Université du Québec Institut National de la Recherche Scientifique Centre Institut Armand-Frappier

# Caractérisation de la communauté bactérienne associée à l'agrile du frêne, *Agrilus planipennis* Fairmaire (Coleoptera : Buprestidae)

Par

Amélie Bergeron

Mémoire présentée pour l'obtention du grade de Maître ès sciences (M.Sc.) en microbiologie appliquée

#### Jury d'évaluation

| Président du jury et examinateur interne | Étienne Yergeau<br>INRS- Institut Armand-Frappier   |
|--|---|
| Examinateur externe                      | Armand Séguin<br>Ressources Naturelles Canada       |
| Directeur de recherche                   | Claude Guertin<br>INRS- Institut Armand-Frappier    |
| Codirecteur de recherche                 | Philippe Constant<br>INRS- Institut Armand-Frappier |

Robert Lavallée Ressources Naturelles Canada

© Droits réservés de (Amélie Bergeron), 2016

Codirecteur de recherche

### REMERCIEMENTS

J'aimerais d'abord grandement remercier mon directeur de recherche Pr Claude Guertin. Merci de m'avoir fait confiance et d'avoir cru en moi. Grâce à tes nombreux conseils, ton support et ta grande écoute, mon apprentissage s'est fait bien au-delà de la science. Tu m'as poussé un peu plus chaque jour parce que tu croyais en mes capacités, bien plus que j'y croyais moi-même. Merci d'avoir toujours gardé ta porte grande ouverte et d'avoir trouvé les mots pour me rassurer et me conseiller lorsque j'en ressentais le besoin. Nos nombreuses sorties sur le terrain toujours plus amusantes les unes que les autres resteront gravées dans ma mémoire. Merci pour ces belles années !

Merci à mes codirecteurs Pr Philippe Constant et Dr Robert Lavallée. Philippe, merci d'avoir cru en moi et de m'avoir poussé à me surpasser chaque jour. Merci pour ton support, ta bonne humeur contagieuse et tes précieux conseils. Robert, merci pour ton support moral et technique. Merci pour ces connaissances transmises lors de nos sorties sur le terrain.

Merci à mes grandes amies Audrey-Anne et Isabelle. Audrey-Anne, rapidement tu es passée de superviseur de stage à un modèle pour moi. Une grande amitié s'est rapidement développée entre nous. Merci pour tous ces moments inoubliables à chanter sous la hotte, à faire des folleries et à fondre en larmes. Isabelle, merci pour toutes ces belles soirées passées à découvrir Montréal ou passées dans mon salon à jaser. Merci pour ta grande écoute et ta bonne humeur contagieuse. Merci à vous deux de m'avoir soutenu, conseillé et aidé dans toutes les petites et grandes décisions que j'ai dû prendre. Merci d'être présentes dans ma vie !

Merci à mes collègues de travail, qui sont également devenus des amis, Jean-Philippe Buffet, Laurence Guertin, Vincent Peck, Valentin Popa et Narin Srei. Merci pour votre aide, votre support et vos encouragements. Merci pour nos moments de folleries. Grâce à vous, mon séjour à l'Institut Armand-Frappier aura été des plus agréables.

Je ne peux passer sous silence l'énorme support de mon amoureux et de ma famille. Merci de toujours m'encourager à poursuivre mes rêves. Sans vos encouragements et votre support, cette réussite n'aurait pas été possible.

Finalement, merci à tous ceux et celles qui ont participé, de près ou de loin à l'élaboration et à la concrétisation de ce projet.

Ш

### RÉSUMÉ

L'agrile du frêne, Agrilus planipennis Fairmaire, est un insecte ravageur causant d'importants dommages à toutes les essences de frênes présentes en Amérique du Nord. Ce ravageur exotique, originaire d'Asie, serait arrivé sur le continent américain dans les années 90. Ce n'est qu'en 2002 qu'on le détecte pour la première fois à la frontière des municipalités de Windsor (ON, Canada) et Détroit (MI, États-Unis). Les insectes sont connus pour être vecteurs d'une communauté microbiologique qui leur permet de survivre et de s'adapter à leur environnement. Cette association entre un insecte et ses microorganismes symbiotiques forme un holobionte, où la structure de la communauté microbienne est influencée par divers processus stochastiques et déterministes. L'objectif de ce projet de recherche était de caractériser la communauté bactérienne associée à l'agrile du frêne. Afin de prendre en compte les processus écologiques qui peuvent influencer cette communauté, des populations d'agriles du frêne en provenance de cinq différentes régions ont été récoltées et les individus ont été séparés selon leur sexe. L'ADN génomique total des échantillons a été extrait et le gène codant pour l'ARNr 16S a été séquencé par la technologie Illumina Miseg paired-end. Après avoir traité les séguences avec le protocole UPARSE, des analyses de diversité ont été réalisées avec le logiciel R. Bien qu'aucune différence significative n'ait été détectée entre les individus mâles et femelles, la structure du microbiome des agriles du frêne permet d'établir des liens entre les populations de l'insectes. Ainsi, une communauté bactérienne de base à toutes les populations a été identifiée, alors que certains genres bactériens semblaient être spécifiques à certains échantillons. Ces divergences, expliquées par divers processus stochastiques et déterministes, ont permis de séparer les échantillons en deux écotypes. Grâce à une analyse de β-diversité, l'importance de certains processus écologiques a pu être mise de l'avant. En effet, la population d'agriles du frêne élevée en conditions contrôlées présente une structure bactérienne plus homogène que les populations en provenance de milieu naturel. De plus, certaines espèces bactériennes ont été identifiées comme étant indicatrices du niveau d'établissement des populations.

**Mots clés** : *Agrilus planipennis*; bactérie; holobionte; microbiome bactérien; écologie microbienne; processus stochastiques et déterministes

### ABSTRACT

Emerald ash borer, Agrilus planipennis Fairmaire, is an insect pest that caused significant damage to all ash species present in North America. This exotic pest, native from Asia, arrived in North America in the 90s. However, it was only detected for the first time in 2002 at the border of Windsor (ON, Canada) and Détroit (MI, United-States). Insects are known to live in association with microorganisms that allow them to survive and adapt to their environment. This association between an insect and its symbiotic microorganisms forms an holobiont, where the microbial community structure is influenced by various stochastic and deterministic pressures. The objective of this study was to characterize the bacterial community associated with the emerald ash borer. To considerer the ecological processes that can influence this community, emerald ash borer were harvested from five different sample sites and insects were separated by gender. Total DNA extraction was done and sequencing on the rRNA 16S gene was performed with the Illumina MiSeq paired-end technology. After sequences processing with UPARSE pipeline, diversity analysis were performed with R software. No significant difference was detected between males and females populations. However, the EAB microbiome structure seems to allow linkage between insect populations. A core microbiome was identified, while some bacterial genera appeared to be more specific to some samples. These differences, explained by various stochastic and determinist processes, allowed to separate the sample into two ecotypes. Analysis of  $\beta$ -diversity shows the importance of ecological processes on the shape of the bacteria community. Laboratory-reared population had a microbial structure more homogenous than the one from natural population. Moreover, some bacterial species were identified as indicative of the establishment population level.

**Key words** : *Agrilus planipennis*; bacteria; holobiont; bacterial microbiome; microbial ecology; stochastic and determinist pressures

## TABLE DES MATIÈRES

| Remerciements   | III              |
|---|------------------|
| Résumé  | V                |
| Abstract  | VII              |
| Liste des figures   | XIII             |
| Liste des tableaux  | XV               |
| SYNTHÈSE  | 1                |
| Chapitre 1 : Agrile du frêne  | 3                |
| 1.1 Cycle de développement  | 4                |
| 1.2 Historique de l'introduction de l'agrile du frêne en Amérique du Nord               | 5                |
| 1.3 Dommages causés par l'agrile  | 6                |
| 1.4 Méthodes de détection   | 8                |
| Chapitre 2 : Microorganismes associés aux insectes                                      | 13               |
| 2.1 Transmission verticale et horizontale   | 13               |
| 2.2 Rôles des microorganismes associés aux insectes                                     | 14               |
| 2.2.1 Nutrition   | 14               |
| 2.2.2 Défense   | 15               |
| 2.2.3 Reproduction et développement   | 17               |
| 2.2.4 Communication   | 17               |
| 2.3 Processus stochastiques et déterministes définissant la structure micro<br>insectes | bienne des<br>18 |
| 2.3.1 Température et précipitations   | 19               |
| 2.3.2 Altitude, latitude et longitude   | 20               |
| 2.3.3 Stade de développement de l'insecte   | 20               |
| 2.3.4 Alimentation  | 21               |
| 2.3.5 Pression exercée par les parasites  | 22               |

| 2.3.6 Niveau de la population des insectes                                  | 22       |
|---|----------|
| Chapitre 3 : Hypothèse et objectif  | 23       |
| ARTICLE   | 25       |
| Chapitre 4 : Bacterial community diversity associated with different popula | tions of |
| Agrilus planipennis Fairmaire (Coleoptera : Buprestidea)                    |          |
| 4.1 Résumé  | 28       |
| 4.2 Abstract  | 28       |
| 4.3 Introduction  | 29       |
| 4.5 Material and Method   | 31       |
| 4.5.1 Site Location   | 31       |
| 4.5.2 Insect collection   | 31       |
| 4.5.3 Sample Preparation and DNA Extraction                                 | 32       |
| 4.5.4 PCR Amplification   | 33       |
| 4.5.5 Illumina MiSeq Paired-End Sequencing                                  | 33       |
| 4.5.6 Sequence Processing Pipeline  | 33       |
| 4.5.7 Diversity analysis  | 34       |
| 4.6 Results   | 35       |
| 4.6.1 Bacterial diversity across Agrilus planipennis populations            | 35       |
| 4.6.2 Taxonomy of EAB bacterial community structures                        |          |
| 4.7 Discussion  | 41       |
| 4.8 Acknowledgements  | 44       |
| Chapitre 5 : Discussion et conclusion                                       | 45       |
| Références bibliographiques   | 49       |
| Annexe I : Matériel supplémentaire (ARTICLE)                                | 59       |
| Annexe II : Procédure UPARSE  | 63       |
| Annexe III : Création d'un fichier csv                                      | 75       |

| Annexe IV : | Script R  |
|-------------|---|
| Annexe V :  | Dendrogramme de similarité représentant les échantillons mâles et |
| femelles    |   |

### LISTE DES FIGURES

**Figure 4.1** : Location of sample sites. SM = Sault-Ste. Marie, Ontario; MC = Middlesex County, Ontario; SBLG = Saint-Basile-le-Grand, Quebec; GAT = Gatineau, Quebec and MTL = Montreal, Quebec. 30

**Figure 4.3** : UPGMA agglomerative clustering of the EAB populations. Populations are separated in two significant clusters represented by a significant node ( $\geq$  95%). Ecotype 1 comprises SSM

**Figure 4.5** : Principal component analysis (PCA) of EAB populations. PC1 and PC2 explained 41.5% and 17.1% of the variance respectively. OTUs out of the circle exerted more weight than the average to explain the position of each EAB population in the reduced space. OTUs 1, 19 and 54 defined the PC1, while OTUs 19, 36 and 40 defined PC2. These OTUs explain the separation of populations in Ecotype 1 (square) and Ecotype 2 (triangle) determined by cluster analysis...39

**Figure 1S** : Rarefaction curves of observed OTUs for each replicate. SSM = Sault-Ste. Marie; MC = Middlesex County; SBLG = Saint-Basile-Ie-Grand; GAT = Gatineau and MTL = Montreal. ....60

### LISTE DES TABLEAUX

# Caractérisation de la communauté bactérienne associée à l'agrile du frêne, *Agrilus planipennis* Fairmaire (Coleoptera : Buprestidae) :

### SYNTHÈSE

## CHAPITRE 1 : AGRILE DU FRÊNE

L'agrile du frêne, *Agrilus planipennis* Fairmaire (Coleoptera : Buprestidae), est un insecte exotique envahissant qui induit d'extrêmes dommages à l'intégralité de toutes les essences de frêne présentes en Amérique du Nord. Il s'agit d'un insecte de l'ordre des coléoptères et de la famille des buprestidés, dont les imagos se caractérisent par une couleur vert métallique (Figure 1.1). En ce qui a trait à leur taille, la longueur des adultes varie de 7,5 à 15 mm et la largeur de 3,1 à 3,4 mm (Lyons *et al.*, 2007). Des caractéristiques morphologiques permettent de différencier les mâles et les femelles (Xiao-Yi Wang *et al.*, 2010). En effet, les mâles sont généralement de plus petites tailles, alors que l'abdomen des femelles est plus large et légèrement bombé.



Figure 1.1 : Imago de l'agrile du frêne, *Agrilus planipennis*, posé sur une feuille d'un frêne noir (*Fraxinus nigra* Marshall) (Photographie de C. Guertin, INRS).

