identify bird species involved in birdstrikes. *Journal of Wildlife Management*, 72, 1231–1236.
20. Eaton MJ, Meyers GL, Kolokotronis SO, Leslie MS, Martin AP, et al. (2010) Barcoding bushmeat: molecular identification of Central African and South American harvested vertebrates. *Conserv Genet* 11: 1389–1404.
21. Eberhardt U (2010) A constructive step towards selecting a DNA barcode for fungi. *New Phytologist*, 187, 266–268.
22. Gaikwad S. S., Ghate H. V., Ghaskadbi S. S., Patole M. S., Shouche Y. S. (2010) DNA barcoding of nymphalid butterflies (Nymphalidae: Lepidoptera) from Western Ghats of India. *Mol Biol Rep* DOI 10.1007/s11033-011-0988-7
23. Gibson WC; Dukes P; Gashumba JK. (1988) Species-Specific DNA Probes for the Identification of African Trypanosomes in Tsetse Flies. *Parasitology*, 97, 63-73.
24. Gleeson D, Holder P, Newcomb R, Howitt R, Dugdale J (2000) Molecular phylogenetics of leafrollers: application to DNA diagnostics. *New Zealand Plant Protection*, 53, 157–162.
25. Greenstone, MH; Vandenberg, NJ; Hu, JH. (2011) Barcode haplotype variation in north American agroecosystem lady beetles (Coleoptera: Coccinellidae)
26. Goldstein PZ, DeSalle R (2011) Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. *Bioessays*, 33, 135–147.
27. Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences, USA*, 103, 968–971.
28. Halloran, S. Michael. (1984) *The Birth of Molecular Biology: An Essay in the Rhetorical Criticism of Scientific Discourse*. *Rhetoric Review* 3.1, 70-83. Print.

29. Hausmann A, Haszprunar G, Hebert PDN (2011) DNA Barcoding the Geometrid Fauna of Bavaria (Lepidoptera) : Successes, Surprises, and Questions. *PLoS ONE* 6(2): e17134. doi:10.1371/journal.pone.0017134
30. Hebert Paul D. N., Cywinska Alina, Ball Shelley L. and DeWaard Jeremy R. (2003a). Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B* 270, 313–321 DOI 10.1098/rspb.2002.2218
31. Hebert Paul D. N.; Ratnasingham Sujeevan; De Waard Jeremy R. (2003b). Barcoding Animal Life: Cytochrome c Oxidase Subunit 1 Divergences among Closely Related Species. *Biological Sciences*, Vol. 270, pp.S96-S99
32. Hebert PDN, Stoeckle MY, Zemplak TS, Francis CM (2004a) Identification of birds through DNA barcodes. *Public Library of Science Biology*, 2, 1657–1663.
33. Hebert P.D.N, Penton E.H, Burns J.M, Janzen D.H, Hallwachs W. (2004b). Ten species in one: DNA barcoding reveals cryptic species in the Neotropical skipper butterfly *Astraptes fulgerator*. *Proc. Natl Acad. Sci.* 101, 14812–14817.
34. Humphries EM, Winker K (2011) Discord reigns among nuclear, mitochondrial and phenotypic estimates of divergence in nine lineages of trans-Beringian birds. *Molecular Ecology*, 20, 573–583.
35. Janzen DH (2004) Now is the time. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 359, 731–732.
36. Janzen, DH; Hajibabaei, M; Burns, JM; Hallwachs, W; Remigio, E; Hebert, PDN. (2005) Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Phil. Trans. R. Soc. B* 360, 1835–1845 doi:10.1098/rstb.2005.1715
37. Janzen DH, Hallwachs W (2011) Joining Inventory by Parataxonomists with DNA Barcoding of a Large Complex Tropical Conserved Wildland in Northwestern Costa Rica. *PLoS ONE* 6(8): e18123. doi:10.1371/journal.pone.0018123
38. Janzen Daniel H; Hallwachs Winnie; Burns John M; Hajibabaei, Mehrdad; Bertrand, Claudia; Hebert, Paul D N. (2011) Reading the complex skipper butterfly fauna of one tropical place. *PLoS ONE* 6(8): e19874. doi:10.1371/journal.pone.0019874
39. Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding *rbcl* gene complements the non-coding *trnH-psbA* spacer region. *PLoS ONE*, 2, e508.
40. Laudénbach V., Mantz J., Desmots J.M. (1999) To Understand Molecular Biology. *Annales Françaises d'Anesthésie et de Réanimation*, Volume 18, Issue 7, Pages 725-747.
41. Lees, DC; Lack, HW; Rougerie, R; Hernandez-Lopez, A; Raus, T; Avtzis, ND; Augustin, S; Lopez-Vaamonde, C. (2011) Tracking origins of invasive herbivores through herbaria and archival DNA: the case of the horse-chestnut leaf miner. *European Journal of Entomology*. Volume: 109 Issue: 1 Pages: 1-6
42. Lowe AJ, Cross HB (2011) The Application of DNA to Timber Tracking and Origin Verification. *Int Assoc Wood Anat J* 32: 251–262.

43. Lunardi J. (2003) Fifty years ago, the double helix gave birth to the molecular biology. *Ann Biol Clin* 2003, 61 : 623-33.
44. Meier R, Kwong S, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Syst Biol* 55: 715–728.
45. Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *Plos Biology*, 3, 2229–2238.
46. Miller KB, Alarie Y, Wolfe GW, Whiting MF (2005). Association of insect life stages using DNA sequences: the larvae of *Philodytes umbrinus* (Motschulsky) (Coleoptera: Dytiscidae). *Syst Entomol*, 30:499-509
47. Monaghan, MT; Balke, M; Gregory, TR; Vogler, AP. (2005) DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. *Phil. Trans. R. Soc. B* 360, 1925–1933
doi:10.1098/rstb.2005.1724
48. Moritz, C., and Cicero, C. (2004). DNA barcoding: promise and pitfalls. *PLoS Biology* 2, e354.
doi:10.1371/journal.pbio.0020354
49. Mutanen Marko, Aarvik Leif, Huemer Peter, Kaila Lauri, Karshlot Ole & Tuck Kevin. (2012) DNA barcodes reveal that the widespread European tortricid moth *Phalonidia manniana* (Lepidoptera: Tortricidae) is a mixture of two species. *Zootaxa* 3262: 1–21
50. Narsinh L. Thakur, Roopesh Jain, Filipe Natalio, Bojan Hamer, Archana N. Thakur, Werner E.G. Müller. (2008) Marine molecularbiology: An emerging field of biological sciences. Volume 26, Issue 3, Pages 233 - 245.
51. Packer L, Grixti JC, Roughley RE, Hanner R (2009) The status of taxonomy in Canada and the impact of DNA barcoding. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, 87, 1097–1110.
52. Pringle, A., Adams, R.I., Cross, H.B. & Bruns, T.D. (2009) The ectomycorrhizal fungus *Amanita phalloides* was introduced and is expanding its range on the west coast of North America. *Molecular Ecology*, 18, 817 – 833.
53. Provan, J., Booth, D., Todd, N.P., Beatty, G.E. & Maggs, C.A. (2008). Tracking biological invasions in space and time: elucidating the invasive history of the green alga *Codium fragile* using old DNA. *Diversity and Distributions*, 14 ,343 – 354.
54. Puillandre N, Lambert A, Brouillet S, Achaz G (2011) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, doi: 10.1111/j.1365-294X.2011.05239.x.
55. Ragupathy, S., Newmaster, S. G., Murugesan, M. and Balasubramaniam, V. (2009) DNA barcoding discriminates a new cryptic grass species revealed in an ethnobotany study by the hill tribes of the Western Ghats in southern India. *Molecular Ecology Resources*, 9: 164–171. doi: 10.1111/j.1755-0998.2009.02641.x
56. Rasmussen RS, Morrissey MT, Hebert PDN (2009) DNA barcoding of commercially important salmon and trout species (*Oncorhynchus* and *Salmo*) from North America. *Journal of Agricultural and Food Chemistry*, 57, 8379–8385.

57. Raupach Michael J, Astrin Jonas J, Hannig Karsten, Peters Marcell K, Stoeckle Mark Y, Wägele Johann-Wolfgang. (2010) Molecular species identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear rDNA expansion segments and DNA barcodes. *Frontiers in Zoology* 2010, 7:26
58. Ratnasingham Sujeevan and Hebert Paul D. N. (2007) BOLD : The Barcode of Life Data System. *Molecular Ecology Notes* : 7, 355–364
59. Riemann Lasse, Alfredsson Hanna, Hansen Michael M, Als Thomas D, Nielsen Torkel G, Munk Petet, Aarestrup Kim, Maes Gregory E, Sparholt Henrik, Petersen Michael I, Bachler Mirjam and Castonguay Martin. (2010) Qualitative assessment of the diet of European eel larvae in the Sargasso Sea resolved by DNA barcoding. *Biol. Lett.* vol. 6 no. 6 819-822
60. Rubinoff D (2006) *Conservation Biology* Volume 20, No. 5, 1548–1549 Society for Conservation Biology DOI: 10.1111/j.1523-1739.2006.00542.x
61. Seifert KA (2009) Progress towards DNA barcoding of fungi. *Molecular Ecology Resources*, 9, 83–89.
62. Shendure Jay; Ji Hanlee. Next-generation DNA sequencing. (2008). *Nat Biotech.* 26, 1135 – 1145.
63. Smith MA, Fisher BL, Hebert PDN (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philos Trans R Soc Lond Biol Sci* 360:1825–1834
64. Saitou N, Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*;4(4):406-25.
65. Takeyama, H; Chow, S; Tsuzuki; Matsunaga, T. (2000) Mitochondrial DNA sequence variation within and between tuna *Thunnus* species and its application to species identification. *Journal of Fish Biology* (2001) 58, 1646–1657 doi:10.1006/jfbi.2001.1575
66. Tänzler R, Sagata K, Surbakti S, Balke M, Riedel A (2012) DNA Barcoding for Community Ecology - How to Tackle a Hyperdiverse, Mostly Undescribed Melanesian Fauna. *PLoS ONE* 7(1): e28832. doi:10.1371/journal.pone.0028832
67. Tautz D, Arctander P, Minelli A, Thomas R H & Vogler A P. (2003) A plea for DNA taxonomy. *Trends Ecol. Evol.* 18, 70–74. (doi:10.1016/S0169-5347(02) 00041-1.)
68. Taylor H. R. and Harris W. E. (2012) An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Molecular Ecology Resources*, doi: 10.1111/j.1755-0998.2012.03119.x
69. Vences M, Thomas M, Bonett RM, Vieites DR (2005) Deciphering amphibian diversity through DNA barcoding: chances and challenges. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 360, 1859–1868.
70. Vogler AP, Monaghan MT (2007) Recent advances in DNA taxonomy. *Journal of Zoological Systematics and Evolutionary Research*, 45, 1–10.
71. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 360, 1847–1857.

72. Wares, J. P. & Cunningham, C. W. (2001) Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution* 12, 2455–2469.
73. Watson J.D. and Crick F.H.C. (1953) A Structure for Deoxyribose Nucleic Acid *Nature* 171, 737-738.



*Estimation de la richesse spécifique des lépidoptères
des Nouragues (Guyane) via le DNA-Barcoding*

Encadrants : Carlos Lopez Vaamonde,
Rodolphe Rougerie

Année 2011 - 2012



Remerciements

Je tenais avant tout à attribuer mes remerciements les plus sincères à toutes les personnes ayant contribué à l'élaboration de cette étude et de ce rapport, notamment :

Carlos Lopez Vaamonde, pour sa gentillesse, son écoute et pour m'avoir encadré et soutenu sans cesse durant ces six mois de stage.

Rodolphe Rougerie, pour son implication, sa gentillesse et pour m'avoir apporté une aide de loin non-négligeable dans la rédaction de ce rapport.

David Lees et Thibaud Decaens, pour m'avoir fait généreusement part de leur connaissances tant au niveau analytique que méthodologique.

Alain Roques ainsi que tous les membres de l'Unité de Recherche de Zoologie Forestière pour m'avoir chaleureusement accueilli au sein du centre de l'INRA (Orléans).

Je souhaiterais remercier également tous les taxonomistes ayant contribué à la réalisation de cette étude, à savoir : **David Adamski, Jérôme Barbut, Don Davi, Jay Sohn, Cees Gielis, Michel Laguerre, Bernard Landry, Antoine Leveque, Joel Minet, Paul Thiaucourt.**

Merci également au **CNRS** (programme "Amazonie") pour avoir financé les trois missions de collectes réalisées.

Sommaire

I) Introduction.....	1
II) Matériels et Méthodes.....	3
A) Milieu d'étude : La Réserve Naturelle des Nouragues.....	3
B) Etude de la richesse spécifique des lépidoptères.....	4
III) Résultats.....	9
A) Identification des individus collectés.....	9
B) Etude de la richesse spécifique des lépidoptères.....	11
IV) Discussion.....	18
V) Bibliographie.....	20
VI) Annexes.....	24

I) Introduction

Comme définie lors de l'acte du congrès américain sur la biodiversité en 1990 (HR1268, 1990), la diversité biologique telle que nous l'entendons se traduit par la variabilité des organismes au sein des espèces (diversité génétique), entre les espèces (diversité spécifique) ainsi qu'entre les écosystèmes (diversité écosystémique). Aujourd'hui, l'érosion de la biodiversité principalement due, à la destruction des habitats, au changement climatique, à l'introduction d'espèces invasives ainsi que les innombrables impacts que cela engendre ont été largement admis (Stuart Chapin III. *et al*, 2000). Il a en effet été démontré maintes fois que cette érosion constante a d'importantes et nombreuses conséquences sur les écosystèmes ainsi que le bien-être et la santé de l'homme (Proença *et al*, 2011 ; Hammen *et al*, 2011). A l'heure où le taux d'extinction des espèces est 100 à 1000 supérieur à la normale (Lawton *et al*, 2005), l'étude de la biodiversité constitue donc un véritable challenge. Dans une grande majorité d'études, l'analyse de la biodiversité est effectuée par la détermination de la richesse spécifique au sein d'une aire géographique donnée qui comprend un ou plusieurs écosystèmes distincts. Cette étude de la richesse spécifique consiste principalement à collecter les êtres vivants dans un lieu donné pour ensuite les identifier et les classer (Moritz and Cicero, 2004). Les taxonomistes jouent ici un rôle des plus prépondérants car ce sont eux qui, depuis de nombreuses décennies, de par leurs collectes, leur expertise, leurs publications fournissent une base de données qui s'avère indispensable dès lors que l'on entreprend l'étude la biodiversité (Janzen & Hallwachs, 2011). L'expertise taxonomique peut s'avérer très coûteuse en temps (Costion *et al*, 2011) et ainsi plusieurs années, voire plusieurs décennies sont nécessaires pour identifier et décrire les espèces présentes sur le terrain. Nous pouvons par exemple citer l'équipe de taxonomistes menée par Daniel Janzen au Costa Rica (Janzen *et al*, 2011), qui durant plus de trente années, a entrepris de collecter et d'élever de nombreuses chenilles en vue de caractériser les stades larvaires de nombreuses espèces ou encore d'établir les relations insectes/parasitoïdes. Depuis des décennies, les espèces sont principalement définies et décrites sur des bases morphologiques, écologiques ou encore comportementales (Cook *et al*, 2010). Dans la grande majorité des cas, le taxonomiste possède des connaissances très pointues sur les espèces étudiées (Raupach *et al*, 2010) et est souvent spécialisé dans un taxon particulier (Ordre, Famille).

Nous pouvons cependant ici mettre en évidence certains facteurs limitant de l'étude de la biodiversité. Premièrement, comme évoqué précédemment, nous pouvons noter que l'expertise taxonomique requiert de nombreuses années de formation pour bien connaître les groupes d'organismes à étudier. Ceci, à l'heure où la biodiversité s'érode de plus en plus rapidement, s'avère

problématique. D'ailleurs, le fait que certaines espèces se seront éteintes avant même leur découverte est aujourd'hui largement reconnu (Hebert *et al.*, 2004).

Deuxièmement, il a été noté que cette expertise taxonomique s'est considérablement raréfiée au cours des dernières décennies. Ce déclin constant des taxonomistes a pour conséquence de ralentir notablement le travail d'identification et de description des espèces et est ainsi considéré comme une véritable entrave à l'étude et à la gestion de la biodiversité (Cook *et al.* 2010 ; Gaikwad *et al.* 2011 ; Wilson, 2011 ; Tanzler *et al.* 2012). Finalement, cette expertise est très souvent spécialisée à un taxon particulier et ne permet pas une identification exhaustive des espèces (Gaikwak *et al.* 2011). Dans ce contexte, pour répondre à ces insuffisances de la taxonomie traditionnelle, fut proposé il y a quelques années un outil d'identification global des espèces basé sur des analyses moléculaires et baptisé DNA barcoding ou « codes-barres ADN » (Hebert *et al.*, 2003). L'idée est de choisir un fragment standard de l'ADN (le gène COI pour le règne animal) dont la variabilité nucléotidique constitue des combinaisons uniques permettant de caractériser chaque espèce. Ces combinaisons nucléotidiques représentent donc un "code" qui est unique pour chaque espèce, tout comme les codes-barres présents dans les systèmes de distribution sont caractéristiques de chaque produit de vente.

La connaissance des milieux les plus riches en biodiversité, dans lesquels s'épanouissent les forêts ombrophiles, est souvent évoquée d'un point de vue assez pessimiste, tant ces habitats abritent des communautés d'espèces très complexes et très diversifiées. Or, comme nous avons pu l'évoquer précédemment, cette connaissance s'avère indispensable dès lors que l'on désire préserver cette biodiversité (Devries *et al.* 2009). Le DNA-Barcoding a ainsi été vu comme étant une des solutions permettant d'améliorer considérablement l'étude et la préservation de ces milieux, aussi complexes soient-ils et fut donc accueilli avec un enthousiasme certain.

Dans ce contexte, les objectifs de ce stage effectué au sein de l'Unité de Zoologie Forestière de l'Institut National de la Recherche Agronomique à Orléans (INRAURZF), ont été de :

- (1) Participer au développement d'une librairie de code-barres ADN pour les lépidoptères de la réserve naturelle des Nouragues (Guyane Française) en zone intertropicale.
- (2) Tester la correspondance entre les groupes génétiques (« clusters » de barcodes) et la taxonomie traditionnelle (basée sur la morphologie) pour les familles dont les espèces sont connues.
- (3) Caractériser la richesse spécifique des lépidoptères de la région des Nouragues à l'aide d'analyses statistiques.

L'étude des lépidoptères dans cette région du globe est particulièrement intéressante, notamment parce que la Guyane abrite un massif forestier admirablement bien préservé dont la biodiversité

estimée est aussi importante que celle du grand ensemble amazonien (Ter Steege *et al.* 2000) mais aussi parce que les lépidoptères constituent un groupe extrêmement diversifié et tout à fait représentatif de cette biodiversité. Bien que le DNA-Barcoding ait permis la révision taxonomique de quelques groupes de lépidoptères guyanais (Hebert *et al.*, 2004; Janzen & Hallwachs 2011; Vaglia *et al.* 2008), l'estimation exhaustive de leur biodiversité à l'aide de ce nouvel outil n'a finalement pas encore été envisagée pour le moment. Cette évaluation de la richesse spécifique en lépidoptères nous permettrait de mettre en avant l'intérêt patrimonial de la Guyane et contribuerait ainsi à la connaissance et finalement à la préservation de la biodiversité présente dans cette région.

II) Matériels et Méthodes

A) Milieu d'étude : La Réserve Naturelle des Nouragues (Guyane)

La Guyane se caractérise avant tout par une imposante forêt ombrophile qui occupe 97% du territoire (DAF 1999, <http://www.cr-guyane.fr>) soit environ 2.3 millions de km² (Gonda *et al.*, 2011). Cet écosystème forestier bénéficie d'un climat tropical tempéré par les alizées, lequel fournit des précipitations annuelles de 1500 à 4000 mm et des températures oscillant de 25°C à 30°C (Hammond, 2005; Gonda *et al.*, 2011). Comme évoqué précédemment, l'étude qui a été menée a plus particulièrement porté sur la réserve naturelle des Nouragues. Cette zone protégée, avec ses 105 000 hectares gérés par l'Office National des Forêts (ONF), constitue une des plus grandes réserves de la Guyane. Après avoir connu par le passé une certaine fréquentation humaine, la zone est aujourd'hui complètement inhabitée et les perturbations d'origine anthropiques sont quasi-nulles (Sarhou *et al.* 2009; Charles-Dominique *et al.* 1998). Ainsi, la région est essentiellement dominée par une végétation primaire dont la canopée peut atteindre 30 à 45 mètres de haut.

Toutefois, la réserve des Nouragues se caractérise aussi et avant tout par une topographie particulière : cette zone est en effet surplombée par une formation granitique très ancienne qui culmine à 410 mètres d'altitude (Sarhou *et al.* 2009). Cette formation granitique, aussi appelée *Inselberg des Nouragues* (de l'allemand «île montagne»), crée ainsi un gradient altitudinal où les conditions environnementales varient en conséquence au point d'être totalement différentes des conditions environnementales présentes sous le couvert végétal. En effet, par exemple, au sommet de l'*Inselberg*, l'hygrométrie et la température peuvent varier énormément au cours de la journée (respectivement 100% à 20% et 18°C à 55°C) (Sarhou *et al.* 2009). Ces conditions environnementales singulières constituent des habitats atypiques de la région dans lesquels de nombreuses espèces animales et végétales y sont aujourd'hui inféodées et donc endémiques (Gasc *et al.* 1998; Sarhou *et al.* 2003). Au niveau végétal par exemple, ces habitats se caractérisent par des formations tout à fait particulières comme la "*Savane-roche*" qui se compose, d'amont en aval, de simples touffes d'herbes

(laïches, graminées, orchidées, ananas sauvages comme le pitcairnia), d'arbustes et d'une forêt basse (Charles-Dominique *et al.* 1998).

En 1986, l'Inselberg et ses alentours ont d'ailleurs été choisis pour accueillir une station scientifique (CNRS) qui attire aujourd'hui de plus en plus de chercheurs. A huit kilomètres de cette station, soit six heures de marche, se situe la seconde station scientifique des Nouragues appelée Saut Pararé/COPAS. Le site, créé en 1979 par le Muséum national d'histoire naturelle, a ici été établi sur un affluent de l'Approuague: l'Arataï. On retrouve sur ce site, situé à 29 mètres d'altitude, une végétation dense typique du climat tropical.

La réserve naturelle des Nouragues concentre un nombre tel d'espèces animales et végétales qu'elle fait partie d'un des 35 hot-spot mondiaux de biodiversité (Simona *et al.*, 2009; Sarthou *et al.* 2008). La biodiversité de ce site naturel, est alors considérée comme ayant une très grande valeur patrimoniale (Parmentier, 2005). C'est ainsi que notre intérêt s'est porté sur l'étude de ce site exceptionnel en concentrant plus particulièrement notre travail sur les lépidoptères.

B) Etude de la richesse spécifique des lépidoptères

Avec à ce jour pas moins de 157 524 espèces décrites (Nieukerken *et al.* 2011), les lépidoptères constituent un des groupes d'insectes les plus diversifiés. Bien que les zones néotropicales comptent aujourd'hui 44 791 espèces décrites (Kristensen *et al.* 2007), on estime qu'au moins 230 000 espèces restent encore à inventorier (www.lepsys.eu). Cependant, comme chez de nombreux groupes animaux et végétaux, le taux d'extinction des espèces de lépidoptère semble bien supérieur à leur taux de description actuel. L'urgence de mettre de nouveaux moyens en place pour les inventorier plus rapidement est donc aujourd'hui bien réelle.

Collectes sur le terrain

Pour réaliser cette étude nous avons à disposition un échantillon composé de 7610 lépidoptères. Cet échantillon provenait de collectes préalablement réalisées sur le terrain selon une méthodologie bien définie.

La collecte des lépidoptères nocturne (ou hétérocères) a été réalisée via l'utilisation de pièges lumineux. Ces pièges lumineux se composaient de lampes à vapeur de mercure (d'une puissance totale s'échelonnant de 125 à 250 watts) lesquelles éclairaient un drap blanc d'une surface assez d'environ 4 m². Ce piégeage lumineux a été réalisé chaque nuit (entre 18h30 et 6h30 du matin) pendant toute la durée de la mission sur le terrain et la date précise de collecte (jj/mm/aa), le nom du collecteur, ainsi que les données GPS ont été notées pour chaque individu capturé.

En parallèle à cela, des campagnes de collectes ont aussi été réalisées le jour pour capturer les rhopalocères (via la chasse à vue) mais aussi pour ramasser les larves d'une famille particulière de lépidoptères, à savoir les Gracillariidae. Les larves appartenant à cette famille se développent à l'intérieur même des feuilles et laissent apparaître par la suite des galeries bien visibles par transparence. La collectes de ces larves a donc consisté à ramasser les feuilles pour lesquelles étaient présentes ces galeries. Cette collecte des larves fut réalisée selon 3 transects principaux : (a) un transect qui débutait à l'Inselberg et qui s'étendait jusqu'au site de Parare, (b) un transect réalisé sur l'Inselberg même et (c) un transect réalisé à mi-haut du sommet de cet Inselberg (zone aussi appelée "Terrasse").

Au total, ces campagnes de collecte furent réalisées en trois missions distinctes qui ont été effectuées à des périodes différentes de l'année:

(1) Lors de la première mission, les lépidoptères furent collectés durant 19 jours consécutifs, lesquels s'étaient du 15 Janvier 2010 au 02 Février 2010. Cette première collecte fut réalisée par trois chercheurs : David Lees (The Natural History Museum, London), Ohshima Issei (Université de Kyoto) et Carlos Lopez-Vaamonde (URZF, Orléans).