#### 1.1 Cycle de développement

Le cycle de développement complet de l'agrile du frêne s'échelonne sur une période d'un à deux ans [voir Lyons et al. (2007)]. Au Canada, l'émergence des adultes se produit entre la fin mai et la fin juillet (Lyons et al., 2007). C'est alors qu'ils s'alimentent des feuilles de leur plantehôte pendant environ une semaine, période durant laquelle il y a une maturation sexuelle (Herms et al., 2014). Ensuite s'enclenche un comportement de reproduction polygame, c'est-à-dire que chaque mâle peut s'accoupler avec plusieurs femelles. Cependant, aucune information n'est disponible sur un caractère polyandrique chez les femelles. Après l'accouplement, les femelles s'alimentent de cinq à sept jours avant la période de ponte. Normalement, celles-ci pondent en moyenne de 40 à 70 œufs, mais ce nombre peut exceptionnellement atteindre 275 (Herms et al., 2014, Lyons et al., 2007). Les œufs sont déposés à la surface de l'écorce des frênes, dans de petites crevasses ou sous des écailles de l'écorce, où ils sont très peu visibles. La période d'incubation des œufs est d'environ deux semaines. À l'éclosion, les larves transpercent l'écorce afin de se rendre sous celle-ci, où elles auront un comportement cryptique. Elles s'alimenteront du phloème et du cambium en creusant des galeries en forme de S. Lorsque nombreuses, ces galeries empêchent la circulation des nutriments dans le phloème provoquant des nécroses et la mort éventuelle des arbres. Le développement larvaire se caractérise par une succession de quatre stades (Figure 1.2). Lors du dernier stade, aussi appelé stade prépupal, la larve creuse une chambre avant de se replier sur elle-même pour former un J, forme sous laquelle elle hiberne la plupart du temps. Comme l'agrile du frêne est holométabole, la pupe est le dernier stade avant la métamorphose et l'émergence des adultes. C'est durant cette transformation que les imagos acquerront leur couleur verte cuivrée caractéristique (Xiao-Yi Wang et al., 2010). Les pupes sont retrouvées sous l'écorce de la fin avril à la mi-juin. Lorsque l'insecte atteint sa maturité, il émerge de l'arbre en forant un trou de sortie en forme de D (Herms et al., 2014, Lyons et al., 2007).



Figure 1.2 : Développement larvaire. Augmentation de la taille des larves au cours du développement jusqu'au repliement en forme de J (A); repliement de la larve en forme de J (B); (Photographie R. Lavallée, RNCan)

#### 1.2 Historique de l'introduction de l'agrile du frêne en Amérique du Nord

*A. planipennis* est un insecte indigène du nord-est de la Chine, de la péninsule coréenne et de l'est de la Russie. Son introduction en Amérique du Nord résulterait d'activités anthropiques. Selon des études dendrochronologiques, ce ravageur serait arrivé sur le continent américain dans les années 90 (Siegert *et al.*, 2014), mais ce n'est qu'en 2002 que la détection officielle du ravageur est faite dans les municipalités frontalières de Détroit (MI, États-Unis) et de Windsor (ON, Canada) (Cappaert *et al.*, 2005). Dans les années qui ont suivi, ce ravageur s'est distribué dans plusieurs municipalités vers le nord-est (Canada) et le sud (États-Unis) de l'épicentre (Natural Resources Canada, 2016). Au Canada, l'agrile du frêne a été retrouvé en 2006 à London et dans plusieurs municipalités limitrophes. En 2007, on le détectait jusqu'à Toronto, et l'année suivante à Sault-Ste-Marie et à Ottawa. Au Québec, la première présence de ce ravageur a été observée en 2008 dans la municipalité de Carignan en Montérégie. En 2009, on décelait sa présence à Saint-Basile-le-Grand. Quelques années plus tard, soit en 2011, cet insecte a été détecté dans les villes de Montréal et Gatineau. À ce jour, l'agrile du frêne est retrouvé dans 24 états américains et deux provinces canadiennes. Une carte mise à jour de la dispersion de l'insecte est disponible pour consultation sur le site internet <u>http://www.emeraldashborer.info/</u>.

#### 1.3 Dommages causés par l'agrile

En Asie, l'agrile du frêne est considéré comme un ravageur secondaire, puisqu'il s'attaque principalement aux arbres affaiblis, malades ou morts. Par contre, en Amérique du Nord, il agit comme ravageur primaire en s'attaquant également aux frênes sains (Herms *et al.*, 2014). Les conséquences directes et indirectes des dommages causés par cet insecte sont nombreuses. Par exemple, en bordure des rivières, la perte des frênes cause l'érosion et le lessivage des sols, en plus de causer un réchauffement de l'eau dû à une plus grande exposition au soleil (Natural Resources Canada, 2016). En milieu forestier, les trous dans la canopée et l'augmentation des débris de bois au sol apportent des modifications au microenvironnement ainsi qu'aux cycles biogéochimiques du sol, en modifiant la concentration de certains éléments (azote, calcium) (Figure 1.3) (Gandhi *et al.*, 2010a). Ces modifications de l'environnement peuvent, entres autres, accroître le risque d'établissement de plantes envahissantes (Gandhi *et al.*, 2010a). De plus, les frênes constituent un habitat et une source de nourriture importante pour de nombreuses espèces d'oiseaux, de petits mammifères et d'insectes (Cappaert *et al.*, 2005). Par exemple, on estime à 282 le nombre d'espèces d'arthropodes se nourrissant du frêne, dont 43 de façon exclusive (Gandhi *et al.*, 2010b).



Figure 1.3 : Forte mortalité causée par l'agrile du frêne sur des arbres d'une plantation de frênes située dans le sud de l'Ontario (Photographie : C. Guertin, INRS).

En milieu urbain, les conséquences n'en sont pas moins néfastes. Pendant plusieurs décennies, les frênes étaient l'essence d'arbre la plus recommandée pour une plantation en milieu urbain grâce à sa tolérance au sel, à la sécheresse, au sol compact et aux variations de pH (MacFarlane *et al.*, 2005). Ces arbres servent d'écran contre le vent, améliorent la qualité de l'air, atténuent les fluctuations de températures et procurent un habitat pour de nombreux animaux (Natural Resources Canada, 2016). En raison du passage de l'agrile du frêne, on compte maintenant plusieurs rues dépourvues d'arbres (Figure 1.4).



Figure 1.4 : Rue de Montréal dont la majorité des arbres ont dû être coupés à la suite des ravages causés par l'agrile du frêne (Photographie : R. Lavallée, RNCan).

Au Canada, on estime les impacts socio-économiques de la présence de l'agrile du frêne à plus de deux milliards de dollars sur les 30 prochaines années pour le traitement, la coupe et le remplacement des frênes en milieu urbain (McKenney *et al.*, 2012). Ceci est sans compter les effets indirects tels que la diminution de la valeur des maisons et les coûts reliés aux problèmes de santé engendrés par des îlots de chaleur et une mauvaise qualité de l'air en milieu urbain. De plus, la création de zones réglementées pour le transport du bois de frêne entraîne des conséquences importantes pour plusieurs entreprises, telles que les scieries, les pépinières et les entreprises faisant l'exploitation du bois de frêne (Herms *et al.*, 2014).

#### 1.4 Méthodes de détection

La découverte de l'agrile du frêne en Amérique du Nord a été réalisée plusieurs années après son introduction, notamment dû au comportement cryptique des larves, mais aussi dû aux symptômes qui peuvent s'apparenter à des maladies affectant les frênes. La phase de latence entre le moment des premières attaques et la présence de symptômes visibles a encouragé le développement de méthodes de détections. Il en existe plusieurs et chacune d'entre elles comporte des avantages et des désavantages.

La détection visuelle de signes et de symptômes associés à la présence de l'agrile du frêne permet de repérer facilement les arbres infestés. La perte du feuillage ainsi que la présence de pousses adventives, de trous de sortie en forme de D, de galeries larvaires, de feuillages mangés par les agriles adultes et des signes d'alimentation de pic-bois et d'écureuils sont tous des indices visuels pouvant être associés à la présence de ce ravageur (Figure 1.5). Par contre, lorsque ces signes sont apparents, il est souvent trop tard pour mettre en place des moyens de lutte pour contrôler la population de l'agrile. Elle s'est déjà amplement dispersée sur l'ensemble de l'arbre (Lyons *et al.*, 2007).



Figure 1.5 : Signes et symptômes pour la détection de l'agrile du frêne. Pousses adventives (A); trou de sortie en forme de D (B); galeries larvaires (C); agrile du frêne s'alimentant de feuillage de frêne (D); signes d'alimentations des pics-bois et des écureuils (E); frêne décimé par les attaques de l'agrile du frêne (F). (Photographies : Robert Lavallée, RNCan)

L'écorçage d'une section du tronc de l'arbre permet de déceler la présence des galeries larvaires de l'agrile du frêne sur des arbres présentant des signes d'attaques (Lyons *et al.*, 2007). Cette méthode consiste en la délimitation d'une fenêtre de 10 cm par 10 cm afin d'en retirer l'écorce jusqu'au cambium, puis d'observer la présence de galeries. Cette méthode a le désavantage d'endommager l'arbre et les risques associés au diagnostic de faux négatifs sont élevés (Ryall *et al.*, 2011), notamment chez les arbres ayant un fort diamètre. Finalement, en

considérant que l'agrile du frêne s'attaque d'abord à la cime de l'arbre, l'application efficace de cette méthode est aussi limitée par des contraintes logistiques du travail en hauteur.

L'annélation est une méthode qui consiste en l'utilisation d'un arbre stressé comme appât (Herms *et al.*, 2014, Marshall *et al.*, 2011). Un anneau d'écorce est prélevé sur toute la circonférence d'un arbre, provoquant ainsi un stress et son affaiblissement. Les agriles, ayant une préférence pour ceux-ci, devraient être attirés vers cet arbre, augmentant ainsi la possibilité de déceler la présence de l'insecte dans la zone ciblée. Par contre, en plus d'endommager les arbres, cette méthode ne permet pas une détection instantanée et doit être couplée aux autres techniques de détection.

L'équipe de Ryall *et al.* (2011) ont développé une technique de détection pour les arbres en milieu urbain. Cette technique consiste à prélever deux branches de frêne de 5 à 8 cm de diamètre au milieu de la canopée et d'écorcer les 50 premiers centimètres à partir de la base afin de détecter la présence de galeries larvaires. Cette méthode permet de déceler la présence de l'agrile du frêne, avant même que les premiers symptômes ne soient apparents sur les arbres. Par contre, elle demande un travail laborieux et implique des coûts importants reliés aux ressources humaines et matérielles. De plus, elle est limitée par la disponibilité de branches de bon calibre. Ce dernier point est certainement un facteur limitant qui peut conduire à l'observation de faux négatifs.

Finalement, plusieurs équipes tentent de mettre au point des pièges de plus en plus efficaces pour la capture des insectes. La couleur du piège, sa forme (*p. ex.* prisme ou entonnoirs multiples), les appâts utilisés (phéromones, kairomones) et l'emplacement du piège dans l'arbre sont autant d'éléments faisant l'objet de plusieurs études (Figure 1.6) (Crook *et al.*, 2012, Herms *et al.*, 2014). Le piège à entonnoirs multiples de type Lindgren comporte de nombreux avantages par rapport au piège prisme, puisqu'il est réutilisable, facilement transportable et permet à l'utilisateur de récupérer rapidement les insectes collectés plutôt que d'inspecter attentivement l'ensemble du piège collant sur le terrain (Francese *et al.*, 2011). De plus, Lyons *et al.* (2012) ont montré que ce piège, sur lequel une couche d'un polymère fluoré était appliquée, pouvait capturer dix fois plus d'insectes que le piège prisme communément utilisé dans les programmes de détection. Tous les efforts déployés pour la détection de l'agrile du frêne sont importants, car ils sont la prémisse au développement de stratégies efficaces de lutte contre ce ravageur exotique envahissant.



Figure 1.6 : Pièges appâtés d'une phéromone et d'une kairomone permettant la capture de l'agrile du frêne. Piège à entonnoirs multiples de type Lindgren à la cime d'un frêne à Gatineau (A); piège prisme prêt à être installé à la cime d'un frêne à Montréal (B) (Photographies : A. Bergeron, INRS (A) et R. Lavallée, RNCan (B)).

# CHAPITRE 2 : MICROORGANISMES ASSOCIÉS AUX INSECTES

Les insectes vivent en étroite association avec des microorganismes tels que des bactéries, des champignons, des levures et des nématodes (Janson *et al.*, 2008, Six, 2013). Selon la théorie de l'hologénome, l'entité subissant l'évolution ne serait pas l'organisme, mais plutôt l'holobionte qui est cette association entre un hôte et ses microorganismes (Zilber-Rosenberg *et al.*, 2008). Chez les insectes, les microorganismes sont indispensables afin de leur permettre de coloniser et de s'adapter à de nouvelles niches écologiques, d'élargir leur éventail alimentaire, de les protéger contre les prédateurs et bien plus encore (Brownlie *et al.*, 2009, Douglas, 2009, Engel *et al.*, 2013, Popa *et al.*, 2012). Ces microorganismes symbiotiques peuvent être des mutualistes, apportant des bénéfices à leur hôte; des commensaux, n'améliorant et ne nuisant pas à leur hôte; ou des parasites, affectant leur hôte pour leurs propres bénéfices (Brownlie *et al.*, 2009). Les prochaines lignes feront état majoritairement des relations symbiotiques entre les bactéries et les insectes.

#### 2.1 Transmission verticale et horizontale

L'association entre les microorganismes symbiotiques et les insectes peut résulter d'une transmission verticale, de la mère à la progéniture, ou d'une transmission horizontale, acquise au cours du développement. La transmission verticale est requise pour les microorganismes obligatoires, c'est-à-dire les microorganismes qui ne peuvent pas survivre en dehors de leur hôte. Bon nombre de ces bactéries, localisées dans des structures spécialisées appelées bactériocytes, se retrouvent principalement associées aux cellules adipeuses et à celles de la paroi de l'intestin des insectes (Gibson *et al.*, 2010, Moran *et al.*, 2008). Ces bactéries endosymbiotiques sont le fruit d'une longue histoire évolutive ayant permis de réduire la taille de leur génome (<1MB) et de conserver uniquement les gènes essentiels à leur hôte (Moran *et al.*, 2008). C'est la symbiose avec les insectes et les autres microorganismes symbiotiques qui assurent à ces bactéries les éléments de synthèses génétiques qui ont été délaissés, mais qui sont requis pour leur survie et leur reproduction. La relation symbiotique entre la mouche pisseuse, *Homalodisca coagula* Say, et deux endosymbiotes obligatoires, soit *Baumania cicadellinicola* et *Sulcia muelleri*, est un bon exemple de cette coévolution (Wu *et al.*, 2006). La bactérie *B. cicadellinicola* consacre

essentiellement ses activivités à la biosynthèse de vitamines et de cofacteurs, alors que *S. muelleri* est en mesure de produire presque tous les acides aminés essentiels pour l'hôte. Cette relation tripartite entre deux symbiotes bactériens et un insecte hôte démontre bien la relation mutualiste qui peut exister au sein d'un holobionte.

En contraste avec la transmission verticale, certaines bactéries sont transmises horizontalement. Ces bactéries sont acquises par l'insecte via son environnement, son alimentation et son comportement social (Gibson *et al.*, 2010). Il est plus difficile de les étudier puisque leur présence et leur abondance peuvent varier d'un individu à l'autre (Moran *et al.*, 2008). Ces bactéries possèdent tous les éléments génétiques nécessaires à leur survie et à leur reproduction. Elles possèdent donc un génome de plus grande taille que celui des endosymbiotes (>1MB). Leur historique de coévolution avec l'hôte serait plus récent et elles auraient la capacité d'habiter différents tissus et types cellulaires de l'insecte. Malgré leurs fonctions souvent présumées facultatives, celles-ci n'en sont pas moins bénéfiques pour leur hôte. En effet, elles sont souvent impliquées dans des rôles de protection contre le stress et les ennemis naturels (Brownlie *et al.*, 2009), dans l'apport d'éléments nutritionnels essentiels (Douglas, 2009), dans la communication entre individus (Dillon *et al.*, 2002) et bien d'autres.

#### 2.2 Rôles des microorganismes associés aux insectes

#### 2.2.1 Nutrition

Les insectes occupent des niches écologiques très diversifiées. Cependant, les éléments nutritifs présents dans leur habitat ne permet pas de toujours répondre à tous leurs besoins alimentaires. Certains nutriments sont présents sous des formes trop complexes pour être assimilables par l'insecte, alors que d'autres, essentiels à leur développement, sont absents. Afin de combler leurs besoins nutritionnels, les insectes vivent en symbiose avec des bactéries qui leur permettent de digérer les éléments complexes et d'acquérir des éléments nutritifs qui, autrement, ne seraient pas accessibles pour eux (Douglas, 2009). D'ailleurs, plusieurs bactéries symbiotiques possèdent les gènes nécessaires à la synthèse d'acides aminés, de vitamines et de cofacteurs essentiels à l'hôte (Shigenobu *et al.*, 2000, Wu *et al.*, 2006). C'est notamment le cas des bactéries du genre *Blattabacterium* associées à la blatte américaine, *Periplaneta americana* 

Linnaeus, qui possèdent, entre autres, les gènes permettant la synthèse de presque tous les acides aminés essentiels à leur hôte (Sabree *et al.*, 2009).

Chez les insectes ayant une alimentation faible en azote, la présence de bactéries capable de le fixer et de le recycler est essentielle (Sabree *et al.*, 2009, Shigenobu *et al.*, 2000). Par exemple, la bactérie *Raoultella terrigena* associée à l'intestin de *Dendroctonus rhizophagus* Thomas et *Dendroctonus valens* LeConte, possède des gènes qui codent pour des enzymes permettant la fixation de l'azote atmosphérique afin de le rendre assimilable pour l'hôte (Morales-Jiménez *et al.*, 2013). De plus, les bactéries *Pseudomonas fluorescens*, *Serratia proteomaculans* et *Rahnella aquatilis*, également présentent dans l'intestin de ces mêmes hôtes, seraient en mesure de recycler l'acide urique.

Plusieurs insectes, particulièrement ceux qui se nourrissent d'arbres et de plantes, requièrent la présence de bactéries capables d'hydrolyser la cellulose et d'autres polymères récalcitrants en molécules organiques simples, afin d'en récupérer le maximum de nutriments (Delalibera *et al.*, 2005, Morales-Jiménez *et al.*, 2009, Vasanthakumar *et al.*, 2008). Cette activité cellulolytique a été démontrée *in vitro* chez plusieurs bactéries, notamment *Stenotrophomonas maltophilia*, *Ponticoccus gilvus* et *Kocuria marina* associées à *D. rhizophagus* (Morales-Jiménez *et al.*, 2012). Plusieurs bactéries associées au ver à soie, *Bombyx mori* Linnaeus, sont également en mesure de dégrader des polymères récalcitrants *in vitro*, comme la cellulose, le xylane, la pectine et l'amidon (Anand *et al.*, 2010). De plus, une espèce de *y*-proteobactérie associée à l'abeille européenne, *Apis mellifera* Linnaeus, s'est spécialisée au cours de l'évolution afin de permettre la dégradation de la paroi de pectine du pollen (Engel *et al.*, 2012, Engel *et al.*, 2013).

#### 2.2.2 Défense

La présence de bactéries mutualistes chez les insectes procure à ces derniers une protection supplémentaire, voire essentielle, contre différents pathogènes. En effet, les insectes sont sans cesse confrontés à la présence de champignons antagonistes (Dillon *et al.*, 1988, Kaltenpoth *et al.*, 2010, Scarborough *et al.*, 2005, Scott *et al.*, 2008), de virus (Hedges *et al.*, 2008, Teixeira *et al.*, 2008), de bactéries pathogènes (Dillon *et al.*, 2005), de parasitoïdes (Degnan *et al.*, 2009, Oliver *et al.*, 2003) et de nématodes (Jaenike *et al.*, 2010). Afin de les contrer, les insectes vivent en association avec des bactéries qui mettent à profil leurs compétences de diverses façons aux bénéfices de leur hôte. Par exemple, la guêpe solitaire, *Philanthus triangulum* Fabricius, cultive une espèce de *Streptomyces* dans les glandes de ses antennes, puis la dépose

sur les parois du couvain avant la ponte (Kaltenpoth *et al.*, 2005). La bactérie, sécrétant potentiellement des antibiotiques, protège la progéniture tout au long de son développement. De plus, une espèce de *Streptomyces* a également été isolée du jardin fongique des fourmis coupe-feuilles, *Acromyrmex octospinosus* Reich. Celle-ci produit de la candicidine afin d'inhiber la croissance du champignon pathogène *Escovopsis* (Currie *et al.*, 2003, Haeder *et al.*, 2009, Mueller, 2012). Le dendroctone de l'épinette, *Dendroctonus rufipennis* Kirby, quant à lui, sécrète oralement des bactéries, tel que *Micrococcus luteus*, afin d'inhiber la prolifération de champignons antagonistes se trouvant dans son habitat (Cardoza *et al.*, 2006). En milieu naturel, la drosophile, *Drosophila neotestacea* Grimaldi, James & Jaenike, était, jusqu'à récemment, régulièrement sous l'effet stérilisant du nématode *Howardula aoronymphium* (Jaenike *et al.*, 2010). Par contre, une association de plus en plus fréquente avec la bactérie *Spiroplasma* protègerait l'insecte contre les effets néfastes de ce nématode.

En plus de devoir combattre plusieurs antagonistes biologiques, les insectes doivent également faire face à l'exposition à certains composés chimiques. Que ce soit contre les défenses naturelles des plantes qu'ils attaquent ou contre les insecticides chimiques utilisés dans différents environnements, les insectes vivent en association avec des bactéries leur permettant de contrer ces obstacles. Les bactéries Pseudomonas, Rahnella, Serratia et Burkholderia associées à Dendroctonus ponderosae Hopkins ont démontré des habiletés à dégrader les terpènes de la plante hôte. Elles permettent donc à l'insecte d'exploiter une niche écologique subcortical qui, autrement, ne serait pas propice à leur établissement (Adams et al., 2013, Adams et al., 2011). En milieu agricole, les producteurs utilisent de nombreux insecticides afin d'éliminer les insectes nuisibles à leur récolte. Par contre, une mauvaise utilisation de ces produits peut mener au développement de résistance chez certains insectes. Par exemple, le fenitrothion est un insecticide communément utilisé partout dans le monde (Kikuchi et al., 2012). Par contre, on constate la prolifération de bactéries résistantes à cet insecticide dans le sol, telle que Burkholderia. Celle-ci est également acquise horizontalement par Riptortus pedestris Fabricius lors de son développement. L'acquisition de la bactérie Burkholderia résistante au fenitrothion entraine une résistance de l'insecte à cet insecticide. À l'opposé, certains insecticides sont efficaces seulement en présence de certaines bactéries. Par exemple, le bioinsecticide Bacillus thuringiensis est efficace contre les larves de la spongieuse seulement si celles-ci vivent en association avec la bactérie Enterobacter (Broderick et al., 2006).

16

#### 2.2.3 Reproduction et développement

La bactérie *Wolbachia* est l'une des bactéries intracellulaires les plus répandues chez les insectes et les arthropodes (Hilgenboecker *et al.*, 2008). Par contre, celle-ci agit, le plus souvent, à titre de parasite en manipulant les capacités reproductives de ses hôtes. En effet, cette bactérie peut induire une féminisation des mâles, une parthénogenèse, la mort de la progéniture masculine ainsi qu'une incompatibilité reproductive entre deux partenaires qui ne sont pas infectés par la même souche de *Wolbachia* (Hilgenboecker *et al.*, 2008).

La bactérie commensale *Lactobacillus plantarum* associée à la drosophile, *Drosophila melanogaster* Meigen, influence les préférences d'accouplement de son hôte. En effet, des populations de drosophile ayant des régimes alimentaires différents ont tendance à se reproduire avec des partenaires ayant le même régime alimentaire qu'elles (Sharon *et al.*, 2010). Cet effet est directement relié à la présence de la bactérie *L. plantarum* dans l'intestin de l'hôte, qui est responsable d'une augmentation de la production de phéromones sexuelles.

De plus, plusieurs études réalisées sur le microbiome de la drosophile ont démontré un lien direct entre la présence de certains symbiotes et la grosseur des insectes, la durée du développement larvaire et la durée de vie (Brummel *et al.*, 2004, Shin *et al.*, 2011, Storelli *et al.*, 2011).

#### 2.2.4 Communication

Certaines bactéries permettent la communication entre les insectes d'une même espèce par l'entremise de phéromones. Le plus souvent, ce sont des phéromones sexuelles permettant d'attirer le sexe opposé pour la reproduction. Par contre, il en existe d'autres types tels que les phéromones d'agrégation. En effet, les bactéries *Pantoea agglomerans, Klebsiella pneumoniae pneumoniae* et *Enterobacter cloacae* retrouvées dans les excréments du criquet pèlerin, *Schistocerca gregaria* Forsskal, sécrètent des précurseurs des phéromones d'agrégation (Dillon *et al.*, 2002). Ces phéromones permettent les attaques massives de ce ravageur. Par contre, chez le criquet migrateur, *Locusta migratoria manilensis* Meyen, une infection par le parasite *Paranosema* (Nosema) *locustae* Canning acidifie l'intestin de l'hôte, empêchant ainsi la croissance des bactéries responsables de la production de ces phéromones (Shi *et al.*, 2014). Le pouvoir communicatif des bactéries n'est pas toujours à l'avantage de l'hôte. En effet, certaines bactéries sécrètent des kairomones qui servent de signaux entre différentes espèces d'insectes.

17

Par exemple, la bactérie *Staphylococcus sciuri* retrouvée dans le miellat du puceron vert du pois, *Acyrthosiphon pisum* Harris, agit à titre d'attractant pour les ennemis naturels de son hôte (Leroy *et al.*, 2011).

Somme toute, les bactéries jouent un rôle primordial dans la vie des insectes à plusieurs niveaux. L'alimentation, la protection, la reproduction, le développement et la communication chez les insectes sont tous influencés par leur microbiome. Toutefois, il est important de considérer qu'une bactérie donnée n'effectue pas nécessairement le même rôle chez tous les insectes et qu'un rôle donné peut être effectué par différentes bactéries (Six, 2013).