(2) la seconde collecte, réalisée sur 6 jours consécutifs, s'est quant à elle située à une tout autre période de l'année puisqu'elle a commencé le 4 Septembre 2010 et s'est terminée le 9 Septembre 2010. Elle fut réalisée par cinq chercheurs : David Lees (The Natural History Museum, London), Ohshima Issei (Université de Kyoto) et Carlos Lopez-Vaamonde (URZF, Orléans), Akito Y. Kawahara (University of Florida) et Atsushi Kawakita (Kyoto University).

(3) Enfin, la troisième et dernière collecte a été réalisée l'année suivante du 28 Janvier 2011 au 10 Février 2011. Deux chercheurs ont ici entrepris les collectes : Rodolphe Rougerie (Université de Rouen) et Alex Smith (University of Guelph).

Une fois collectés, une grande part des lépidoptères ont été tués dans des flacons à cyanure ou par injection d'ammoniaque. Une autre partie à quant à elle été placée au congélateur. Par la suite, tous les lépidoptères récoltés et tués ont été placés au sein de grands séchoirs dans l'optique de les déshydrater un maximum et ce, dans les plus brefs délais. Ce processus de déshydratation est très important notamment pour préserver l'ADN au sein des tissus. En effet, cette molécule complexe est très instable et se dégrade rapidement en milieu propice aux réactions chimiques tel que le milieu aqueux. Nous verrons que cette conservation de l'intégrité de certains fragments d'ADN va être essentielle pour la suite de nos expérimentations. Par la suite, tous les individus furent placés dans des papillotes constituées de papier glace pour éviter que ceux-ci ne s'abiment.

Séquençage de l'ADN en laboratoire

L'objectif premier du protocole expérimental visait à obtenir la séquence ADN du gène mitochondrial COI (658pb) pour la totalité des individus collectés. Ce travail a été réalisé en collaboration avec le Centre Canadien du DNA-Barcoding (<http://www.ccdb.ca>) qui, est spécialisé dans le séquençage à haut débit du fragment du gène COI utilisé comme code-barres ADN standard (capacité du CCDB supérieure à 200 000 spécimens par an) pour une grande variété d'organismes vivants.

L'obtention de la séquence du gène COI pour chacun des individus, par le biais de cet institut, s'est réalisée selon une procédure précise (cf. Annexe 1). Elle a surtout consisté à préparer un échantillon de tissu de chaque individu pour qu'ensuite le CCDB en extraie l'ADN et entreprenne le séquençage du gène COI (cf protocole Annexe 2). Dans le cadre de mon travail, les échantillons de tissu de 1765 individus furent préparés et envoyés au CCDB. Le reste des individus à préalablement été traité par Carlos Lopez (URZF), Rodolphe Rougerie (Université de Rouen) et David Lees (The Natural History Museum, London).

Au final, sur les 7349 individus récoltés, 6891 séquences ont été obtenues, dont une grande majorité (98.5 %) étaient de longueur complète (658 paires de bases (bp)).

Identification des individus collectés

L'étape suivante du protocole expérimental visait ensuite à identifier (si possible jusqu'au niveau espèce) chacun des spécimens associés aux séquences du gène COI obtenues. Comme nous l'avons évoqué en introduction, nous avons utilisé la variabilité nucléotidique de ce gène COI pour caractériser et ainsi déterminer le nombre d'espèces présentes dans notre échantillon. Pour ce faire, nous avons utilisé certains outils analytiques de BOLD.

- [Comparaison de séquences](#)

Nous avons tout d'abord tenté d'identifier nos séquences en comparant celles-ci aux séquences déjà publiées sur le site de BOLD. Rien que pour les lépidoptères, 634 870 séquences ont à ce jour été publiées en lignes et rendues publiques. Ce total de 634 870 séquences dans lesquelles ont été identifiées 74 503 espèces de lépidoptères constitue finalement une véritable bibliothèque de référence.

Pour réaliser ces identifications, nous avons utilisé le moteur d'identification de BOLD dans lequel nos séquences inconnues peuvent être entrées par un simple copier-coller. Nous avons tenu compte

du degré de similarité de ces séquences et nous avons considéré comme conspécifiques les individus pour lesquels cette similarité était supérieure à 98%. Ce pourcentage correspond en fait au pourcentage moyen minimum de similarité retrouvé au sein des espèces de lépidoptères. Nous avons aussi complété ces identifications avec l'analyse d'arbres phénétiques ainsi qu'avec la consultation de données photographiques (Annexe 3).

Le moteur d'identification de BOLD nous a permis de caractériser un certain nombre d'espèces parmi nos barcodes génétiques.

Ces identifications n'auraient finalement pas été possibles si les individus présents dans la bibliothèque de référence de BOLD n'avaient pas été soumis à une expertise taxonomique des systématiciens. Par extension, ce sont donc ces systématiciens qui nous ont permis d'identifier nos séquences. Nous le considérerons comme tel dans la suite de cette étude.

Cependant comme l'expertise taxonomique des séquences présentes dans la bibliothèque de référence n'est pas exhaustive et que la bibliothèque en elle-même est loin d'être complète, une grande part de nos séquences n'ont pas pu être identifiées jusqu'au rang d'espèce. Il a alors fallu procéder différemment caractériser les espèces dans le reste de notre jeu de données.

Nous avons notamment sollicité l'expertise de certains taxonomistes basés en France ou à l'étranger et chacun spécialisés dans une famille de lépidoptères distincte. Voici la liste des collaborateurs ainsi que la famille dans laquelle ils sont spécialisés :

David Adamski (Smithsonian, Washington, **Blastobasidae**) ; Jérôme Barbut (Muséum d'Histoire Naturelle de Paris, **Noctuidae**) ; Don Davis & Jay Sohn (Smithsonian, Washington, **Lyonettidae & Opostegidae**) ; Cees Gielis, (**Pterophoridae**) ; Michel Laguerre (**Arctiidae**) ; Bernard Landry (Muséum Genève, **Pyralidae**) ; Antoine Leveque (Muséum d'Histoire Naturelle de Paris, **Geometridae** et **Hedylidae**) ; Joel Minet (Muséum d'Histoire Naturelle de Paris, **Microlepidopteres**), Josef Razowski (**Tortricidae**), Paul Thiaucourt (**Notodontidae**).

Néanmoins, malgré ce soutien de ces experts taxonomistes, bon nombre de nos spécimens n'ont pu être identifiées jusqu'à l'espèce. Nous avons donc eu recours à un dernier moyen pour caractériser les espèces encore non-identifiées.

- [Taxonomie moléculaire : utilisation des "Barcode Index Numbers" \("BIN"\)](#).

Pour tenter de caractériser les espèces pour le reste de nos séquences, nous avons utilisé un autre outil de BOLD permettant de caractériser et nommer automatiquement les groupes ou « clusters » de séquences. Cet outil permet de discriminer des unités appelées BIN (Barcode Index Number) d'une manière comparable à ce que propose la taxonomie moléculaire en caractérisant des Unités Moléculaires Taxonomiques Opérationnelles (MOTU) (Tautz *et al.* 2003; Blaxter, 2004; Blaxter *et al.* 2005). En l'absence d'identification morphologique, ces BINs peuvent être utilisés comme un proxy des espèces au sein de notre jeu de données. L'algorithme de « clusterisation » permettant de générer ces unités n'a pas encore été publié mais il a été présenté et discuté lors de la dernière conférence internationale sur le barcoding à Adelaïde (Ratnasimgham, & Hebert, 2011). Nous avons utilisé cet outil du fait de son implémentation dans BOLD qui rend son usage très pratique et immédiat et après vérification, pour les familles les mieux connues, de la bonne correspondance entre ces BINs et les espèces morphologiques (voir plus bas). Lorsque les espèces sont caractérisées via cet outil, ce n'est pas un nom scientifique binomial mais un numéro de référence, un BIN, qui leur est attribué.

Analyse des résultats

L'identification spécifique et l'attribution d'un numéro BIN à chacun de nos individus nous a finalement permis de déterminer le nombre total de clusters génétiques présents dans notre échantillon. La connaissance de cette richesse spécifique au sein de notre échantillon, combiné à des analyses statistiques adaptées, nous a ensuite permis d'établir une estimation de la richesse spécifique totale des lépidoptères présents sur les lieux de collectes. Ces analyses statistiques ont été réalisées en plusieurs étapes, dont voici les détails.

- [Réalisation de courbes de régression](#)

L'analyse de ces données a dans un premier temps consisté à réaliser des courbes de raréfaction, ou courbes d'accumulation des espèces ici caractérisées. Ces courbes d'accumulation résultent du nombre cumulé d'espèces rencontrées par unité d'effort d'échantillonnage (Chao *et al.* 2005; Chao *et al.* 2009). Dans notre cas, l'effort d'échantillonnage est ici défini par le nombre d'individus que nous avons collecté toute période de l'année confondue.

Les courbes d'accumulation suivent une fonction bien précise et atteignent un moment donné une asymptote. Cette asymptote indique au final le nombre total d'espèces estimé dans la zone de récolte. Il va de soit alors que plus cette courbe d'accumulation est rectiligne et pentue, moins l'asymptote sera atteinte rapidement et donc plus le nombre total d'espèces estimé sera grand.

Pour réaliser nos courbes d'accumulation nous avons utilisé ici la version 7.72 du logiciel *Ecosym* réalisé par Nicholas J. Gotelli (University of Vermont). Nous avons ensuite utilisé certains scripts (Thibaud Deacaens, Université de Rouen) sous R 2.13.1 (Annexe 4) pour réaliser des graphiques représentatifs de ces courbes.

- [Estimation du nombre total d'espèces](#)

Par la suite, nous avons utilisé deux tests statistiques différents qui permettent d'évaluer chacun le nombre total d'espèces présentes sur la zone de collecte. Ces tests statistiques ont, à partir des données brutes, permis de calculer l'asymptote des courbes d'accumulation. Ce calcul de l'asymptote atteinte a été réalisé pour chacun des deux tests selon un algorithme différent.

Test ACE :

Le premier test statistique, nommé ACE (Chao et al, 2005), a tenu compte de l'abondance relative de chacune des espèces et les a scindé en deux catégories : espèces "rares" et espèces "abondantes". Ce statut, rare ou abondant, a été fonction du nombre d'individus rencontré pour chaque espèce, lequel était défini par un paramètre K. Les espèces pour lesquelles $K < 10$, étaient considérées comme rares, tandis que les espèces pour lesquelles $K > 10$, étaient considérées comme abondantes.

Test Chao1 :

Le second test statistique utilisé opérait de la même manière que le test ACE à la différence près que celui considérait comme "rares" les espèces représentées uniquement par un ou deux individus (aussi appelées respectivement *singletons* et *doubletons*).

Pour mettre en pratique ces tests statistiques, nous avons aussi utilisé certains scripts (Thibaud Deacaens, Université de Rouen) réalisés sous R 2.13.1 (Annexe 4).

III) Résultats

A) Identification des individus collectés

Le séquençage en laboratoire de nos échantillons (7610 spécimens) au CCDB nous a permis d'obtenir un total de 7279 barcodes. Ceci représente un pourcentage de réussite du séquençage d'environ 95%.

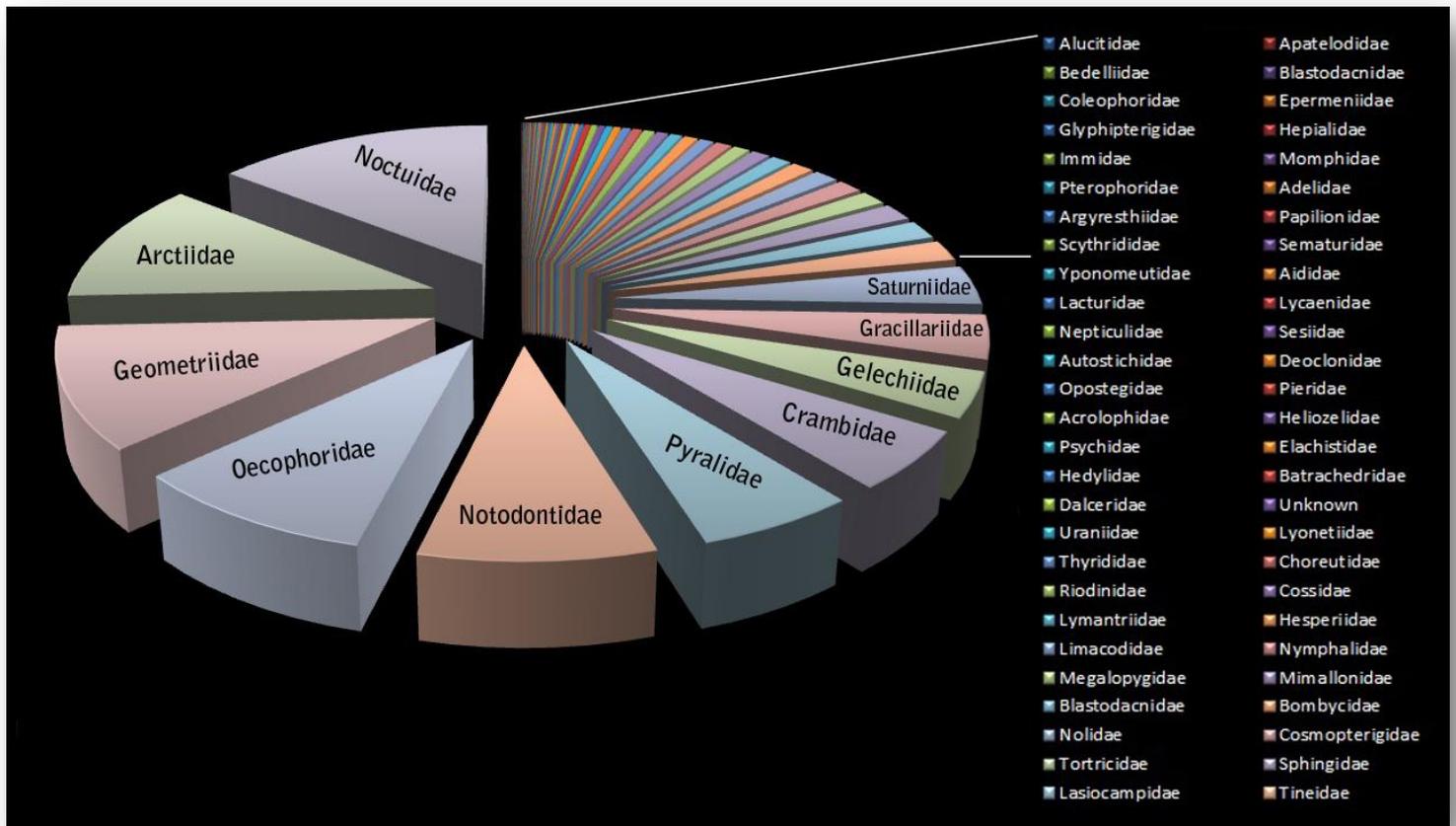
Grâce aux résultats combinés du Barcode Index Number ainsi que de la taxonomie traditionnelle, nous avons pu caractériser parmi tous ces barcodes, un total de 3146 espèces. Ces espèces se répartissaient au sein de 64 familles différentes parmi lesquelles dix furent prépondérantes et totalisaient à elles seules 2457 espèces (fig. 1). La taxonomie de certaines de ces 10 familles était très bien établie, à tel point que la quasi-totalité des séquences que nous avons obtenues étaient déjà