# 2.3 Processus stochastiques et déterministes définissant la structure microbienne des insectes

La structure de la communauté microbienne des insectes est définie par des processus stochastiques et déterministes qui influencent leur niche écologique. Les processus stochastiques sont le fruit du hasard. Par exemple, une colonisation par chance ou une extinction aléatoire de certaines espèces microbiennes dans la niche écologique sont des processus stochastiques. Pour sa part, le déterminisme est défini comme étant un processus écologique qui implique des mécanismes non aléatoires. Ces mécanismes sont des filtres imposés par l'habitat. Par exemple, le climat, la composition du sol et la présence d'autres espèces animales ou microbiennes sont des facteurs induisant des processus déterministes.

Plusieurs études ont démontré l'impact de l'environnement sur la structure de la communauté microbienne (Corby-Harris *et al.*, 2007, Montagna *et al.*, 2015a, Rani *et al.*, 2009a, Xiang *et al.*, 2006). Par exemple, neuf populations de *Homalodisca vitripennis* Germar (synonyme de *H. coagula*), originaire de différents vignobles, ont démontré des différences significatives dans la composition de leur communauté bactérienne en fonction de leur provenance (Welch *et al.*, 2015). D'autres études ont mis de l'avant les différences dans la structure microbienne d'insectes en provenance de milieu naturel *versus* ceux élevés en laboratoire (Bansal *et al.*, 2014, Montagna *et al.*, 2015b, Rani *et al.*, 2009a, Xiang *et al.*, 2006). En effet, les populations naturelles sont sujettes à de nombreuses variations de conditions ambiantes, alors que les populations d'élevage sont dans des environnements contrôlés. Les événements stochastiques sont présents à plus grande échelle et les événements déterministes varient davantage en milieu naturel. Par exemple, plusieurs espèces bactériennes associées au puceron du soja sont retrouvées en plus faible

abondance dans les populations en provenance d'élevage de laboratoire comparativement aux populations naturelles (Bansal *et al.*, 2014).

Les processus déterministes peuvent être reliés à des facteurs abiotiques ou biotiques. La température, les précipitations, l'altitude, la latitude et la longitude sont tous des facteurs abiotiques, alors que le stade de développement de l'insecte, l'alimentation, la présence de prédateurs ou de pathogènes et le niveau de la population de l'insecte sont des facteurs biotiques.

#### 2.3.1 Température et précipitations

Les changements climatiques ont des effets importants sur le comportement des insectes, notamment en causant une modification de leur dispersion géographique (Musolin, 2007). Cette modification pourrait, par exemple, entrainer la mort de nouveaux arbres cibles qui n'ont pas d'historique de co-évolution avec l'insecte et qui, ainsi, n'ont pas développé de moyens de défence appropriés contre ces ravageurs. Comprendre comment les changements climatiques affecteront les interactions entre les insectes et leurs symbiotes pourrait permettre de mieux prévoir les impacts engendrés par de tels changements. Plusieurs études se sont intéressées aux effets du changement de température sur la communauté bactérienne des insectes (Jia *et al.*, 2013). Par exemple, une augmentation de température de 5 °C réduit significativement le nombre de symbiotes bactérien, détectable par PCR, de l'intestin d'une population d'*Acrosternum hilare* Say et d'une population de *Murgantia histrionica* Hahn, maintenues en laboratoire (Prado *et al.*, 2010).

Les différences observées au niveau du microbiome entre différentes populations d'insectes sont souvent le résultat d'une combinaison de plusieurs facteurs. En effet, la flore bactérienne de l'intestin des larves d'*Holotrichia parallela* Motschulsky varie selon la provenance des populations (Huang *et al.*, 2013). Ces différences sont expliquées, entre autres, par une variation au niveau de la température et des précipitations moyennes annuelles ainsi qu'un changement de la composition du sol (pH, azote et carbone) pour chacun des sites d'échantillonnage. Des résultats similaires ont été observés dans plusieurs populations naturelles d'*A. pisum* (Tsuchida *et al.*, 2002). En effet, la présence de l'endosymbiote PAUS (*Pea Aphid U-type Symbiont*) varie selon la position géographique des différentes populations de l'insecte, et semble être positivement corrélée avec la température et les précipitations moyennes annuelles ainsi qu'avec l'espèce de la plante hôte.

#### 2.3.2 Altitude, latitude et longitude

Bien qu'elles soient souvent liées à un changement de climat local, l'altitude, la latitude et la longitude influencent la microflore bactérienne et fongique associée aux insectes. En effet, la comparaison de cinq populations du complexe *Cryptocephalus marginellus* Olivier (Coleoptera : Chrysomelidae), a démontré des divergences au sein de la communauté bactérienne de ces insectes en relation avec l'altitude à laquelle les individus ont été récoltés (Montagna *et al.*, 2015b).

Dans une autre étude mettant en relation *D. ponderosae* et ses symbiotes fongiques, un changement d'abondance relative de deux des trois champignons ophiostomatoïdes symbiotiques a été observé le long d'un gradient de latitude et, de façon moins prononcée, le long d'un gradient d'altitude (Roe *et al.*, 2011). En effet, l'abondance relative de *Leptographium longiclavatum* augmente lors de l'accroissement de la latitude, alors que celle de *Grosmannia clavigera* est plutôt associée à l'élévation du site d'échantillonnage. D'autre part, Adams *et al.* (2010) se sont intéressés à la communauté bactérienne associée à huit populations de *D. valens*, le long d'un gradient de longitude. Les variations observées au niveau du microbiome entre les populations étaient directement corrélées avec la distance entre les sites d'échantillonnage, c'est-à-dire que les insectes des sites les plus rapprochés avaient une structure microbienne plus similaire que ceux provenant de sites plus isolés.

#### 2.3.3 Stade de développement de l'insecte

La communauté bactérienne et fongique varie au cours du développement de l'insecte (Delalibera *et al.*, 2007, Rani *et al.*, 2009a). Ces modifications pourraient être en lien avec un changement physiologique de l'insecte ainsi qu'avec un mode de vie et des besoins nutritionnels différents. Chez l'agrile du frêne, 43 et 22 espèces bactériennes sont retrouvées exclusivement en association avec, respectivement, les prépupes et les adultes (Vasanthakumar *et al.*, 2008). La communauté bactérienne associée aux larves est, quant à elle, un mélange d'espèces bactériennes associées aux deux autres stades. Ces résultats suggèrent que l'insecte conserve certains symbiotes tout au long de son développement, formant une communauté bactérienne de base, alors que d'autres sont spécifiques aux stades de développement. Chez le scolyte *D. rhizophagus*, la flore bactérienne de l'intestin de la pupe est moins diversifiée que celle des larves et des adultes (Morales-Jiménez *et al.*, 2012). En effet, seulement la bactérie *Rahnella aquatilis* 

a été observée chez la pupe. Cette observation pourrait être le résultat d'un ralentissement de l'activité métabolique reliée à ce stade de métamorphose. Enfin, la variation du microbiome ne s'effectue pas seulement entre les stades de développement de l'insecte, mais également au cours du développement larvaire. En effet, la flore bactérienne de l'intestin du premier stade larvaire de *H. parallela* est moins complexe que celle du deuxième et troisième stade, alors que peu de différences sont observées entre ces derniers (Huang *et al.*, 2013). Ces variations sont expliquées par une plus grande alimentation et une augmentation de la taille des individus au cours du développement.

#### 2.3.4 Alimentation

Il est bien connu que les bactéries symbiotiques intestinales exercent des rôles indispensables à la digestion et à l'assimilation des nutriments (Douglas, 2009). Il n'est donc pas surprenant de constater des variations au niveau de la flore microbienne des insectes en fonction de leur alimentation. Par exemple, des larves de *Lymantria dispar* Linnaeus ont été alimentées, en conditions contrôlées, avec une diète artificielle stérile, de tremble, de mélèze, de chêne blanc ou de saule (Broderick *et al.*, 2004). Des différences significatives ont été observées au niveau de la structure de la communauté bactérienne associée aux insectes en fonction du régime alimentaire. Des résultats similaires ont été observés chez des populations de *Phylloxera notabilis* Pergande en provenance de pacaniers et de pacaniers aquatiques (Medina *et al.*, 2011). En effet, les bactéries *Pantoea agglomerans* et *Serratia marcescens* sont les plus abondantes dans la population en provenance des pacaniers, alors qu'elles sont totalement absentes dans la population de pacaniers aquatiques.

La source d'alimentation des insectes est elle-même porteuse d'une communauté bactérienne qui colonise l'intestin des individus qui s'en alimente. Par exemple, 12 genres bactériens sur 13 associés aux feuilles de palmier sont également retrouvés dans l'intestin du charançon rouge des palmiers, *Rhynchophorus ferrugineus* (Olivier) (Montagna *et al.*, 2015a). Il en est de même pour la flore bactérienne de la noctuelle de la tomate, *Helicoverpa armigera* Hübner (Priya *et al.*, 2012). Les individus s'alimentant de différentes espèces de plantes ont des communautés bactériennes divergentes, alors que le microbiome de la plante hôte ressemble davantage à celui de l'insecte.

Certaines espèces d'insectes, plus particulièrement les moustiques, ont des habitudes alimentaires qui varient en fonction du sexe des individus. Par exemple, la femelle d'*Anopheles*
*stephensi* Liston est anautogène, c'est-à-dire qu'elle doit s'alimenter de sang avant l'oviposition, alors que les mâles ne s'en alimentent jamais (Rani *et al.*, 2009a). Ces différentes habitudes alimentaires résultent en des divergences au niveau de leur flore bactérienne (Rani *et al.*, 2009a, Zouache *et al.*, 2011).

#### 2.3.5 Pression exercée par les parasites

La fréquence d'un symbiote au sein des populations et au cours des générations peut augmenter, si celui-ci apporte un bénéfice net à l'insecte. Par exemple, le puceron vert du pois vit en association avec le symbiote *Serratia symbiotica* ainsi qu'avec le symbiote facultatif, *Hamiltonalla defensa*, qui le protègent contre le parasitoïde *Aphidius ervi* (Oliver *et al.*, 2008). Lorsque le puceron du pois est soumis à une exposition répétée à *A. ervi*, l'abondance des deux symbiotes augmente. Par contre, en absence d'exposition, la fréquence des deux bactéries diminue significativement.

#### 2.3.6 Niveau de la population des insectes

Des études ont suggéré que le niveau de la population, c'est-à-dire une population endémique *versus* épidémique, pourrait influencer la communauté fongique associée à différents scolytes. En effet, la fréquence d'association du champignon *Leptographium abietinum* et d'une espèce de *Pesotum* avec le dendroctone de l'épinette varie significativement entre les populations endémiques, épidémiques et transitoires (Aukema *et al.*, 2005). Chez le bostryche typographe, *lps typographus* Linnaeus, certaines espèces fongiques sont présentes chez les populations endémiques et épidémiques, alors que d'autres sont spécifiques à la population endémique (Giordano *et al.*, 2013).

# CHAPITRE 3 : HYPOTHÈSE ET OBJECTIF

La détection de l'agrile du frêne en milieu urbain et forestier pose un défi de taille aux municipalités qui souhaitent contrôler ce ravageur. Son comportement subcortical rend difficilement possible sa détection précoce, ce qui favorise sa propagation. Un grand intérêt est alors porté à l'identification des populations d'agrile du frêne sur une base géographique, afin de mieux comprendre la dispersion dans le temps et l'espace de ce ravageur. En tenant compte du grand potentiel d'information relié à l'étude des microorganismes ainsi qu'aux facteurs stochastiques et déterministes qui influencent ceux-ci chez les insectes, un intérêt est porté sur la caractérisation des bactéries associées à l'agrile du frêne selon le niveau d'établissement de différentes populations ainsi qu'au niveau du sexe des individus.

Ainsi, l'hypothèse suivante a été formulée :

La diversité du microbiome d'*Agrilus planipennis* Fairmaire (Coleoptera : Buprestidae) varie en fonction du sexe des individus et est un indicateur des liens entre les populations de l'insecte provenant de différentes régions.

Afin de répondre à cette hypothèse, les objectifs suivants ont été posés :

- 1. Caractérisation de la communauté bactérienne d'*Agrilus planipennis* par séquençage à haut débit.
- 2. Comparaison de la diversité et de la composition bactérienne entre différentes populations d'agrile du frêne.
- 3. Comparaison de la diversité et de la composition bactérienne entre les mâles et les femelles de cet insecte.

# Caractérisation de la communauté bactérienne associée à l'agrile du frêne, *Agrilus planipennis* Fairmaire (Coleoptera : Buprestidae) :

## ARTICLE

## CHAPITRE 4 :

# BACTERIAL COMMUNITY DIVERSITY ASSOCIATED WITH DIFFERENT POPULATIONS OF *AGRILUS PLANIPENNIS* FAIRMAIRE (COLEOPTERA : BUPRESTIDEA)

### DIVERSITÉ DE LA COMMUNAUTÉ BACTÉRIENNE ASSOCIÉE À DIFFÉRENTES POPULATIONS D'AGRILUS PLANIPENNIS FAIRMAIRE (COLEOPTERA : BUPRESTIDEA)

#### Auteurs

Amélie Bergeron<sup>1</sup>, Robert Lavallée<sup>2</sup>, Philippe Constant<sup>1</sup>, Claude Guertin<sup>1</sup>

#### Affiliation

<sup>1</sup> INRS- Institut Armand-Frappier, Laval, QC, Canada, H7V 1B7

<sup>2</sup> Natural Resources Canada, Canadian Forest Service, Quebec, QC, Canada, G1V 4C7

#### **Contribution des auteurs**

AB a participé à l'échantillonnage sur le terrain, aux traitements des échantillons en laboratoire, aux traitements des données et à l'analyse des résultats. RL, PC et CG ont élaboré le projet de recherche. RL a participé à l'échantillonnage sur le terrain. PC a participé l'analyse des résultats. CG a participé à l'échantillonnage sur le terrain et à l'analyse des résultats. L'article a été écrit par AB, puis corrigé par CG, PC et RL.

#### Journal

ISME journal

#### État actuel

Soumis le 28 avril 2016

#### 4.1 Résumé

L'agrile du frêne (ADF), Agrilus planipennis Faimaire (Coleoptera: Buprestidae), est un insecte exotique envahisseur en Amérique du Nord et il forme un holobionte avec ses microorganismes associés. Ce microbiome joue un rôle crucial dans la biologie de l'insecte et peut contribuer à son établissement sur un arbre hôte. Peu d'informations sont connues sur le microbiome de l'ADF. Le but de cette étude était d'identifier les communautés bactériennes associées à quatre populations naturelles et une population élevée en laboratoire d'ADF, en utilisant la technologie de séquençage Illumina MiSeq du gène codant pour l'ARNr 16S. La composition des communautés microbiennes des échantillons en provenance de populations naturelles semble plus stochastique que les réplicas en provenance de populations élevées en laboratoire suggèrant que les conditions d'élevage exercent un effet déterministe (basé sur la niche) sur la structure de la communauté microbienne. Un dendrogramme de similarité Unweighted Pair Group Method with Arithmetic (UPGMA) sépare les réplicas des populations d'ADF en deux écotypes distincts, basés sur leur profil de distribution des OTU. Une structure microbienne commune à tous les échantillons a été définie et elle comprend les genres bactériens Erwinia, Serratia, Tatumella, Hafnia, Enterobacter, Pantoea, Kluyvera, Pseudomonas, Enterococcus, Corynebacterium, Neptunomonas et Sporosarcina. Cependant, des OTU indicateurs ont été identifiés pour chacun des écotypes. Il y a une évidence claire que la niche et les processus neutres influencent l'association des bactéries avec leur insecte hôte. Des facteurs biotiques et abiotiques sont proposés pour expliquer la différence observée dans l'assemblage des communautés bactériennes associée avec l'ADF.

#### 4.2 Abstract

The emerald ash borer (EAB), *Agrilus planipennis* Fairmaire (Coleoptera: Buprestidae) is an exotic pest in North America and forms a holobiont with its associated microorganisms. These play crucial roles in the insect biology, and may contribute to its establishment success on its host. Little is known about the EAB microbiome. The aim of this study was to identify bacterial communities, associated with EAB from four natural and one laboratory reared populations through the construction of a library of 16S rRNA genes using Illumina MiSeq sequencing of PCRamplified fragments. The composition of microbial communities among samples from natural populations appeared more stochastic than from laboratory population suggesting that rearing conditions seem to exert deterministic (niche-based) effect on the microbial community structure. An Unweighted Pair Group Method with Arithmetic mean agglomerative clustering method separated replicates of EAB populations in two distinct ecotypes based on their operational taxonomic units distribution profiles. A core microbiome has been defined and comprised bacterial genera *Erwinia*, *Serratia*, *Tatumella*, *Hafnia*, *Enterobacter*, *Pantoea*, *Kluyvera*, *Pseudomonas*, *Enterococcus*, *Corynebacterium*, *Neptunomonas*, and *Sporosarcina*. However, OTU indicators have been identified for each ecotype. There is clear evidence that niche and neutral processes influenced bacterial association with their insect host. Biotic and abiotic factors are proposed to explain the differences observed in bacterial community assemblages associated with EAB.

#### 4.3 Introduction

The phloem-feeding beetle emerald ash borer (EAB), *Agrilus planipennis* Fairmaire (Coleoptera: Buprestidae), is a highly destructive invasive species. It has already killed several millions of ash trees (*Fraxinus* spp.) since introduced in North America in the 1990s (Siegert *et al.* 2014). The insect was first detected in 2002, several years after its introduction from Asia, in Detroit, Michigan (United States) and Windsor, Ontario (Canada) (Haack *et al.* 2002). During the following years, EAB spread northeast and, in 2008 its first occurrence was reported in the province of Quebec in Carignan (Figure 4.1). To date, 24 American states and two Canadian provinces are affected by the presence of this exotic insect (Anonymous 2016). The larvae chew on phloem and xylem tissues of the host tree and are responsible for the most extensive damages conducive to the death of ash trees (Bauer et al. 2004; Cappaert et al. 2005; Duan JJ *et al.* 2010;).



Figure 4.1 : Location of sample sites. SM = Sault-Ste. Marie, Ontario; MC = Middlesex County, Ontario; SBLG = Saint-Basile-le-Grand, Quebec; GAT = Gatineau, Quebec and MTL = Montreal, Quebec.

It is now widely accepted that insect-associated microbes are beneficial to several aspects of survival and fitness of their host (Weiss and Aksoy 2011). In microbial ecology, the concept of holobiont is defined as the host and his associated microorganisms (Six 2013). It is now clearly known that the microbiome of insects is essential for their successful establishment on a host plant (Adams et al. 2013, Frago et al. 2012). Some associated microorganisms play crucial roles in the protection and the fitness of insect (Popa et al. 2012, Rajagopal 2009) by having, for example, important nutritional functions (Douglas 2009, Morales-Jiménez et al. 2012, Vasanthakumar et al. 2008); like the cellulolytic activity required for cellulose digestion in Agrilus planipennis seems to be mediated by bacteria found in the gut, such as Streptomyces spp., Erwinia spp. and Burkholderia cepacia (Vasanthakumar et al. 2008). Bacteria may also contribute to detoxification of tree defense molecules especially under subcortical environment as observed in Dendroctonus ponderosae (Adams et al. 2013). They may also secrete antibiotics (Seipke et al. 2013) or antifungal (Cardoza et al. 2006) compounds that may provide protection to the insect against antagonistic bacteria and fungi. Furthermore, bacteria can facilitate communication between insects by producing aggregative pheromone (Dillon et al. 2002). In addition, as indicated by Adam (2010), microorganisms promote the adaptation of invasive insects in their new environment, but the environment can also influence the microbiome of insects.

Little is known about the EAB microbiome. Gut microbial communities of fourth-instar larvae, prepupae, and adults of EAB have been identified using 16S rRNA gene clone libraries

and culture-based methods (Vasanthakumar *et al.* 2008). To date, no study has applied highthroughput sequencing on EAB adult populations to discover its microbiome. The aim of this study was to identify and compare bacterial communities associated with the emerald ash borer from different locations and environments. We hypothesized that a combination of stochastic and niche selection pressures influences the taxonomic structure of the microbiome of the whole EAB, resulting in a biogeographical distribution pattern of microbial communities associated with the insect. To challenge this hypothesis, we sampled four EAB natural populations in the provinces of Ontario and Quebec, and one population reared under laboratory conditions and we studied the taxonomic structure of their respective microbiomes, using Illumina MiSeq sequencing of PCRamplified bacterial 16S rRNA genes. We report for the very first time the characterization of EAB whole body microbiome, while suggesting potential underlying abiotic and biotic drivers.

### 4.5 Material and Method

#### 4.5.1 Site Location

Natural populations of the EAB were collected from four infested sites in Canada. The first three locations were in the province of Quebec and the fourth one in Ontario (Figure 4.1). The first site was found at the Montreal Botanical Garden, QC (MTL; 45.560, -73.556), where several native and non-native ash species are present. The first detection of EAB in Montreal was done in 2011 (Natural Resources Canada 2016). The second site was an urban park situated in the municipality of Gatineau, QC (GAT; 45.504, -75.653) where EAB was detected for the first time in 2011 (Natural Resources Canada 2016), and at about 6.5km from Ottawa (ON) where the insect was first recorded in 2008 (Natural Resources Canada 2016). The third site was located at Saint-Basile-le-Grand, QC (SBLG; 45.506, -73.254) on ash trees positioned on the shore of the Richelieu River; this site was less than 5km from Carignan, the first village where EAB occurrence was reported in the province of Quebec in 2008 (CQEEE 2014). Finally, the fourth site was in Sault Ste. Marie, ON (SSM; 46.515, -84.290), where the insect was first reported in 2009 (Lyons 2010).

#### 4.5.2 Insect collection

In the summer of 2014, at each natural population sites, three Lindgren green 12-funnel traps were baited with a (Z)-3-Hexenol lure (Synergy Semiochemicals Corp., Burnaby (BC),

Canada) and 0.1mg of (Z)-3-Lactone synthesized as described by Silk *et al.* (2011) and loaded on rubber septa (Wheaton Scientific, Millville, NJ, USA). Traps were placed, using single rope techniques, in the canopy of randomly selected ash trees, corresponding to three replicates per site. By the end of June, EAB adults were collected twice weekly and individually conserved in 2.0 ml microcentrifuge tubes at 4°C.

Laboratory population corresponded to EAB newly emerged adults from ash log maintained in controlled conditions. Ash logs came from Middlesex County, ON (MC; 43.214, - 81.643) (Figure 4.1). They were cut from woodlot owned by the Ausable Bayfield Conservation Authority (Wright Thompson Tract) on fall of 2013 and stored for the winter in an outdoor container. The logs were brought indoors in early spring of 2014 and setup in cages in controlled conditions (24-28°C, L:D 16:8 photoperiod, and 20-40% RH), until adults emerged. Emerging insects from three different cages (replicates) were individually placed in 2.0 ml microcentrifuge tubes at 4°C until the sample preparation.

#### 4.5.3 Sample Preparation and DNA Extraction

Based on morphological characters, males and females were separated as described by Rodriguez-Saona *et al.* (2007). For each replicate, ten EAB (males and females) were randomly selected, and insects were individually crushed using a mortar pestles in 200ul of extraction buffer (50 mM Tris-HCl, 5 mM EDTA-2Na, 3% SDS, pH 8.0) containing 20  $\mu$ g/ml RNase A in a 2.0 ml microcentrifuge tube. For each gender, beetles were then pooled per replicate into 2.0 ml screw cap tube containing 200 mg of 0.1 mm glass beads (BioSpecs, Bartlesville, OK, USA).

Total DNA extraction was performed on each pool using mechanical lysis method as described by Durand *et al.* (2015). Briefly, two cycles of cell lysis at 4 m/s for 50 s were performed using FastPrep®-24 Instrument (MP Biomedicals, Solon, OH, USA) followed by 5 min on ice. The tubes were then centrifuged at 16 800 x g for 5 min and the supernatants were collected respectively into new tubes. A second cell lysis cycle was performed on the tissue pellets by adding 1 ml of extraction buffer containing 20  $\mu$ g/ml RNase A to the tubes, followed by two cycles of cell lysis as previously described. The supernatants of the two cell lysis cycles were combined and treated as follow: ammonium acetate was added to each sample at a final concentration of 2 M and briefly agitated by inversion before placed on ice for 5 min utes; the tubes were centrifuged at 20 800 x g for 15 min at 4°C, and transferred on ice for 5 min before a second centrifugation under the same conditions. Supernatants were then recovered into a new tube and an equal

volume of isopropyl alcohol (2-Propanol) was added to each tube for overnight DNA precipitation at 4°C; DNA was pelleted by centrifugation at 20 800 x g at 4°C for 30 min, and the supernatants were discarded; DNA pellets were washed twice with cold EtOH 70% followed by a centrifugation at 20 800 x g for 15 min at 4°C between each wash; the pellets were air-dried and dissolved in sterile water. DNA concentration was estimated using the Quant-iT<sup>™</sup> PicoGreen® dsDNA Assay Kit (Invitrogen, Life Technologies, Burlington, ON, Canada) following the manufacturer instructions.

#### 4.5.4 PCR Amplification

To confirm the presence of bacterial DNA in each sample, a PCR has been performed to amplify the 16S rRNA gene using universal primer 28F (5' GAG TTT GAT CNT GGC TCA G 3') and 519R (5' GTN TTA CNG CGG CKG CTG 3') as described by Durand *et al.* (2015). All 50 µl PCR reactions of contained 0.2 mM mM MgCl<sub>2</sub>, 10 µg BSA, 0.2 mM dNTPs, 0.4 µM of each primer, 2.5 U Taq DNA polymerase, and 5X ThermoPol® buffer (New England Biolabs, Whitby, ON, Canada). The initial denaturation step of 5 min at 94°C was followed by 30 amplification cycles (94°C for 30 s, 58°C for 30 s, 72°C for 30 s) and a final extension step at 72°C for 10 min. The presence of PCR amplification fragments was confirmed on a 1.5% agarose gel electrophoresis stained with ethidium bromide and visualized under UV light.

#### 4.5.5 Illumina MiSeq Paired-End Sequencing

The V6-V8 region of the 16S rRNA gene was sequenced using the primers B969F-CS1 5'-ACGCGHNRAACCTTACC-3' and BA1406R-CS2 5'-ACGGGCRGTGWGTRCAA-3' (Comeau *et al.* 2011). The sequencing technology used was Illumina MiSeq 250 bp paired-ends at the Quebec Genome Innovation Centre (McGill University, Montreal, QC, Canada). Raw data are available on NCBI under BioProject PRJNA317317.