représentées dans la bibliothèque de référence de BOLD. Ce fut notamment le cas pour la famille des Sphingidae et celle des Saturniidae.

Certaines autres familles, comme celles des Arctiidae ou des Lasiocampidae présentaient un cas intermédiaire dans lequel une part importante des espèces (jusqu'à la moitié) n'a pu être caractérisée que par l'utilisation de la taxonomie moléculaire (BIN).

Enfin, chez certaines familles comme celle des Gracillariidae, des Tortricidae ou encore des Tineidae le recours à la taxonomie moléculaire était quasi-exclusif. En effet, ces familles, classées dans la catégorie des micro-lépidoptères n'ont été que très peu étudiées et la taxonomie à leur sujet reste aujourd'hui encore très peu connue avec beaucoup d'espèces nouvelles pour la science.

Figure 1 : Ci-dessous : schéma représentatif de la proportion relative de chaque famille de lépidoptères présente au sein de notre échantillon



B) Etude de la richesse spécifique des lépidoptères

Comme les familles des Saturniidae et celle des Sphingidae étaient ici assez bien représentées dans notre échantillon, mais aussi avant tout parce que ce sont des groupes dont la taxonomie est bien établie, nous avons comparé l'estimation de leur richesse spécifique totale en nous basant d'une part sur l'expertise taxonomique traditionnelle et d'autre part en nous basant sur les nombres BIN définis par l'approche de taxonomie moléculaire. Si les résultats s'avéraient être similaires, ils constitueraient ainsi un argument de poids pour estimer la richesse spécifique au sein de familles dont la majorité des espèces a pour l'instant été caractérisée uniquement via la taxonomie moléculaire. Ceci nous conforterait finalement dans le choix de prendre en compte notre échantillon en entier pour estimer ainsi la richesse spécifique de tous les lépidoptères des Nouragues.

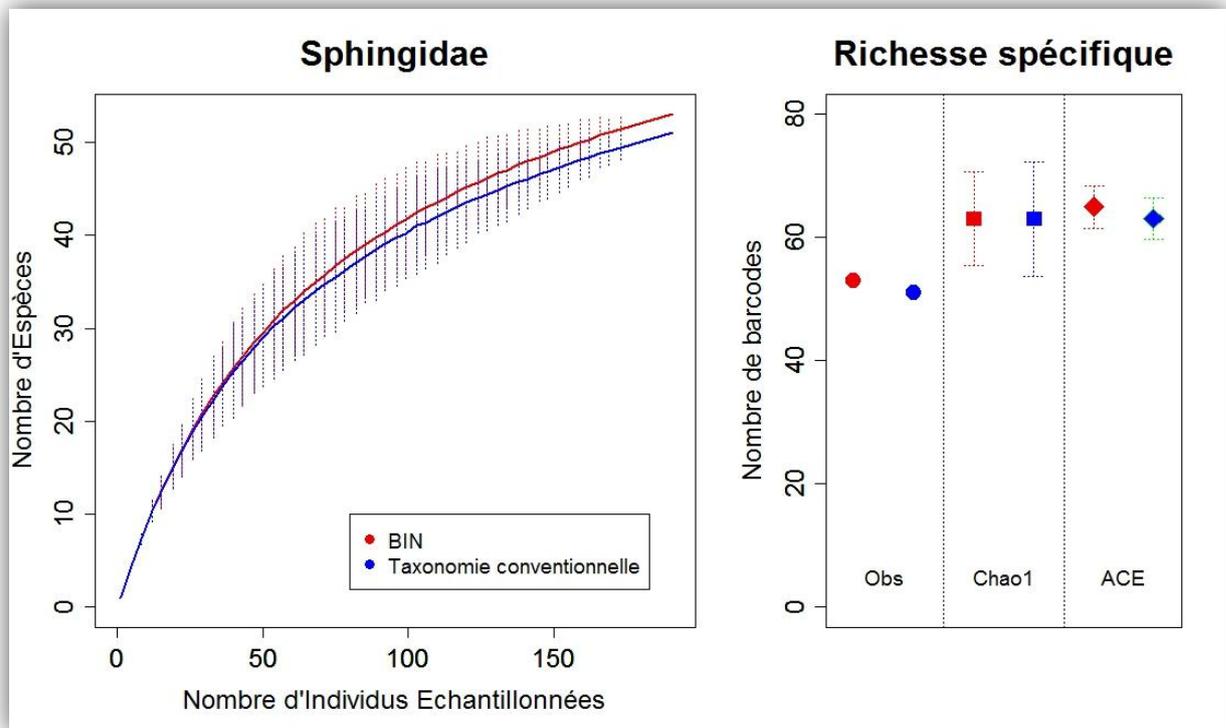
Dans la suite de ce rapport, dans le but de faciliter la lecture, sera attribué le nom "BIN(s)" aux espèces ayant justement été caractérisées via cet outil de la taxonomie moléculaire. Les espèces déterminées sur les bases de la taxonomie traditionnelle ont quant à elles tout simplement été nommées "espèce(s)".

- Estimation de la richesse spécifique totale pour la famille des Sphingidae

Nous pouvons noter, chez la famille des Sphingidae, que les deux courbes de raréfactions propres aux espèces et aux BINs décrivent sensiblement la même trajectoire (fig. 2). Outre le fait que le nombre de BINs est légèrement plus important (53 BINs pour 51 espèces), la pente des deux courbes semble malgré tout quasi-identique et l'on s'attend donc à ce que l'estimation de la richesse spécifique totale donne des résultats similaires.

Nous pouvons en effet lire sur la figure de droite, que via le test Chao1, le nombre total d'espèces et le nombre total de BINs estimés sont tous deux centrés autour de 63, avec un écart type s'étalant respectivement de 55 à 70 et 53 à 73. Le test ACE donne lui aussi des résultats similaires à savoir un nombre total d'espèces et de BINS respectivement centré à 63 et 64, avec un écart type s'étalant respectivement de 62 à 68 et 59 à 65. Cette première similarité des résultats nous encourage donc à nous baser la taxonomie moléculaire pour estimer la richesse spécifique totale chez les familles dont la plupart des espèces n'ont jusqu'alors pas encore été caractérisées via la taxonomie traditionnelle.

Figure 2 : Ci-dessous : Courbes de raréfaction et estimations du nombre total d'espèces pour la famille des Sphingidae. Résultats récapitulatifs montrant d'une part les courbes de raréfaction respectives des espèces et des BINs et d'autre part, l'estimation de leur richesse totale via les tests Chao1 et ACE.

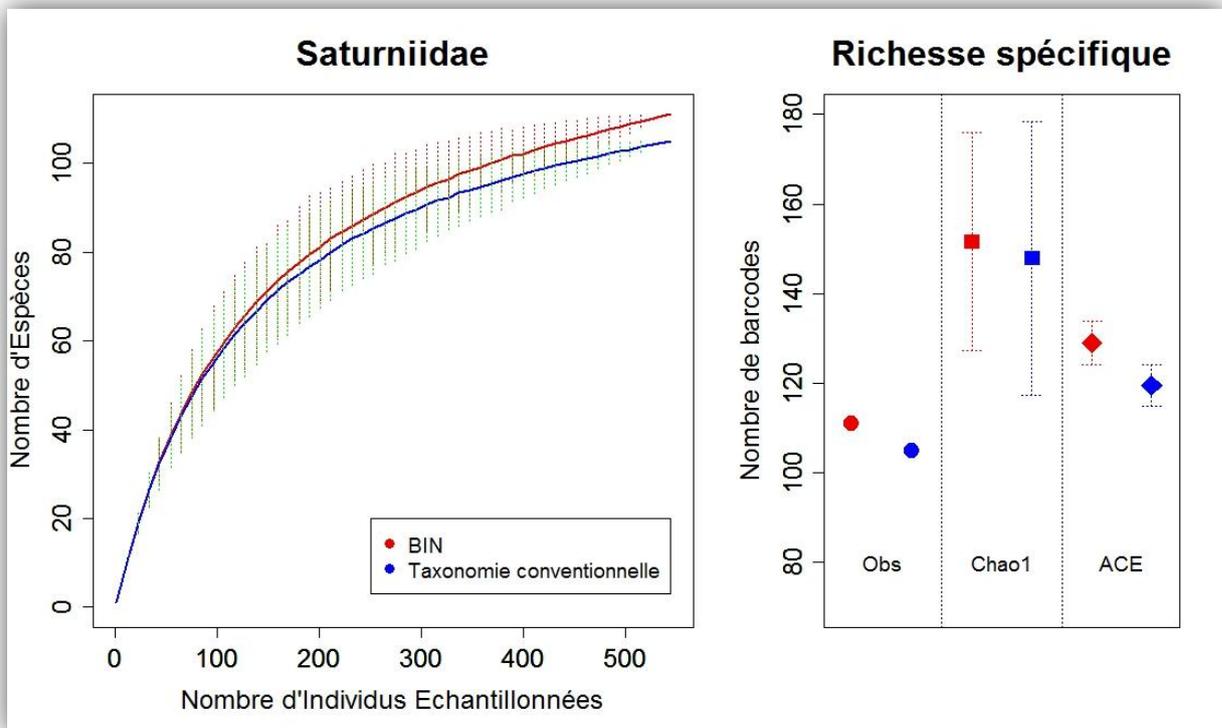


- **Estimation de la richesse spécifique totale pour la famille des Saturniidae**

Les courbes de raréfactions et l'estimation de la richesse spécifique pour la famille des Saturniidae montrent des résultats tout à fait similaires à ceux observés chez la famille des Sphingidae (fig.3). En effet, les courbes de raréfaction propres aux espèces et aux BINs décrivent elles aussi des trajectoires quasi-identiques. Là encore, le nombre de BINs calculé est un peu plus important que le nombre d'espèces (103 contre 110) mais au final, la pente des deux courbes ne diffère que très peu.

Les résultats faisant apparaître l'estimation du nombre total d'espèces et de BINs sont eux aussi très similaires. En effet, pour le test Chao1, bien que la marge d'erreur soit relativement importante (respectivement de 127 à 177 et 117 à 179), nous pouvons noter que l'estimation du nombre total d'espèces ainsi que celle des BINs sont respectivement centrées autour de 151 et 148. Pour le test ACE, cette estimation du nombre total d'espèces et le nombre total de BINs estimé est respectivement de 118 et de 128, avec un écart type s'étalant respectivement de 122 à 134 et 116 à 122. Ici encore, nous pouvons donc noter une similarité satisfaisante des résultats pour les BINs et les espèces.

Figure 3 : Ci-dessous : Courbes de raréfaction et estimations du nombre total d'espèces pour la famille des Saturniidae. Résultats récapitulatifs montrant d'une part les courbes de raréfaction respectives des espèces et des BINs et d'autre part, l'estimation de leur richesse totale via les tests Chao1 et ACE.