#### 4.5.6 Sequence Processing Pipeline

Sequences were processed using the software USEARCH 64 bits following the UPARSE pipeline (Edgar 2013). Briefly, the paired-end sequences were merged with a minimum length overlap of 20 bp. Since no significant difference has been observed between male and female

bacterial populations (data not shown), their merged sequences were combined for each replicate. Quality control was then applied using a maximum expected error value of 1.0. All singleton sequences were removed before the clusterization of the remaining sequences in operational taxonomic unit (OTU) set at 97% identity. An extra chimera check was realized using both the UCHIME algorithm (Edgar *et al.* 2011) and the ChimeraSlayer 'Gold' reference database (Haas *et al.* 2011). A preliminary taxonomic assignment was realized with the Ribosomal Database Project (RDP) classifier V. 11 trained on 16S rRNA training set 14 (Wang *et al.* 2007) to remove OTU identified as archaea. The software Quantitative Insights Into Microbial Ecology (QIIME) V. 1.8.0 (Caporaso *et al.* 2010) was used to remove sequences represented by less than 0.005% of the total number of reads per library. All libraries were also equalized with QIIME for equal sampling depth. The final taxonomic assignment was done with RDP with a minimum of 50% confidence level.

#### 4.5.7 Diversity analysis

Rarefaction curves were generated using the software QIIME V. 1.8.0 (Caporaso et al. 2010) to assess the sufficiency of the sequencing effort; all samples were rarefied at 243 859 sequences to obtain equal sampling depth. All other diversity analysis were carried out with the software R V. 3.1.3 (R Core Development Team 2015). The abundance-based coverage estimator (ACE) was estimated using the package "fossil" (Vavrek 2011). Variance analysis, ANOVA, was realized with the software Jumping (JMP®, V. 10.0.0. SAS Institute Inc., Cary, NC, 1989-2007). Hellinger transformation of the OTU table was performed to standardize the data. An Euclidean distance matrix allowed to perform β-diversity analysis, average agglomerative clustering (UPGMA) and principal component analysis (PCA).  $\beta$ -diversity analysis was measured as the homogeneity of multivariate dispersions of the OTU using the "betadisper" function (999 permutations) implemented in the "vegan" package V. 2.3-3 (Oksanen et al. 2015). UPGMA agglomerative clustering was computed using "hclust" function in the "stats" package (R Core Development Team 2015) to verify similarity among sample site. To identify significant nodes, the similarity profile tool (SIMPROF) implemented in the "clustsig" package (Douglas and Mary 2014) was used to perform 999 permutations. PCA was computed using the "vegan" package V. 2.3-3 (Oksanen et al. 2015). Based on UPGMA cluster, the indicator species analysis procedure implemented in the "indicspecies" package (De Caceres and Legendre 2009) was used to identify indicators OTU (999 permutations). To illustrate the relative abundance of indicators OTU among samples, a heat map was constructed using "gplots" and "RColorBrewer" packages (Neuwirth 2014, Warnes *et al.* 2015). All statistical tests were performed with  $\alpha$  level of 0.05.

### 4.6 Results

#### 4.6.1 Bacterial diversity across Agrilus planipennis populations

Bacterial community of five *Agrilus planipennis* populations was characterized using highthroughput sequencing. A total of 8 924 489 sequences (434.6 base pairs (bp) +/- 17.5) were obtained succeeding the merge of paired-end raw sequences. After sequences processing, 243 859 (418.9  $\pm$  35.5 bp) high-quality sequences remained for each sample. Following clusterization at 97% pairwise-identity threshold, 94 OTUs were obtained. As shown by the rarefaction curves (Annexe1, Figure 1S), a sigmoidal curve was observed for all samples, suggesting that almost all bacterial diversity was covered in the 15 samples and that sampling depth was suitable for this study. Moreover, specific richness for each sample was estimated using the abundance-based coverage estimator (ACE) (Table 1) to compare the  $\alpha$ -diversity. No significant difference was observed between ACE estimator of the microbial communities associated with EAB populations (ANOVA test; F=1.5482; *p* = 0.2616), and, furthermore, there was no significant difference between ACE estimator and observed OTU values (*t*-Test; *p* < 0.05).

Table 4.1 : Specific richness associated with EAB populations. The abundance-base coverage estimator (ACE) and the observed OTU were used to compared  $\alpha$ -diversity between EAB populations. SSM = Sault-Ste. Marie; MC = Middlesex County; SBLG = Saint-Basile-le-Grand; GAT = Gatineau and MTL = Montreal; sd = standard deviation.

| Sample | ACE ± sd | Observed OTUs ± sd |
|--------|----------|--------------------|
| SSM    | 74 ± 7   | 66 ± 5             |
| МС     | 75 ± 4   | 73 ± 5             |
| SBLG   | 72 ± 3   | 70 ± 5             |
| GAT    | 64 ± 8   | 70 ± 12            |
| MTL    | 75 ± 8   | 73 ± 9             |

For each sampling site, heterogeneity among EAB associated bacterial communities was also analyzed using multivariate dispersion metrics (Figure 4.2). Based on these results, significant  $\beta$ -diversity variations were observed between EAB populations (ANOVA test; F=9.0082; *p* < 0.05). According to the distance to centroid, bacterial communities associated with natural populations of EAB (SSM, SBLG, GAT, MTL) were significantly more heterogeneous than the laboratory population (MC). Among the natural populations, a significant difference was also observed between MTL and GAT populations (pairwise comparison; *p* < 0.05). MC was more constrained since all replicates were closely related to the centroid. Bacterial communities of EAB populations sampled at SSM, SBLG, and GAT were more broadly dispersed. Some replicates of these populations were close to their centroid while others were more distant. On the other hand, MTL population appeared more stochastic, since all replicates were distant to the centroid.



Figure 4.2 : Multivariate dispersion measure of EAB population. Replicates are group by populations. The boxplots show calculated multivariate distance, where the lower and upper values are represented with the whiskers and the median is represented with the black line. Distance to centroid is correlated with the heterogeneity of sample among populations (999 permutations; p < 0.05). SSM = Sault-Ste. Marie; MC = Middlesex County; SBLG = Saint-Basile-le-Grand; GAT = Gatineau and MTL = Montreal.

A UPGMA agglomerative clustering was computed to classify EAB populations based on their bacterial community structure (Figure 4.3). Two significantly ( $\geq$  95%) distinct groups were obtained, defined as Ecotype 1 and Ecotype 2 clusters, respectively, SSM, MC, SBLG\_2, and MTL\_1, and GAT, MTL\_2, MTL\_3, SBLG\_1, and SBLG\_3 samples. Surprisingly, as it would have been expected in the case of the biogeographical distribution pattern of microbial communities, SBLG\_2 and MTL\_1 were not clustered with related EAB samples, indicating important variability between samples collected in these two locations.



Figure 4.3 : UPGMA agglomerative clustering of the EAB populations. Populations are separated in two significant clusters represented by a significant node ( $\geq$  95%). Ecotype 1 comprises SSM (Sault-Ste. Marie samples); MC (Middlesex County samples); SBLG\_2 (sample 2 of Saint-Basile-le-Grand); and MTL\_1 (sample 1 of Montreal). Ecotype 2 includes GAT (Gatineau samples); SBLG\_1 and SBLG\_3 (sample 1 and 3 of Saint-Basile-le-Grand); MTL\_2 and MTL\_3 (sample 2 and 3 of Montreal).

#### 4.6.2 Taxonomy of EAB bacterial community structures

Figure 4.4 shows the taxonomical identification at the genus level of all identified OTUs. Among them, 37 OTUs were shared between all 15 samples. These OTUs were represented by *Pseudomonas* (6 OTUs), *Serratia* (3 OTUs), *Enterococcus* (1 OTU), *Erwinia* (1 OTU), *Corynebacterium* (1 OTU), *Neptunomonas* (1 OTU), *Hafnia* (1 OTU), *Enterobacter* (1 OTU), *Tatumella* (1 OTU), *Sporosarcina* (1 OTU), *Pantoea* (1 OTU), *Kluyvera* (1 OTU). Additionally, 19 shared OTUs were unclassified.





Differences between the two ecotypes are explained by the relative abundance of specific bacterial genera (*t*-Test, p < 0.1) (Annexe 2, Figure S2). OTUs belonging to the *Pseudomonas* genus were significantly more abundant in Ecotype 1, with relative abundance ranging from 30% to 83% between samples. On the other hand, *Kluyvera* genus was more abundant in Ecotype 2 (10% to 55%) than Ecotype 1. *Stenotrophomonas* genus was only present in some samples of Ecotype 1 (1% to 6% of relative abundance), where *Tatumella* genus was associated with some

samples of Ecotype 2 (0% to 10% of relative abundance). To further compare the bacterial diversity among samples, a principal component analysis (PCA) was computed (Figure 4.5). The proportion of the variance accounted for the first two principal components was about 58.6% with, respectively, 41.5% and 17.1% for PC1 and PC2. Adding the "Ecotype" as a discrete variable, the result indicated that the PC1 component explained mostly the separation between samples belonging to Ecotype 1 and Ecotype 2. This separation was brought negatively by OTU 1 with an eigenvalue of -0.8709 and positively by OTU 54 having an eigenvalue of 0.5788 associated respectively with *Pseudomonas* and *Pantoea* genus. OTU 19 associated with *Kluyvera* genus was positively linked with PC1 and negatively with PC2 with eigenvalues of, respectively, 0.7402 and -0.8141. PC2 is also positively associated with OTU 36 (eigenvalue 0.7554) and OTU 40 (eigenvalue 0.5794) associated with *Serratia* and *Hafnia* genus, respectively.



Figure 4.5 : Principal component analysis (PCA) of EAB populations. PC1 and PC2 explained 41.5% and 17.1% of the variance respectively. OTUs out of the circle exerted more weight than the average to explain the position of each EAB population in the reduced space. OTUs 1, 19 and 54 defined the PC1, while OTUs 19, 36 and 40 defined PC2. These OTUs explain the separation of populations in Ecotype 1 (square) and Ecotype 2 (triangle) determined by cluster analysis.

In addition to species that are discriminant of the two ecotypes, indicator OTUs were identified with the aim to determine which OTU were restricted to each ecotype (Figure 4.6). Nine OTUs were identified as indicators of Ecotype 1: OTU 52 (*Erwinia* sp.) was found in all samples of Ecotype 1, except for SSM\_2; OTU 61 (*Stenotrophomonas* sp.) was present in all samples of this ecotype, except for SSM\_1; OTU 16 (*Brevundimonas* sp.) and OTU 70 (*Vagococcus* sp.) were identified, respectively, in six and five samples of Ecotype 1 and some indicator OTUs such as OTU 13 (*Pseudomonas* sp.), OTU 31 (*Plantibacter* sp.), OTU 46 (*Pseudomonas* sp.), OTU 55 (*Sanguibacter* sp.) and OTU 74 (*Pigmentiphaga* sp.) were only present in less than four samples. Six indicator OTUs were identified for Ecotype 2: OTU 11 (*Citrobacter* sp.) and OTU 38 (*Enterobacter* sp.), OTU 59 (*Propionibacterium* sp.), and OTU 88 (*Pelomonas* sp.) were only found in few samples of Ecotype 2; OTU 17 (*Bacillus* sp.) was present in high abundance only in MTL\_3 sample.





#### 4.7 Discussion

The aim of this study was to characterize the bacterial community associated with the EAB. Insects were collected from four natural and one laboratory reared populations. Thirty-seven OTUs belonging to 12 bacterial genera were shared in all of the 15 samples, which form the core microbiome associated with *A. planipennis* adults (Figure 4.4). This core microbiome consists of bacterial genera *Erwinia*, *Serratia*, *Tatumella*, *Hafnia*, *Enterobacter*, *Pantoea*, and *Kluyvera* belonging to Enterobacterioaceae family, and *Pseudomonas*, *Enterococcus*, *Corynebacterium*, *Neptunomonas*, and *Sporosarcina* from, respectively, Pseudomonadaeae, Enterococcaceae, Corynebacteriaceae, Oceanospirillaceae, and Planococcaceae family. All these genera are more likely to play a symbiotic role within the insect since they are present in all samples.

Erwinia sp. and Pantoea sp. have already been identified in EAB larvae gut while Enterococcus sp. and Corynebacterium sp. have been identified previously only in EAB prepupae gut (Vasanthakumar et al. 2008). Erwinia isolated from EAB has already demonstrated a cellulolytic activity (Vasanthakumar et al. 2008). Despite the fact that Pantoea sp., Enterococcus sp. and Corynebacterium sp. were known associates of EAB, their functions have not been identified so far. Functions of some bacteria forming the core microbiome have already been determined in others insect. For example, in the fungus garden of leaf-cutter ant, Pantoea participates in the biosynthesis of essential amino acids and vitamins (Aylward et al. 2012), and is involved in polymer degradation (Suen et al. 2010). Pantoea agglomerans and Enterobacter cloacae associated with fecal pellets of the desert locust [Schistocerca gregaria (Forsskål)], produce phenol compounds that are found in cohesion pheromone (Dillon et al. 2002). These phenol compounds are also involved in antifungal activities according to Dillon and Charnley (1995). Moreover, Serratia sp. and Pseudomonas sp. associated with the oral secretion of Dendroctonus rufipennis Kirby has shown the ability to inhibit antagonistic fungi (Cardoza et al. 2006). In Balanogastris kolae (Desbrocher des Loges), Corynebacterium sp. and Serratia sp. show proteolytic and amylolytic activities (Femi-Ola and Babalola 2012). Enterococcus sp. in Harpalus pennsylvanicus (Degeer) facilitates the digestion of seeds (Schmid et al. 2014). Despite the fact that Hafnia sp., Tatumella sp., Kluyvera sp. and Sporosarcinia sp. have already been found in insects microbiome, their functions have yet to be described (Gupta et al. 2014, Lundgren et al. 2007, Maleki-Ravasan et al. 2014, Muratoglu et al. 2011, Silk et al. 2011). The genus Neptunomonas comprises two species isolated from marine sediments, N. naphthovorans and N. japonica, which can oxidize polycyclic aromatic hydrocarbon (Hedlund and Costa 2010).

Considering the strength of the taxonomic assignment, this is the first time an association between *Neptunomonas* sp. and an insect is reported. In all cases, more studies are required to determine a more precise taxonomic level of identification of the bacteria and to investigate their function in association with EAB.

ACE estimator, which defines the  $\alpha$ -diversity or specific richness, is a good measure of sample coverage based on OTUs obtained from high-throughput sequencing (Gotelli and Colwell 2011). The results indicated that all populations have the same specific richness (no significant difference). This finding did not support other studies showing dissimilarities in specific richness between insect microbiome collected from natural and laboratory populations of the red palm weevil [Rhynchophorus ferrugineus (Olivier)] (Montagna et al. 2015), a mosquitoes species (Anopheles stephensi Liston) (Rani et al. 2009), and the cotton bollworm [Helicoverpa zea (Boddie)] (Xiang et al. 2006). Beside, Durand et al. (2015) have demonstrated higher specific richness in endomicrobiome than ectomicrobiome of natural populations of the eastern larch beetle [Dendroctonus simplex (LeConte)]. It is admitted that the microbiome associated to the cuticle of an insect should be more exposed to stochastic events than from the gut where nichebased processes shaped community assemblage (Douglas 2015), and may explain the difference observed between the endo and ectomicrobiome of the bark beetle. In the present work, data of the bacterial community structure of the EAB is not focusing only on the insect gut microbiome but considered the whole body, which might rationalize why no significant difference is observed among α-diversity values. Moreover, normal insect-rearing techniques involve thorough conditions such as specific artificial diet including antibiotics, disinfection and sanitization processes, controlled environmental conditions as well as reproductive selective pressures that may trigger more deterministic effects on the microbiome community structure than in natural populations (Blum et al. 2013, Dillon and Dillon 2004, Weiss and Aksoy 2011). However, as previously mentioned, the EAB laboratory insects were harvested from logs randomly selected from infested trees. The only controlled factors affecting these samples were those related to the temperature, relative humidity, and photoperiod. The pressure exerted by these abiotic factors was not shown to discriminate bacterial richness between EAB field and laboratory populations.

The  $\beta$ -diversity can be defined as the variability in OTUs composition observed among sample sites (Anderson *et al.* 2006). Multivariate dispersion analysis was used to evaluate heterogeneity among replicates of a population and higher distance to centroid suggests a higher heterogeneity within a sample site. The results showed that bacterial communities associated with EAB samples harvested from logs maintained in the laboratory are more homogeneous than those related to

samples of natural populations. As previously mentioned, natural populations are subject to many environment pressures that can shape the structure of insect microbiome. These results support the hypothesis that the immediate environment may have a significant impact on insect bacterial communities. Both  $\alpha$ - and  $\beta$ -diversity results suggest that EAB, which is a recent exotic insect pest, may carry a variety of bacteria. However, the steadiness in specific richness may also indicate that the number of species required to maintain bacterial function seems to be limited.

The EAB bacterial community structure is probably involved in stochastic and determinist processes. These pressures on the microbiome are induced by abiotic and biotic environmental features (Adams et al. 2010, Corby-Harris et al. 2007, Welch et al. 2015, Zouache et al. 2011). Host tree species (Medina et al. 2011), climate/temperature (Evans et al. 2011, Six 2013), elevation (Roe et al. 2011), distance between populations (Adams et al. 2010, Roe et al. 2011), sex (Rani et al. 2009) and life stage (Vasanthakumar et al. 2008) are among factors that have been demonstrated to affect the microbiome community. Moreover, the dissemination of microorganisms thriving in the phyllosphere or airborne particles should also be a factor to investigate. In this study, the absence of a strong biogeographical pattern was impelled by two samples from Quebec, which shared more similarity with Ontario samples. Results suggested that other factors might influence the taxonomic organization of EAB microbiome. One proposition is that the beetle population level of a specific tree may contribute to shape the microbial community structure. The detection of EAB in Quebec (MTL, 2011; SBLG, 2009; GAT, 2011) is more recent than in Ontario (SSM, 2008; MC, 2006), and the insect density at the tree level may affect the community structure of each sample (Figure 4.3). In fact, ash tree defense mechanism involve several changes in the phytochemistry such as high activities of peroxidase, trypsin inhibitor, and an increase of soluble phenolic concentrations (Cipollini et al. 2011), and might account as a factor affecting the shape of the bacterial community. For example, we found that Ecotype 1 is related to EAB samples mostly harvested in Ontario, where insect populations were established several years ago compared to Ecotype 2 which included samples collected in Quebec, where the establishment is still in progression. This progression process may explain why some samples from Quebec are clustered within Ecotype 1. More investigations should be achieved to evaluate the potential relationship that may exist between EAB establishment levels and the bacterial community structure.

Results of the principal component analysis show that some OTUs powered the distribution of the samples into two ecological Ecotypes (Figure 4.5). OTU\_1 (*Pseudomonas* sp.) and OTU\_54 (*Pantoea* sp.) drag some samples in Ecotype 1; while OTU\_19 (*Kluyvera* sp.) and also OTU\_54

43

(*Pantoea* sp.) pull samples of Ecotype 2 along the first component, suggesting that these OTUs are probably associated with the establishment level of EAB populations. Moreover, some OTUs are closely linked to each ecotype (Figure 4.6). OTU\_52 and OTU\_61 for Ecotype 1 and OTU\_11 and OTU\_38 for Ecotype 2 are the most abundant OTUs assigned as indicative OTU specie. Using these molecular markers, the insect establishment level of an EAB population can be easily assigned, which may provide noteworthy information for pest management authorities.

This work was the first to characterize the bacterial communities associated with different emerald ash borer populations. Bacteria present in all samples that we identified represent the core microbiome. Microbial variation among samples has been observed due to a combination of stochastic and determinist environmental factors shaping microbial communities. Insect establishment levels of EAB populations seem to be a major factor explaining the variations among bacterial communities. However, more experiments are necessary to support this hypothesis and to investigate the benefit or the cost linked to *A. planipennis* associated bacteria.

#### 4.8 Acknowledgements

We would like to thank Gene Jones (NRCan, Sault-Ste. Marie, ON), Sylvie Perron (Montreal Botanical Garden, Montreal, QC), Robert Werbiski (National Defense, Canada), Cloé Corbin-Bélanger (Saint-Basile-le-Grand municipality, QC), Alexandre Dumas (Gatineau municipality, QC) and John Enright (Upper Thames River Conservation Authority, London, ON) for assistance in insect collection. We would also thank Jean-Philippe Buffet, Audrey-Anne Durand, Laurence Guertin, Natalie Lavallée, Louis-Philippe Oliveira, Vincent Peck and Narin Srei for technical assistance. This project was supported by National Defense contract to CG and NRCAN funds to RL. AB was supported by NSERC Fellowship.

## CHAPITRE 5 : DISCUSSION ET CONCLUSION

L'objectif de ce projet de recherche était de caractériser la communauté bactérienne associée à l'agrile du frêne afin de comparer la diversité microbienne entre différentes populations ainsi qu'entre les différents sexes de ce ravageur. En considérant que seulement 10 % des bactéries sont cultivables avec les techniques de mise en culture actuelles (Stewart, 2012), une méthode de séquençage à haut débit, soit la technologie Illumina Miseq, a été choisie pour faire cette caractérisation. Cette méthode permet d'identifier une plus grande, voir presque la totalité, de la diversité bactérienne des échantillons. De plus, les résultats sont obtenus plus rapidement qu'avec les méthodes traditionnelles. De nombreux logiciels et procédures existent pour faciliter le traitement des données, tels que Mothur (Schloss *et al.*, 2009), QIIME (Caporaso *et al.*, 2010) et UPARSE (Edgar, 2013). La sélection de la procédure UPARSE (Annexe 2) a permis d'analyser les résultats rapidement grâce à un algorithme qui effectue la détection des chimères en même temps que le regroupement des séquences sous forme d'unité taxonomique opérationnelle (OTU) (Edgar, 2013). Les analyses de diversité bactérienne permettant la comparaison entre les échantillons ont été effectuées avec le logiciel R. Les scripts utilisés pour effectuer ces analyses sont disponibles en annexe 3 et 4.

Lors de l'élaboration du plan expérimental, cinq sites d'échantillonnage en milieu naturel provenant de différentes villes avaient été sélectionnés, soit London, Sault-Ste-Marie, Gatineau, Saint-Basile-le-Grand et Montréal. Par contre, arrivée au site d'échantillonnage de London, la plantation de frêne dans laquelle les agriles devaient être récupérés avait été complètement rasée quelques jours plus tôt (Figure 5.1). Face à cette mauvaise nouvelle, des collaborateurs ont proposé de nous fournir des échantillons en provenance du comté de Middlesex (Ontario). Les échantillons du site de London ont donc été remplacés par ceux-ci. À la suite des analyses de dispersion multivariée (Figure 4.2), l'hypothèse que les échantillons en provenance du compté de Middlesex soient issus d'une population d'élevage en laboratoire a été émise. Les résultats démontrent que les réplicas provenant de populations en milieu naturel sont hétérogènes alors que ceux du comté de Middlesex sont très homogènes. Effectivement, après avoir contacté les collaborateurs, ceux-ci nous ont confirmé que les agriles du frêne en provenance de Middlesex avaient émergé de bûches conservées en conditions contrôlées. Ce résultat nous démontre la puissance de cet outil de mesure de  $\beta$ -diversité. De surcrôt, ce résultat a permis de confirmer

qu'étudier les microorganismes associés aux insectes provenant d'élevage de laboratoire n'est pas toujours représentatif des populations naturelles.



Figure 5.1 : Plantation de frênes de London rasée quelques jours avant notre arrivée pour la récolte des agriles du frêne (Photographie : A, Bergeron, INRS).

Comme mentionné dans le chapitre 2, aucune différence significative au niveau de la communauté bactérienne n'a été observée entre les mâles et les femelles des différentes populations (Annexe 4). Les séquences des échantillons mâles et femelles ont alors été regroupées par site d'échantillonnage au début du traitement des séquences. En revanche, une variation dans la composition et la diversité bactérienne associée aux différentes populations d'agriles du frêne a permis d'établir un lien entre celles-ci. En effet, les analyses de diversité ont permis de séparer les populations d'agriles en deux groupes selon leur structure bactérienne. L'hypothèse proposée pour expliquer cette séparation est la densité de la population de l'insecte au niveau de l'arbre. Les échantillons regroupés dans l'Écotype 1 sont majoritairement des échantillons en provenance des sites d'échantillonnage de l'Ontario, où la détection de l'agrile du frêne date de plusieurs années, alors que les échantillons de l'Écotype 2 proviennent des sites d'échantillonnage du Québec, où l'établissement des populations est en progression. Dans ce dernier cas, on peut s'attendre à des densités de populations plus faibles que dans les populations

plus anciennes. Cette progression expliquerait le regroupement de certains réplicas en provenance de populations québécoises avec les populations ontariennes.

Les analyses de diversité ont également permis d'identifier les espèces bactériennes qui influencent la séparation de ces populations (Figure 4.5). De plus, des espèces bactériennes indicatrices ont été identifiées pour chacun de ces groupes (Figure 4.6). La confirmation de ces espèces indicatrices pourrait permettre le développement de marqueur moléculaire permettant d'assigner un niveau d'établissement aux populations d'agriles du frêne. Ces informations pourraient être l'objet d'un outil d'aide à la prise de décision pour les autorités responsables de la gestion des insectes ravageurs. En effet, connaître le niveau de la population d'agriles pour un site précis permettrait aux gestionnaires de choisir le meilleur moyen de lutte possible. Par exemple, en cas de grande infestation, on pourrait décider d'abattre les arbres afin de diminuer autant que possible la propagation de l'agrile. Alors que dans le cas de faibles infestations, on pourrait décider de traiter les arbres pour essayer de les sauver.