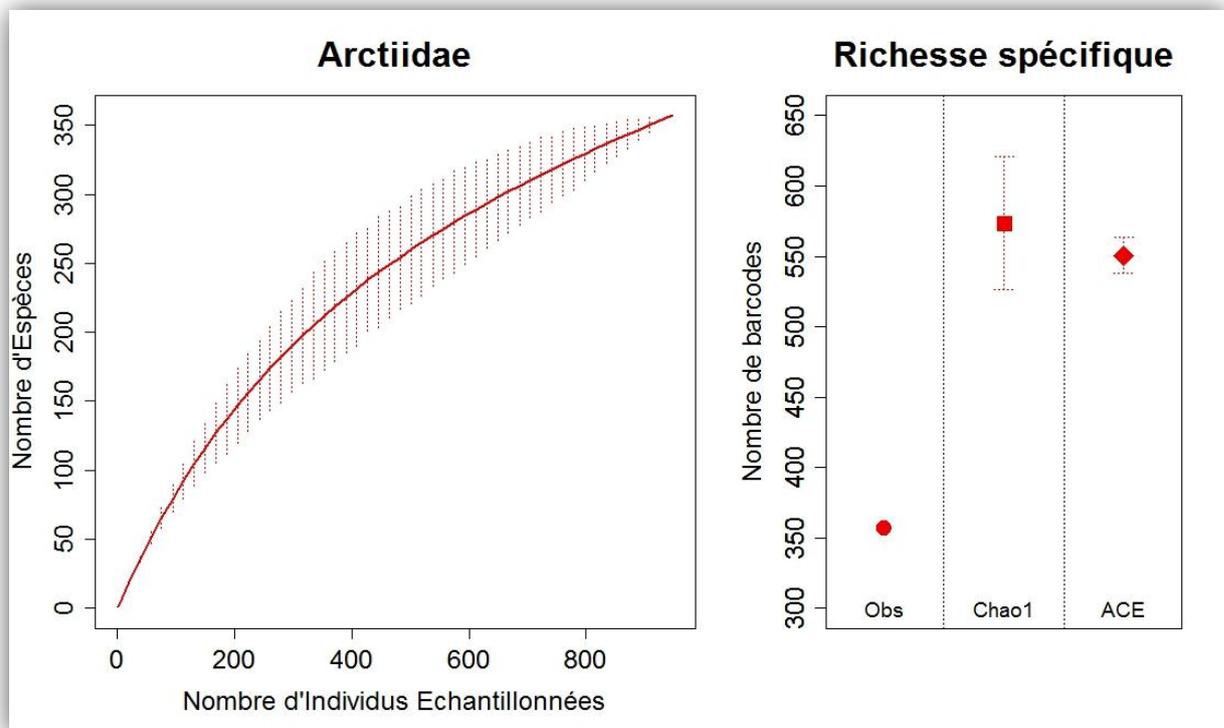
La similitude des résultats obtenus nous a donc conforté dans l'utilisation de la taxonomie moléculaire pour caractériser et finalement estimer le nombre total d'espèces présentes parmi les familles de lépidoptères dont l'expertise taxonomique est partielle ou quasi-inexistante. Nous avons donc arbitrairement choisi deux familles bien représentées dans notre échantillon (à savoir Arctiidae et Gracillariidae) pour estimer leur richesse spécifique et la comparer à celle des groupes déjà bien connus comme les Sphingidae ou les Saturniidae.

- **Estimation de la richesse spécifique totale pour la famille des Arctiidae**

Nous pouvons observer ici que la courbe de raréfaction calculée pour les espèces appartenant à la famille des Arctiidae décrit une trajectoire plus pentue et plus rectiligne que celle des Sphingidae et des Saturniidae (fig.4). En conséquence, le calcul de la diversité spécifique totale réalisée via les tests Chao1 et ACE montre une richesse spécifique bien supérieure à celles observées pour ces deux autres familles. En effet, le nombre total d'espèces estimé ici est respectivement de 575 pour le test Chao1 et de 550 pour le test ACE (avec un écart type s'étalant respectivement de 152 à 208 et 226 à

252) et vient donc ajouter jusqu'à plus de 200 espèces supplémentaires au nombre d'espèces présentes dans notre échantillon initial (350 espèces environ). Pour les familles des Sphingidae et des Saturniidae, cette proportion d'espèces supplémentaires estimée est relativement moindre puisqu'elle est d'environ 5 pour les Sphingidae et est au maximum de 33 pour les Saturniidae (fig. 2-3).

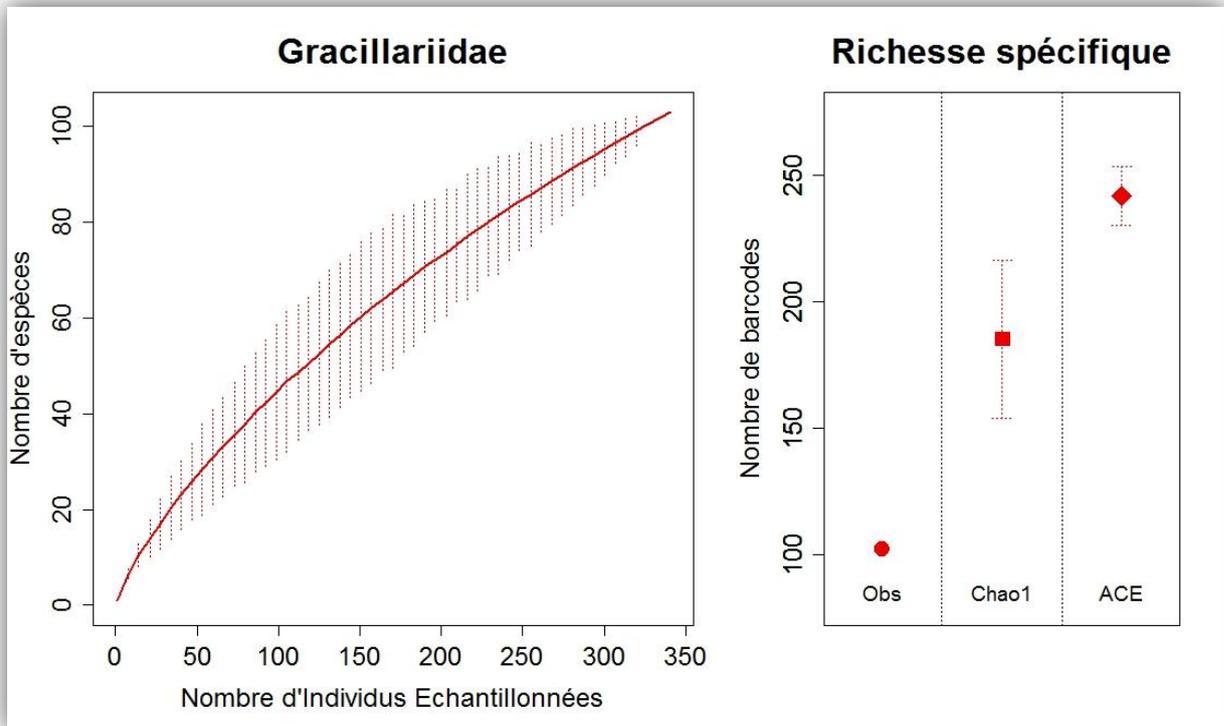
Figure 4 : Ci-dessous : Courbes de raréfaction et estimations du nombre total d'espèces pour la famille des Arctiidae. Résultats récapitulatifs montrant d'une part les courbes de raréfaction respectives des espèces et des BINs et d'autre part, l'estimation de leur richesse totale via les tests Chao1 et ACE.



- [Estimation de la richesse spécifique totale pour la famille des Gracillariidae \(microlépidoptères\).](#)

La courbe de raréfaction réalisée pour les espèces appartenant à la famille des Gracillariidae montre aussi une trajectoire relativement plus pentue et plus rectiligne que celle des Sphingidae et des Saturniidae (fig. 8). Là aussi, le calcul de la diversité spécifique totale pour cette famille montre des résultats bien supérieurs. En effet, le nombre total d'espèces estimé est respectivement centré autour de 177 pour le test Chao1 et de 240 pour le test ACE (avec des écarts type s'étalant respectivement de 152 à 218 et 226 à 253) et vient donc ajouter 74 à 137 espèces supplémentaires à notre échantillon initial (composé de 103 espèces).

Figure 8 : Ci-dessous : Courbes de raréfaction et estimations du nombre total d'espèces pour la famille des Gracillariidae. Résultats récapitulatifs montrant d'une part les courbes de raréfaction respectives des espèces et des BINs et d'autre part, l'estimation de leur richesse totale via les tests Chao1 et ACE.

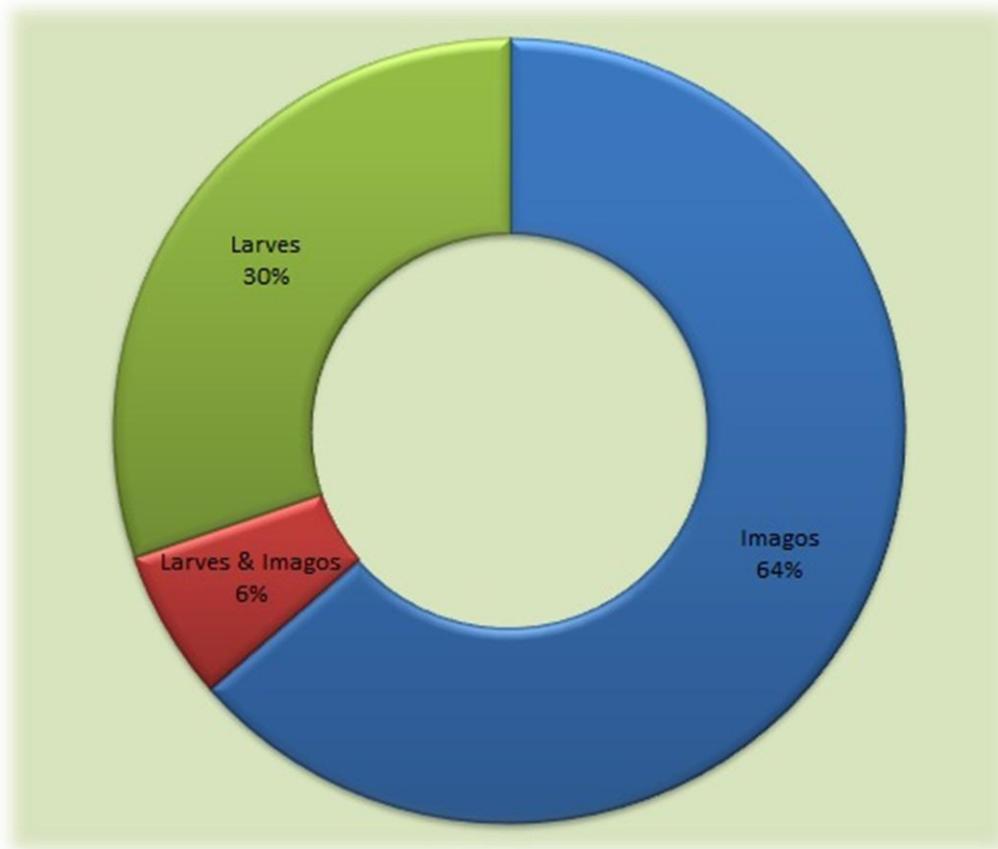


A elles seules, les familles des Gracillariidae et des Arctiidae semblent donc posséder ici une richesse spécifique beaucoup plus importante que celle des Sphingidae et Saturniidae.

- [Proportion relative du nombre d'espèces collectées chez les larves et les imagos de Gracillariidae](#)

Par ailleurs, nous avons jugé intéressant de calculer pour la famille des Gracillariidae le pourcentage relatif du nombre d'espèces présentes parmi les imagos collectés aux pièges lumineux durant la nuit ainsi que le pourcentage relatif du nombre d'espèces présentes parmi les larves collectées le jour via les feuilles minées (fig. 9). Il s'avère alors que 30% (31 BINs sur 103 au total) des BINS de Gracillariidae sont représentés uniquement par les larves et 64% (66 BINs sur 103) sont représentés par des imagos. De plus, seulement un 6% (6 BINs sur 103) des BINS sur la centaine d'espèces de Gracillariidae collectées au total, sont à la fois représentées par les larves et par les imagos.

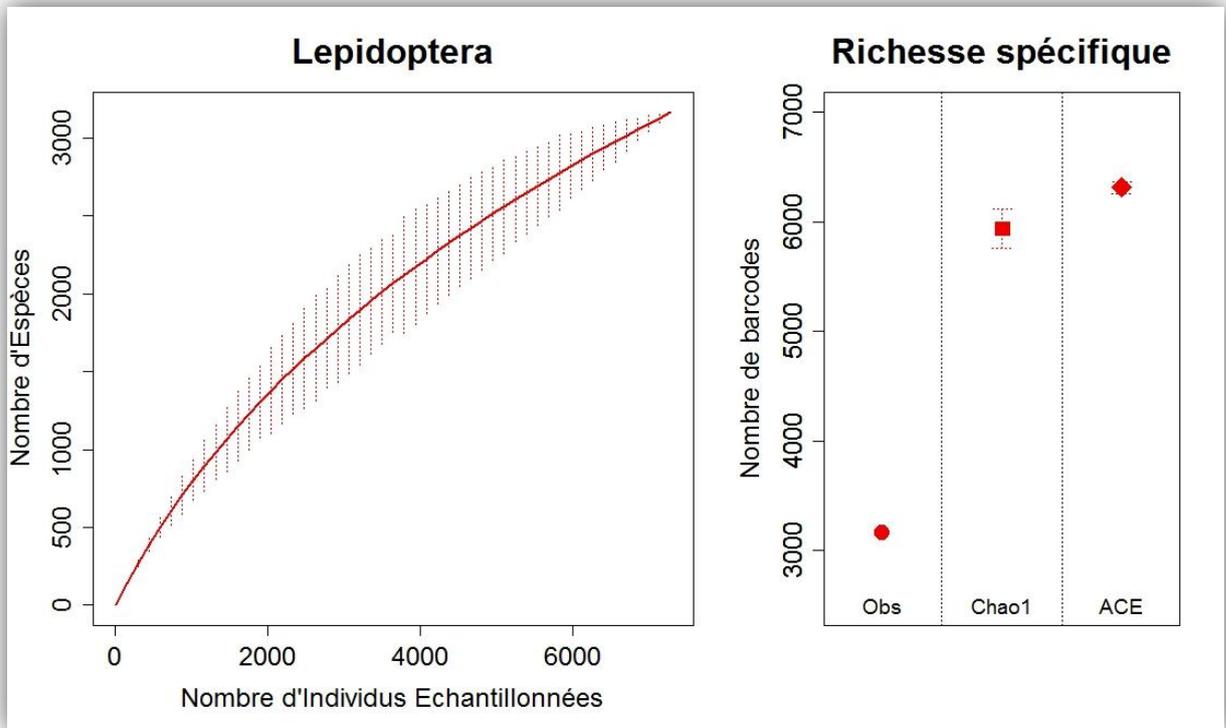
Figure 9 : Schéma récapitulatif nous montrant la proportion du nombre d'espèces récoltées parmi les larves et les imagos de Gracillariidae.



- **Estimation de la richesse spécifique totale des lépidoptères des Nouragues**

Par la suite, nous avons donc réalisé la courbe de raréfaction ainsi que les tests Chao1 et ACE en prenant en compte cette fois-ci les 7279 lépidoptères collectés (fig. 10). Comme nous pouvons le voir sur la figure 10, le nombre total d'espèce estimé via le test Chao1 est centré autour de 5937, avec un écart type s'étalant de 5847 à 6028. Le test ACE, avec un nombre total d'espèce estimé à 6312 et un écart type s'étalant de 6288 à 6338, donne des résultats cohérents à ceux du test Chao1.

Figure 10 : Ci-dessous : Courbes de raréfaction et estimations du nombre total d'espèces pour tous les lépidoptères des Nouragues. Résultats récapitulatifs montrant d'une part les courbes de raréfaction respectives des espèces et des BINs et d'autre part, l'estimation de leur richesse totale via les tests Chao1 et ACE.



IV) Discussion

Les analyses statistiques réalisées pour estimer la richesse spécifique des lépidoptères présents au sein de la réserve naturelle des Nouragues nous ont donné de nombreux résultats que nous avons ici tenté d'interpréter.

Tout d'abord, nous avons pu noter que les familles de lépidoptères partiellement ou très peu connues telles que les Gracillariidae possèdent une diversité considérable avec quasiment autant d'espèces estimées pour les Nouragues (177-240 espèces) que pour toute l'Europe (257 espèces, De Prins & De Prins, 2012).

Ces familles, dont les espèces sont loin d'avoir une morphologie spectaculaire (petite voire très petite taille, couleurs ternes) n'ont finalement que très peu suscité l'intérêt des taxonomistes. Même en Europe, les connaissances relatives aux microlépidoptères restent encore très partielles. D'ailleurs, nous avons pu constater qu'une part importante des espèces non-identifiées chez les Arctiidae appartenaient à la sous-famille des Lithosiinae et contient aussi beaucoup d'espèces de petite taille. Dans l'optique de connaître et protéger la biodiversité et ce, dans les plus brefs délais, l'étude privilégiée de ces microlépidoptères à l'aide d'une approche barcoding serait donc à encourager vivement.

Aussi, nous avons pu noter parmi ces résultats traitant des Gracillariidae, que la collecte active de feuilles minées de larves permettait d'ajouter un nombre d'espèces de loin non négligeable à notre échantillon. Ceci pourrait être interprété par le fait qu'une part importante de ces espèces de Gracillariidae ne serait finalement pas ou peu attirée par les sources lumineuses au stade imago. Cette hypothèse est notamment encouragée par le fait qu'il a déjà été mis en évidence le caractère lucifuge de certaines espèces lépidoptères comme par exemple *Cucullia lucifuga* en Europe (Robineau *et al.* 2007). Nous pourrions renforcer ou contrecarrer cette hypothèse en étalant notamment l'effort d'échantillonnage aux stades larvaires d'autres familles de microlépidoptères. Si cette hypothèse s'avérait renforcée, il serait alors logique de l'étendre à tous les lépidoptères en général et entreprendre donc la collecte active des stades larvaires de toute famille confondue.

Cette collecte active de larves viendrait sans doute ajouter un nombre non négligeable d'espèces de lépidoptère à l'estimation totale qui a été réalisée. Cette estimation, faisant état de pas moins de 5937 espèces présentes sur les lieux de collectes est par ailleurs considérable.

En effet, les pièges lumineux ayant finalement une portée d'attraction assez réduite (quelques centaines de mètres), cette richesse spécifique estimée ne s'appliquait en définitive qu'à quelques

hectares. Ce nombre d'espèces estimé pour quelques hectares est cependant tout aussi important que les quelques 5200 espèces de lépidoptères comptabilisés pour la France métropolitaine (Lafranchis, 2007; Robineau *et al.* 2007) et deux fois plus important que les 2995 espèces décrites pour le Royaume-Unis (<http://ukmoths.org.uk/index.php>). Il nous montre à quel point l'entomofaune de la Guyane est riche. Il aurait été d'ailleurs intéressant de connaître la portée exacte de ces pièges lumineux pour additionner à cette surface de collecte, la surface des zones de prospection le jour et ainsi estimer combien d'espèces de lépidoptère sont présentes à l'hectare aux Nouragues.

Un autre point intéressant de cette étude aurait été de mettre en évidence la variation spatiale de ce nombre d'espèces, notamment en réalisant des estimations de richesse spécifique pour les individus collectés sur le site de l'Inselberg (notamment au sein du biotope savane-roche) et pour les individus collectés sur le site de Parare. Cependant l'effort d'échantillonnage actuel pour ces deux sites n'est pas encore suffisant.

Bien que ce nombre total d'espèces de lépidoptères estimé soit considérable, il se peut aussi qu'il ait été grandement sous-estimé. En effet, bien que les collectes de lépidoptères fussent réalisées à deux périodes principales de l'année (janvier-février et septembre), il est cependant fort probable que certaines espèces de lépidoptères volaient à des périodes tout autres et ont finalement échappé aux piégeages et aux captures. Pour mettre en évidence ce Turn-over temporel des espèces, il aurait notamment été intéressant de comparer l'estimation du nombre total d'espèces sur la base des campagnes de collectes réalisées en Janvier-Février et celle réalisée en Septembre.

Aussi, ce nombre total d'espèces peut notamment être remis en cause par certains biais de la méthode d'estimation utilisée (Colwell & Coddington, 1994; Bunge & Fitzpatrick, 1993).

Ainsi, cette étude nous a permis de caractériser la richesse spécifique des lépidoptères des Nouragues en montrant notamment des résultats semblables entre la taxonomie traditionnelle et la taxonomie moléculaire. Elle a ainsi participé à la connaissance et par extension, à la préservation de la biodiversité présente aux Nouragues. Il est cependant évident que des études supplémentaires doivent être réalisées pour caractériser la biodiversité présente au sein des familles les moins connues mais aussi au sein des familles déjà bien étudiées, notamment via des aspects écologiques. En effet, par exemple nous avons pu dénombrer au sein de la famille des Saturniidae, 17 clusters génétiques appartenant à un même genre (*Hylesia*). Il serait intéressant de savoir si ces 17 clusters partagent la même niche écologique ou si au contraire ils sont dépendants de facteurs environnementaux bien distincts. Aussi complexe soit-elle, la connaissance totale de la biocénose des Nouragues doit être envisagée pour entreprendre des mesures de sauvegardes les plus adaptées.

V) Bibliographie

- 1) Blaxter, M. L. (2004). "The promise of a DNA taxonomy." *Philosophical Transactions of the Royal Society B: Biological Sciences* 359(1444): 669-679.
- 2) Blaxter, M., J. Mann, et al. (2005). "Defining operational taxonomic units using DNA barcode data." *Philosophical Transactions of the Royal Society of London (Series B)* 360: 1935-1943.
- 3) Bunge, J. and Fitzpatrick, M. (1993). Estimating the number of species: a review. *J. Amer. Statist. Ass.*, 88, 364-373.
- 4) Charles-Dominique Pierre, Blanc Patrick, Larpin Denis, Ledru Marie-Pierre, Ricra Bernard, Sarthou Corinne, Servant Michel, Tardy Christophe. (1998) Forest perturbations and biodiversity during the last ten thousand years in French Guiana. *Acta Oecologica* 19 (3) 295-302.
- 5) Colwell, R. K. and Coddington, J. A. (1994). Estimating terrestrial biodiversity through extrapolation. *Philos. Trans. Royal Soc., London, Series B*, 345, 101-118.
- 6) Cook L. G., Edwards R. D., Crisp M. D. and Hardy N. B. (2010). Need morphology always be required for new species descriptions? *Invertebrate Systematics*, 24, 322 – 326.
- 7) Costion C, Ford A, Cross H, Crayn D, Harrington M, et al. (2011) Plant DNA Barcodes Can Accurately Estimate Species Richness in Poorly Known Floras. *PLoS ONE* 6(11): e26841. (doi:10.1371/journal.pone.0026841)
- 8) Devries, P. J., Austin, G. T. and Martin, N. H. (2009), Estimating species diversity in a guild of Neotropical skippers (Lepidoptera: Hesperidae) with artificial lures is a sampling problem. *Insect Conservation and Diversity*, 2: 125–134. doi: 10.1111/j.1752-4598.2009.00047.x
- 9) De Prins J, De Prins W. (2012). Global Taxonomic Database of Gracillariidae (Lepidoptera). World Wide Web electronic publication (<http://www.gracillariidae.net>).
- 10) Gaikwad S. S., Ghatge H. V., Ghaskadbi S. S., Patole M. S., Shouche Y. S. (2010) DNA barcoding of nymphalid butterflies (Nymphalidae: Lepidoptera) from Western Ghats of India. *Mol Biol Rep* DOI 10.1007/s11033-011-0988-7.
- 11) Gasc J.P., Sarthou C., Garrouste R., Villiers J.F., Cremers G., Thiollay, J.M. (1998). Inselbergs et savanes-roches en Guyane: biodiversité et conservation des milieux associés aux affleurements granitiques. *JATBA* 40, 311–327.

- 12) Gonda Valéry, Freycon Vincent, Molinob Jean-François, Brunauxc Olivier, Ingrassiac Florent, Joubertc Pierre, Pekeld Jean-François, Prévoste Marie-Françoise, Thierronf Viviane, Trombec Pierre-Julien, Daniel Sabatier. (2011) Broad-scale spatial pattern of forest landscape types in the Guiana Shield. *International Journal of Applied Earth Observation and Geoinformation* 13, 357–367.
- 13) Hammen V. C., Biesmeijer J. C., Bommarco R., Budrys E., T. Christensen R., Fronzek S., Grabaum R., Jaksic P., Klotz S., Kramarz P., Kröel-Dulay G., Kühn I., Mirtl M., Moora M., Petanidou T., . Pino J, Potts S. G., Rortais A., Schulze C. H., Steffan-Dewenter I., Stout J., Szentgyörgyi H., Vighi M., Vujic A., Westphal C., Wolf T., Zavala G., Zobel M., Settele J., Kunin W. E. (2011). Establishment of a cross-European field site network in the ALARM project for assessing large-scale changes in biodiversity. *Environ Monit Assess* (2010) 164:337–348. DOI 10.1007/s10661-009-0896-7.
- 14) H.R.1268 : National Biological Diversity Conservation and Environmental Research Act (1990).
- 15) Hebert Paul D. N., Cywinska Alina, Ball Shelley L. and DeWaard Jeremy R. (2003a). Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B* 270, 313–321 313 DOI 10.1098/rspb.2002.2218
- 16) Hebert P.D.N, Penton E.H, Burns J.M, Janzen D.H, Hallwachs W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the Neotropical skipper butterfly *Astrartes fulgerator*. *Proc. Natl Acad. Sci.* 101, 14 812–14 817.
- 17) Janzen DH, Hallwachs W (2011) Joining Inventory by Parataxonomists with DNA Barcoding of a Large Complex Tropical Conserved Wildland in Northwestern Costa Rica. *PLoS ONE* 6(8): e18123. doi:10.1371/journal.pone.0018123
- 18) Janzen Daniel H; Hallwachs Winnie; Burns John M; Hajibabaei, Mehrdad; Bertrand, Claudia; Hebert, Paul D N. (2011) Reading the complex skipper butterfly fauna of one tropical place. *PLoS ONE* 6(8): e19874. doi:10.1371/journal.pone.0019874
- 19) KRISTENSEN N.P., SCOBLE M.J. & KARSHOLT O. 2007: Lepidoptera phylogeny and systematics: the state of inventorying moth and butterfly diversity. *Zootaxa* 1698: 699–747.
- 20) Lawton J.H. and May R.M. (2005) *Extinction rates*. Oxford University Press, Oxford, Insight Review Articles.
- 21) LAFRANCHIS T. (2007) Papillons d'Europes. *Diatheo*.
- 22) Moritz, C., and Cicero, C. (2004). DNA barcoding: promise and pitfalls. *PLoS Biology* 2, e354. doi:10.1371/journal.pbio.0020354.

- 23) Nieukerken, EJ, L Kaila, JJ Kitching, et al. (2011). Order Lepidoptera Linnaeus, 1758. In: Zhang, Z.-Q. (Ed.), *Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness*. Zootaxa 3148:212–221.
- 24) Pogue M.G. (2009) Three new cryptic species within the *Dargida procinctus* (Grote) complex (Lepidoptera: Noctuidae: Hadeninae) from the Neotropics. *Proceedings of the Entomological Society of Washington*. 111:686-697.
- 25) Proença, V., Pereira, H.M. & Vicente, L. (2010) Resistance to wildfire and early regeneration in natural broadleaved forest and pine plantation. *Acta Oecologica* 36: 626-633.
- 26) Ratnasingham, S. R., Hebert, P. D. N. (2011) Barcode Index Numbers : An Interim Taxonomic System For Animals. Fourth International Barcode of Life Conference.
- 27) Raupach Michael J, Astrin Jonas J, Hannig Karsten, Peters Marcell K, Stoeckle Mark Y, Wägele Johann-Wolfgang. (2010) Molecular species identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear rDNA expansion segments and DNA barcodes. *Frontiers in Zoology* 2010, 7:26
- 28) Robineau R. & Al. (2007) *Guide des Papillons Nocturnes de France*. Delachaux & Niestlé.
- 29) Sarthou, C., Villiers, J.F., Ponge, J.F., 2003. Shrub vegetation on tropical granitic inselbergs in French Guiana. *J. Veg. Sci.* 14, 645–652.
- 30) Sarthou Corinne, Kounda-Kiki Charlotte, Vaçulik Anne, Mora Philippe, Ponge Jean-François. (2009) Successional patterns on tropical inselbergs: A case study on the Nouragues inselberg (French Guiana). *Flora* 204, 396–407.
- 31) Simona Marcelo F, Gretherc Rosaura, de Queirozd Luciano P., Skemae Cynthia, R. Penningtone Toby, and Hughes Colin E. (2009) Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by in situ evolution of adaptations to fire. doi: 10.1073/pnas.0903410106.
- 32) Stuart Chapin III F, Zavaleta Erika S, Eviner Valerie T, Naylor Rosamond L, Vitousek Peter M, Reynolds Heather L, Hooper David U, Lavorel Sandra, Sala Osvaldo E, Hobbie Sarah E, Mack Michelle C & Díaz Sandra. *Consequences of Changing Biodiversity*. *Nature* 405, 234-242. 2000.
- 33) Tänzler R, Sagata K, Surbakti S, Balke M, Riedel A (2012) DNA Barcoding for Community Ecology - How to Tackle a Hyperdiverse, Mostly Undescribed Melanesian Fauna. *PLoS ONE* 7(1): e28832. doi:10.1371/journal.pone.0028832
- 34) Tautz, D., P. Arctander, et al. (2003). "A plea for DNA taxonomy." *Trends in Ecology and Evolution* 18(2): 70-74.

- 35) Ter Steege H., Sabatier D., Castellanos H., van Andel T., Duivenvoorden J., Oliveira A.A. de, Ek R., Lilwah R., Maas P. and Mori S.A. 2000 - An analysis of the floristic composition and diversity of Amazonian forests including those of the Guiana Shield. *Journal of Tropical Ecology* 16: 801-828.
- 36) Vaglia T, Haxaire J, Kitching I J, Meusnier I & Rougerie R. (2008). Morphology and DNA barcoding reveal three cryptic species within the *Xylophanes neoptolemus* and *loelia* species-groups (Lepidoptera: Sphingidae). *Zootaxa* 1923: 18–36.
- 37) Wilson JJ (2011) Assessing the Value of DNA Barcodes for Molecular Phylogenetics: Effect of Increased Taxon Sampling in Lepidoptera. *PLoS ONE* 6(9):e24769. doi:10.1371/journal.pone.0024769

VI) Annexes

Annexe 1 : Texte récapitulatif du protocole établi pur réaliser le séquençage du gène COI en laboratoire.

Pour réaliser le séquençage du gène COI, il a tout d'abord fallu extraire l'ADN des tissus préalablement collectées en utilisant le kit GenElute Mammalian Genomic DNAMiniprep (Sigma Genosys). La séquence de ce gène a ensuite été amplifiée par la PCR et ce, en utilisant une paire d'amorces spécifiques aux lépidoptères : LEP-F1, (5-ATTCAACCAATCATAAAGATAT-3) et LEP-R1, (5TAAACTTCTGGATGTCCAAAAA-3). Au final, le mélange était contenu dans un volume de 50 µl et contenait 2.5 mM de MgCl₂, 5 pmol de chaque amorce, 20 µM de MdNTPs, 10 mM de Tris-HCl (pH 8.3), 50 mM de KCl, 10 à 50 ng d'ADN extraite ainsi qu'une unité d'ADN Taq polymérase. Le thermocycleur était ici réglé à un cycle d'une minute à 94°C, six cycles d'une minute à 94°C, une minute trente secondes à 45°C et une minute 15 secondes à 72°C, le tout suivit ensuite par 36 cycles d'une minute à 72°C avec une étape finale de 5 minutes à 72°C. Les produits issus de cette PCR furent soumis à l'électrophorèse dans un gel d'agarose à 1.0%, colorés au bromide d'ethidium et observées finalement sous éclairage UV. Deux microlitres (20-50 ng) du produit de la PCR furent séquencés en boucle en utilisant les amorces précédemment citées, le kit d'ABI Prism TaqFS dye terminator (Applied Biosystems) ainsi que le logiciel BIG DYE (version 3.1). Le produit de ce séquençage avait un volume total de 10 µl et incluait 10 pmol de chaque primer. Le protocole d'amplification par PCR consistait en un cycle d'une minute à 96°C, suivit ensuite par 30 cycles de 10 secondes à 96°C, de 5 secondes à 55°C et de 4 minutes à 60°C.

Source : Hebert et al, 2003.

Annexe 2 : Détail des procédures exigées lors du séquençage du gène COI via le Centre Canadien sur le DNA-Barcoding (CCDB)

Le séquençage du gène COI pour chacun de nos individus consistait tout d'abord à prélever un morceau de tarse ou un morceau d'antenne (riche en ADN mitochondrial) pour ensuite disposer celui-ci dans un tube Eppendorf. Pour ce faire, le CCDB nous a fait parvenir au préalable un certain nombre de plaques numérotées et constituées chacune de 96 tubes Eppendorf accolés et disposés en 8 lignes de 12 tubes chacune (schéma et photo ci-contre). A chaque ligne est attribuée une lettre de l'alphabet (A à H), tandis que les puits au sein de chaque ligne sont numérotés de 1 à 12. Chaque tube Eppendorf était préalablement rempli de 30 µl d'alcool à 95° de sorte que les morceaux de tissu disposés au fond de ces tubes ne soient pas éjectés au moment de l'ouverture des plaques, notamment à cause de phénomènes d'électricité statique. Nous avons pris soin aussi de nettoyer les pinces entomologiques à chaque prélèvement de tissus dans le but d'éviter toute contamination d'ADN d'un tube à l'autre.

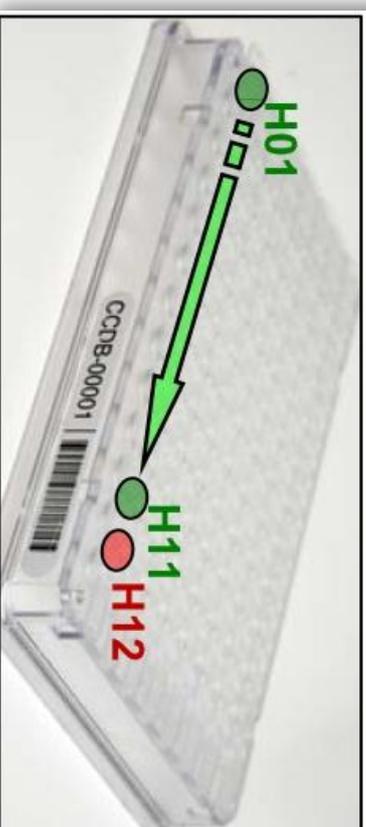
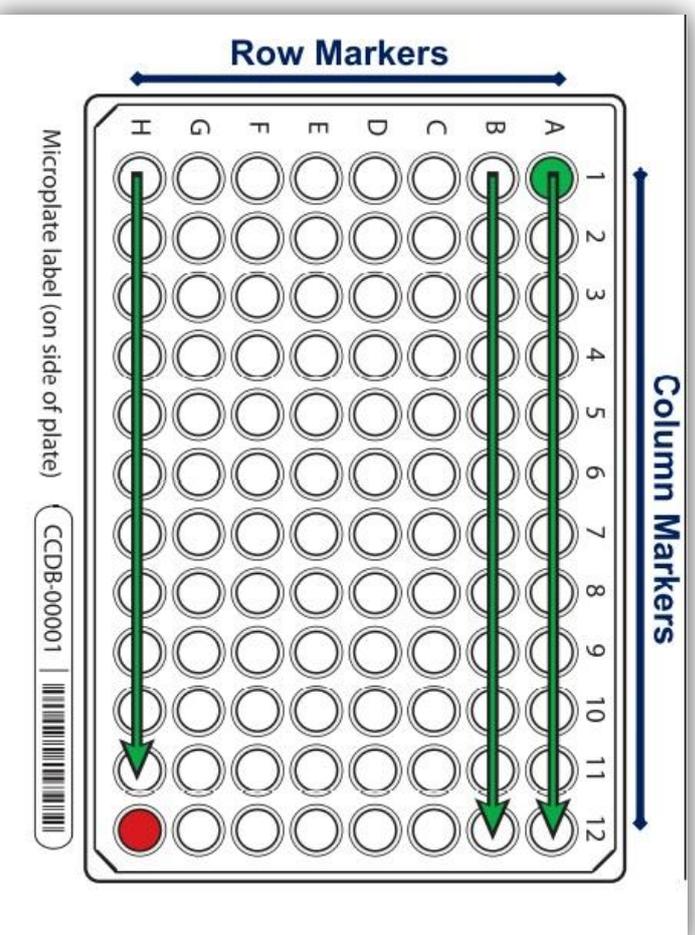
Dans le même temps, nous avons dû préparer un fichier excel propre à chaque plaque dans lequel chacun de ces tubes est affecté d'un numéro de référence unique (ex. CLV5130) qui définit sa position dans la plaque (ex. D08). Ce numéro de référence, aussi appelé Sample ID, est finalement attribué à la séquence obtenue pour chaque tube ainsi qu'au spécimen associé.

Sachant que le dernier tube (H12) sert de tube témoins, 95 individus pouvaient ainsi être traités par plaque.

En parallèle à cette préparation de tissus, nous avons dû préparer un second fichier excel qui compilait toutes les informations relatives à chacun des spécimens associés. Ces informations étaient connectées au spécimen via le Sample ID et se divisaient en quatre grandes catégories, lesquelles sont : (1) Les informations sur le spécimen, à savoir le Sample Id ou encore la personne ayant échantillonné le tissu, (2) La taxonomie de l'individu en question (ordre, famille, genre, espèce) dans la mesure du possible, (3) Les détails propres à l'individu comme le stade de développement, le sexe etc, (4) Les données de récolte, à savoir le collecteur, la date, le pays, la province, les coordonnées GPS etc.

Aussi, chaque spécimen échantillonné devait être accompagné d'une ou plusieurs photographies. Les photographies facilitant l'identification de ces spécimens (spécimen étalé, genitalia) étaient bien sûr préférées. Là encore, un fichier excel permettant la prise en charge de ces photographies a été réalisé.

Les séquences obtenues étaient accessibles et analysables en ligne dans BOLD (Barcode of Life DataSystems - www.boldsystems.org) où elles étaient associées à toutes les données précédemment compilées.



Annexe 3 : Procédure détaillée de l'identification d'espèces via la comparaison de séquences sur BOLD

Le moteur d'identification de BOLD nous renvoyait un résultat composé de différents éléments :

Il présentait notamment un graphique (a) comportant en abscisses les 100 premiers individus référencés

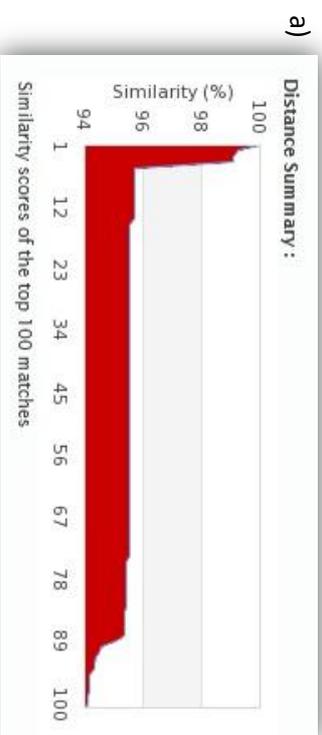
pour lesquels le pourcentage de similarité avec la séquence soumise était le plus important. En ordonnées était présentée la valeur de ce pourcentage de similarité. Ce graphique nous a permis de déterminer si la séquence soumise est déjà plus ou moins représentée parmi la bibliothèque de référence.

Un second élément (b) nous indiquait les éventuels noms d'espèces associés à chacune de ces séquences de référence ainsi que leur pourcentage de similarité avec la séquence soumise. C'est cet élément qui nous permettait principalement d'identifier nos séquences. Nous réalisons ici les identifications en prenant en compte le fait que, chez les lépidoptères, ce pourcentage de similarité au sein d'une même espèce est rarement inférieur à 98% (Hebert et al. 2003).

Si nécessaire, nous pouvions aussi générer des résultats complémentaires à cette comparaison de séquences comme par exemple la possibilité de réaliser un arbre phénétique basé lui aussi sur le pourcentage de similarité entre la séquence soumise et les séquences de la bibliothèque de référence (c).

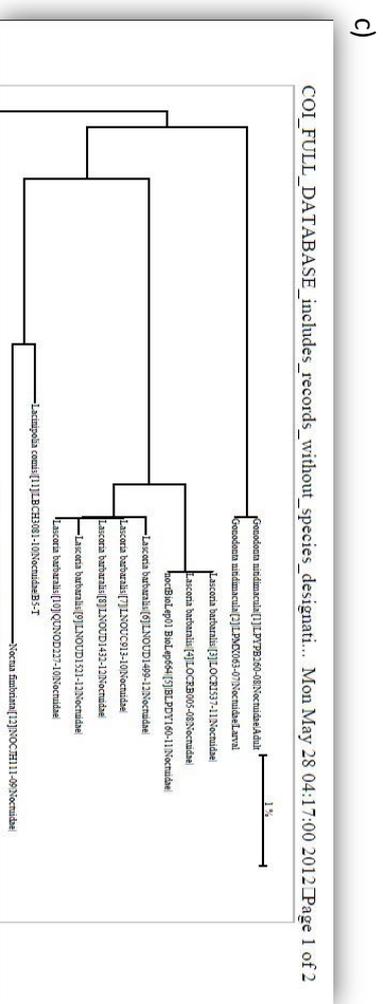
Cet arbre phénétique regroupait les individus appartenant à la même espèce au sein de clusters distincts. L'utilisation de cet arbre phénétique s'avérait bénéfique pour associer une espèce dont le pourcentage de similarité était finalement inférieur à 98%.

Outre à cela, nous avons aussi la possibilité de consulter les photographies des individus associées à chacune des séquences (fig. 3d) et vérifier ainsi s'il y avait une concordance morphologique des individus comparés.



b)

Phylum	Class	Order	Family	Genus	Species	Specimen Similarity (%)
Arthropoda	Insecta	Lepidoptera	Arctidae	Epidesma	<i>ursula</i>	100
Arthropoda	Insecta	Lepidoptera	Arctidae	Epidesma	<i>ursula</i>	99.24
Arthropoda	Insecta	Lepidoptera	Arctidae	Epidesma	<i>ursula</i>	99.08
Arthropoda	Insecta	Lepidoptera	Arctidae	Epidesma	<i>ursula</i>	99.08
Arthropoda	Insecta	Lepidoptera	Arctidae	Epidesma	<i>ursula</i>	95.72
Arthropoda	Insecta	Lepidoptera	Arctidae	Epidesma	<i>ursula</i>	95.72



d)

Number of records: 101

File by resolution: By default, a single image is listed for each specimen. File options override the default behavior.

Genosiana andersoni [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31] [32] [33] [34] [35] [36] [37] [38] [39] [40] [41] [42] [43] [44] [45] [46] [47] [48] [49] [50] [51] [52] [53] [54] [55] [56] [57] [58] [59] [60] [61] [62] [63] [64] [65] [66] [67] [68] [69] [70] [71] [72] [73] [74] [75] [76] [77] [78] [79] [80] [81] [82] [83] [84] [85] [86] [87] [88] [89] [90] [91] [92] [93] [94] [95] [96] [97] [98] [99] [100]

Genosiana andersoni [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31] [32] [33] [34] [35] [36] [37] [38] [39] [40] [41] [42] [43] [44] [45] [46] [47] [48] [49] [50] [51] [52] [53] [54] [55] [56] [57] [58] [59] [60] [61] [62] [63] [64] [65] [66] [67] [68] [69] [70] [71] [72] [73] [74] [75] [76] [77] [78] [79] [80] [81] [82] [83] [84] [85] [86] [87] [88] [89] [90] [91] [92] [93] [94] [95] [96] [97] [98] [99] [100]

Lixocera andersoni [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31] [32] [33] [34] [35] [36] [37] [38] [39] [40] [41] [42] [43] [44] [45] [46] [47] [48] [49] [50] [51] [52] [53] [54] [55] [56] [57] [58] [59] [60] [61] [62] [63] [64] [65] [66] [67] [68] [69] [70] [71] [72] [73] [74] [75] [76] [77] [78] [79] [80] [81] [82] [83] [84] [85] [86] [87] [88] [89] [90] [91] [92] [93] [94] [95] [96] [97] [98] [99] [100]

Xanthopoda [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31] [32] [33] [34] [35] [36] [37] [38] [39] [40] [41] [42] [43] [44] [45] [46] [47] [48] [49] [50] [51] [52] [53] [54] [55] [56] [57] [58] [59] [60] [61] [62] [63] [64] [65] [66] [67] [68] [69] [70] [71] [72] [73] [74] [75] [76] [77] [78] [79] [80] [81] [82] [83] [84] [85] [86] [87] [88] [89] [90] [91] [92] [93] [94] [95] [96] [97] [98] [99] [100]

Annexe 4 : Script de base réalisé et utilisé pour établir les courbes de régression ainsi que les tests statistiques Chao 1 et ACE sous R 2.13.1

```
require(vegan)
require(plotrix)

Lepi <- read.table("donnees brutes.txt", h=T)

estbin<-estimateR(Lepi[,2])

x<-cbind(estbin)

Spin <- read.table("resultats ecosym.txt", h=T)

window$width = 25, height = 15)

layout(matrix(1:2, 1, 2), widths=c(3/5,2/5))

plot(Sbin[,1],Sbin[,2],col="red2", type="l",xlab="Nombre d'individus Echantillonnées",ylab="Nombre d'Espèces", main="Lepidoptera",
      cex.lab=1.5,cex.main=2,cex.axis=1.5,lwd=2,xlim=c(0,max(Sbin[,1])),ylim=c(0,max(Sbin[,2])))

plot(Sbin[,1],Sbin[,4],scol="red2",.pch=NA, col="red2",.pch=NA, col="red2",.sly=3, sfrac=0, add=TRUE)

plot(1:5,x[1,],pch=21, col=c("red2","blue"),bg=c("red2","blue"),cex=2,cex.lab=1.5,cex.main=2,cex.axis=1.5,ylab="Nombre de barcodes",xlab=NA,xaxt='n',xlim=c(0,75,6.25),ylim=c(2500,7000),main="Richesse spécifique")

par(new=TRUE)

plot(3:5,x[2,],x[3,],pch=22, col=c("red2","blue"),.pt.bg=c("red2","blue"),cex=2,cex.axis=1.5,sly=3, sfrac=0.02,ylab=NA,xlab=NA,xaxt='n',xlim=c(0,75,6.25),ylim=c(2500,7000),main=NA)

par(new=TRUE)

plot(5:5,x[4,],x[5,],pch=23, col=c("red2","blue"),.pt.bg=c("red2","blue"),cex=2,cex.axis=1.5,sly=3, sfrac=0.02,ylab=NA,xlab=NA,xaxt='n',xlim=c(0,75,6.25),ylim=c(2500,7000),main=NA)

abline(v=4.5,lty=3);abline(v=2.5,lty=3)

text(1.5,2500,"Obs",cex=1.2);text(3.5,2500,"Chao1",cex=1.2);text(5.5,2500,"ACE",cex=1.2)

write.table(estbin)
```

Source : Thibaud Decaens (Université de Rouen)

Résumé

Le DNA Barcoding est un outil de la biologie moléculaire destiné à l'identification des espèces. Il a été proposé comme très prometteur pour caractériser la biodiversité présente au sein de faunes encore très peu connues telles que les faunes tropicales. Les bibliothèques de barcodes ADN peuvent aider à caractériser les espèces pour établir ainsi la richesse spécifique des milieux et mettre en évidence un turnover spatial et phénologique de cette biodiversité.

Cette étude s'inscrit dans un projet dont l'objectif principal fut de barcoder toutes les espèces de lépidoptères de la station de recherche des Nouragues en Guyane Française, avec un intérêt tout particulier pour les microlépidoptères. La Réserve Naturelle des Nouragues fut visitée au cours de trois expéditions réalisées à deux saisons différentes (Janvier/Février et Septembre, 2010-2011). Faisant face à une faune très diversifiée et largement inexplorée, nous avons identifié les spécimens collectés via les données de référence présentes en ligne (BOLD Systems). Nous avons aussi bénéficié de l'aide de plusieurs experts taxonomistes.

Pour l'ensemble de ces expéditions, 7610 lépidoptères ont pu être collectés, 7279 ont pu être barcodés et 3020 ont pu être identifiés jusqu'à l'espèce.

Nous également avons utilisé un outil de la taxonomie moléculaire (Barcode Index Number, BIN) pour caractériser les espèces n'ayant pas pu être identifiées et réaliser ensuite nos analyses. Nous nous sommes basés sur deux familles (Sphingidae et Saturniidae) très bien connues d'un point de vue taxonomique pour montrer que les clusters génétiques caractérisés par le BIN concordent très bien avec les espèces déterminées via la taxonomie traditionnelle. Nous avons alors utilisé le BIN pour déterminer le nombre de clusters présents au sein des familles les moins connues et nous avons pu ainsi mettre en évidence une diversité remarquable notamment chez les Arctiidae (550 à 575 BINs) et les Gracillariidae (117 à 240 Bins).

Nos données sont particulièrement riches en microlépidoptères, notamment la famille des Oecophoridae avec 9.25% du nombre total de BINs, les Gelechiidae (4.16%) et les Gracillariidae (3.27%) pour lesquelles la grande majorité des espèces n'a pas encore été décrite.

Au total, parmi les 7279 barcodes, 3146 BINs ont pu être caractérisés (30.75% singletons). L'estimation à 5937 (Chao1, 1.95% cIs) du nombre total d'espèces présente sur le site présentes sur le site d'étude révèle une richesse considérable représentant près du double de la faune de Lépidoptères du Royaume Uni. Cette étude représente une des premières démarches dans l'estimation de la biodiversité totale des lépidoptères en région tropicale.

Abstract

DNA barcoding has great potential for carrying out rapid biodiversity assessments of relatively unknown tropical faunas. DNA barcode libraries can be used to examine species richness/identities, and thus spatial and phenological turnover between surveys. The Nouragues nature reserve (French Guiana) has been the focus of three expeditions in two different seasons (January/February and September, 2010-2011). The aim of this project was to barcode all Lepidoptera species with a particular focus on microlepidoptera. Facing an extremely diverse, largely unexplored biota, we identified the specimens by querying the DNA barcode reference library. We were also helped by several expert taxonomists.

Overall, 7610 individuals were collected, of which 7279 were successfully barcoded, and xx were identified to species level. We employed a barcode clustering system (Barcode Index number, BIN) to provide uniquely numbered units for faunal analysis. We used two families, Sphingidae and Saturniidae that are well known taxonomically to show that BIN clusters are a good match to the species delimitations based on traditional taxonomy. We then used BINs as a proxy to species for measuring the number of clusters in families with many unidentified species. Our findings reveal an outstanding level of diversity with 550 to 575 BINs for Arctiidae and 117 to 240 for Gracillariidae.

Our data set is particularly rich in Microlepidoptera, especially Oecophoridae with 9.25 % of all BINs, Gelechiidae (4.16 %) and Gracillariidae (3.27 %) for which most of the species are likely to be undescribed.

In total, the 7279 sequences obtained represent 3146 BINs (30.75% singletons). Our estimates of total species richness on the study site is as high as 5937 species (Chao 1, 95% cIs), thus representing nearly twice the number of species of Lepidoptera in the United Kingdom alone. This study represents one of the first attempts to quantify the complete diversity of Lepidoptera of a Neotropical rainforest.