En conclusion, une caractérisation de la communauté bactérienne associée à l'agrile du frêne a été réalisée. Certains genres bactériens étaient présents dans tous les échantillons, formant ainsi la structure de base du microbiome de l'agrile du frêne. Tandis que d'autres bactéries étaient spécifiques à certains échantillons. Ces divergences, guidées par des processus stochastiques et déterministes, ont permis d'établir des liens entre les populations sur la base du niveau d'établissement de celles-ci. Davantage d'études sont nécessaires afin d'établir le rôle de ces bactéries lorsqu'elles vivent en association avec l'agrile du frêne. Ce projet de recherche a permis de mettre de l'avant l'importance de considérer les facteurs stochastiques et déterministes pouvant influencer la structure microbienne lors de l'étude de microbiome d'insecte.

## **RÉFÉRENCES BIBLIOGRAPHIQUES**

- Adams AS, Adams SM, Currie CR, Gillette NE & Raffa KF (2010) Geographic variation in bacterial communities associated with the red turpentine beetle (Coleoptera: Curculionidae). *Environmental Entomology* 39(2):406-414.
- Adams AS, Aylward FO, Adams SM, Erbilgin N, Aukema BH, Currie CR, Suen G & Raffa KF (2013) Mountain pine beetles colonizing historical and naive host trees are associated with a bacterial community highly enriched in genes contributing to terpene metabolism. *Applied and Environmental Microbiology* 79(11):3468-3475.
- Adams AS, Boone CK, Bohlmann J & Raffa KF (2011) Responses of bark beetle-associated bacteria to host monoterpenes and their relationship to insect life histories. *Journal of Chemical Ecology* 37(8):808-817.
- Anand AAP, Vennison SJ, Sankar SG, Prabhu DIG, Vasan PT, Raghuraman T, Geoffrey CJ & Vendan SE (2010) Isolation and characterization of bacteria from the gut of *Bombyx mori* that degrade cellulose, xylan, pectin and starch and their impact on digestion. *Journal of Insect Science* 10(1):107.
- Anderson MJ, Ellingsen KE & McArdle BH (2006) Multivariate dispersion as a measure of beta diversity. *Ecology Letters* 9(6):683-693.
- Anonymous (2016) *Emerald ash borer information network.* USDA Forest Service and Michigan State University, <u>www.emeraldashborer.info</u> (Consulté le 22-02-2016)
- Aukema BH, Werner RA, Haberkern KE, Illman BL, Clayton MK & Raffa KF (2005) Quantifying sources of variation in the frequency of fungi associated with spruce beetles: Implications for hypothesis testing and sampling methodology in bark beetle– symbiont relationships. *Forest Ecology and Management* 217(2):187-202.
- Aylward FO, Burnum KE, Scott JJ, Suen G, Tringe SG, Adams SM, Barry KW, Nicora CD, Piehowski PD & Purvine SO (2012) Metagenomic and metaproteomic insights into bacterial communities in leaf-cutter ant fungus gardens. *The ISME Journal* 6(9):1688-1701.
- Bansal R, Mian M & Michel AP (2014) Microbiome diversity of *Aphis glycines* with extensive superinfection in native and invasive populations. *Environmental Microbiology Reports* 6(1):57-69.
- Blum JE, Fischer CN, Miles J & Handelsman J (2013) Frequent replenishment sustains the beneficial microbiome of *Drosophila melanogaster*. *MBio* 4(6):e00860-00813.
- Broderick NA, Raffa KF, Goodman RM & Handelsman J (2004) Census of the bacterial community of the gypsy moth larval midgut by using culturing and culture-independent methods. *Applied and Environmental Microbiology* 70(1):293-300.
- Broderick NA, Raffa KF & Handelsman J (2006) Midgut bacteria required for *Bacillus thuringiensis* insecticidal activity. *Proceedings of the National Academy of Sciences* 103(41):15196-15199.

- Brownlie JC & Johnson KN (2009) Symbiont-mediated protection in insect hosts. *Trends in Microbiology* 17(8):348-354.
- Brummel T, Ching A, Seroude L, Simon AF & Benzer S (2004) *Drosophila* lifespan enhancement by exogenous bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 101(35):12974-12979.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK & Gordon JI (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7(5):335-336.
- Cappaert D, McCullough DG, Poland TM & Siegert NW (2005) Emerald ash borer in North America: a research and regulatory challenge. *American Entomologist* 51(3):152-165.
- Cardoza YJ, Klepzig KD & Raffa KF (2006) Bacteria in oral secretions of an endophytic insect inhibit antagonistic fungi. *Ecological Entomology* 31(6):636-645.
- Cipollini D, Wang Q, Whitehill JG, Powell JR, Bonello P & Herms DA (2011) Distinguishing defensive characteristics in the phloem of ash species resistant and susceptible to emerald ash borer. *Journal of Chemical Ecology* 37(5):450-459.
- Comeau AM, Li WK, Tremblay J-É, Carmack EC & Lovejoy C (2011) Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. *PLoS One* 6(11):e27492.
- Corby-Harris V, Pontaroli AC, Shimkets LJ, Bennetzen JL, Habel KE & Promislow DE (2007) Geographical distribution and diversity of bacteria associated with natural populations of *Drosophila melanogaster*. *Applied and Environmental Microbiology* 73(11):3470-3479.
- CQEEE (2014) *Agrile du frêne.* Conseil Québecois des Espèces Exotiques Envahissantes, Mont-Saint-Hilaire, QC, <u>http://agrile.cqeee.org/?page\_id=1467</u> (Consulté le 25-01-2016)
- Crook DJ, Khrimian A, Cossé A, Fraser I & Mastro VC (2012) Influence of trap color and host volatiles on capture of the emerald ash borer (Coleoptera: Buprestidae). *Journal of Economic Entomology* 105(2):429-437.
- Currie C, Bot A & Boomsma JJ (2003) Experimental evidence of a tripartite mutualism: bacteria protect ant fungus gardens from specialized parasites. *Oikos* 101(1):91-102.
- De Caceres M & Legendre P (2009) Associations between species and groups of sites: indices and statistical inference. *Ecology* 90(12):3566-3574.
- Degnan PH, Yu Y, Sisneros N, Wing RA & Moran NA (2009) *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proceedings of the National Academy of Sciences* 106(22):9063-9068.
- Delalibera I, Handelsman J & Raffa KF (2005) Contrasts in cellulolytic activities of gut microorganisms between the wood borer, *Saperda vestita* (Coleoptera: Cerambycidae), and the bark beetles, *Ips pini* and *Dendroctonus frontalis* (Coleoptera: Curculionidae). *Environmental Entomology* 34(3):541-547.

- Delalibera I, Vasanthakumar A, Burwitz BJ, Schloss PD, Klepzig KD, Handelsman J & Raffa KF (2007) Composition of the bacterial community in the gut of the pine engraver, *Ips pini* (Say)(Coleoptera) colonizing red pine. *Symbiosis* 43(2):97.
- Dillon R & Charnley A (1988) Inhibition of *Metarhizium anisopliae* by the gut bacterial flora of the desert locust: characterisation of antifungal toxins. *Canadian Journal of Microbiology* 34(9):1075-1082.
- Dillon R & Charnley A (1995) Chemical barriers to gut infection in the desert locust: in vivo production of antimicrobial phenols associated with the bacterium *Pantoea agglomerans*. *Journal of Invertebrate Pathology* 66(1):72-75.
- Dillon R & Dillon V (2004) The gut bacteria of insects: nonpathogenic interactions. *Annual Review of Entomology* 49(1):71-92.
- Dillon R, Vennard C, Buckling A & Charnley A (2005) Diversity of locust gut bacteria protects against pathogen invasion. *Ecology Letters* 8(12):1291-1298.
- Dillon R, Vennard C & Charnley A (2002) A note: gut bacteria produce components of a locust cohesion pheromone. *Journal of Applied Microbiology* 92(4):759-763.
- Douglas (2009) The microbial dimension in insect nutritional ecology. *Functional Ecology* 23(1):38-47.
- Douglas (2015) Multiorganismal insects: diversity and function of resident microorganisms. *Annual Review of Entomology* 60:17-34.
- Douglas W & Mary C (2014). clustsig: Significant Cluster Analysis. R package version 1.1.
- Durand A-A, Bergeron A, Constant P, Buffet J-P, Deziel E & Guertin C (2015) Surveying the endomicrobiome and ectomicrobiome of bark beetles: The case of *Dendroctonus simplex*. *Scientific Reports* 5:17190.
- Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature methods* 10(10):996-998.
- Edgar RC, Haas BJ, Clemente JC, Quince C & Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics (Oxford)* 27(16):2194-2200.
- Engel P, Martinson VG & Moran NA (2012) Functional diversity within the simple gut microbiota of the honey bee. *Proceedings of the National Academy of Sciences* 109(27):11002-11007.
- Engel P & Moran NA (2013) The gut microbiota of insects–diversity in structure and function. *FEMS Microbiology Reviews* 37(5):699-735.
- Evans L, Hofstetter R, Ayres M & Klepzig K (2011) Temperature alters the relative abundance and population growth rates of species within the *Dendroctonus frontalis* (Coleoptera: Curculionidae) community. *Environmental Entomology* 40(4):824-834.
- Femi-Ola TO & Babalola AG (2012) Microbiology of the gut of the kola nut weevil, *balanogastris kolae. Journal of Insect Science* 12(1).
- Frago E, Dicke M & Godfray HCJ (2012) Insect symbionts as hidden players in insect-plant interactions. *Trends in Ecology and Evolution* 27(12):705-711.

- Franceschi VR, Krokene P, Christiansen E & Krekling T (2005) Anatomical and chemical defenses of conifer bark against bark beetles and other pests. *New Phytologist* 167(2):353-376.
- Francese JA, Fraser I, Lance DR & Mastro VC (2011) Efficacy of multifunnel traps for capturing emerald ash borer (Coleoptera: Buprestidae): effect of color, glue, and other trap coatings. *Journal of economic entomology* 104(3):901-908.
- Gandhi KJ & Herms DA (2010a) Direct and indirect effects of alien insect herbivores on ecological processes and interactions in forests of eastern North America. *Biological Invasions* 12(2):389-405.
- Gandhi KJ & Herms DA (2010b) North American arthropods at risk due to widespread *Fraxinus* mortality caused by the alien emerald ash borer. *Biological Invasions* 12(6):1839-1846.
- Gibson CM & Hunter MS (2010) Extraordinarily widespread and fantastically complex: comparative biology of endosymbiotic bacterial and fungal mutualists of insects. *Ecology letters* 13(2):223-234.
- Giordano L, Garbelotto M, Nicolotti G & Gonthier P (2013) Characterization of fungal communities associated with the bark beetle *Ips typographus* varies depending on detection method, location, and beetle population levels. *Mycological Progress* 12(1):127-140.
- Gotelli NJ & Colwell RK (2011) Estimating species richness. *Biological diversity: frontiers in measurement and assessment,* Vol 12. p 39-54.
- Gupta AK, Rastogi G, Nayduch D, Sawant SS, Bhonde RR & Shouche YS (2014) Molecular phylogenetic profiling of gut-associated bacteria in larvae and adults of flesh flies. *Medical and Veterinary Entomology* 28(4):345-354.
- Haack RA, Jendak E, Houping L, Marchant KR, Petrice TR, Poland TM & Ye H (2002) The emerald ash borer: a new exotic pest in North America. *Newsletter of the Michigan Entomological Society* 47(3-4):1-5.
- Haeder S, Wirth R, Herz H & Spiteller D (2009) Candicidin-producing *Streptomyces* support leaf-cutting ants to protect their fungus garden against the pathogenic fungus *Escovopsis. Proceedings of the National Academy of Sciences* 106(12):4742-4746.
- Hedges LM, Brownlie JC, O'Neill SL & Johnson KN (2008) *Wolbachia* and virus protection in insects. *Science* 322(5902):702-702.
- Hedlund BP & Costa K (2010) *Neptunomonas. Handbook of Hydrocarbon and Lipid Microbiology*, Springer. p 1773-1779.
- Herms DA & McCullough DG (2014) Emerald ash borer invasion of north america: History, biology, ecology, impacts, and management. *Annual Review of Entomology* 59:13-30.
- Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A & Werren JH (2008) How many species are infected with *Wolbachia*?–a statistical analysis of current data. *FEMS Microbiology Letters* 281(2):215-220.

- Huang S & Zhang H (2013) The impact of environmental heterogeneity and life stage on the hindgut microbiota of *Holotrichia parallela* larvae (Coleoptera: Scarabaeidae). *PloS One* 8(2):e57169.
- Jaenike J, Unckless R, Cockburn SN, Boelio LM & Perlman SJ (2010) Adaptation via symbiosis: recent spread of a *Drosophila* defensive symbiont. *Science* 329(5988):212-215.
- Janson EM, Stireman JO, Singer MS & Abbot P (2008) Phytophagous insect–microbe mutualisms and adaptive evolutionary diversification. *Evolution* 62(5):997-1012.
- Jia S, Zhang X, Zhang G, Yin A, Zhang S, Li F, Wang L, Zhao D, Yun Q & Wang J (2013) Seasonally variable intestinal metagenomes of the red palm weevil (*Rhynchophorus ferrugineus*). *Environmental Microbiology* 15(11):3020-3029.
- Kaltenpoth M, Göttler W, Herzner G & Strohm E (2005) Symbiotic bacteria protect wasp larvae from fungal infestation. *Current Biology* 15(5):475-479.
- Kaltenpoth M, Schmitt T, Polidori C, Koedam D & Strohm E (2010) Symbiotic *streptomycetes* in antennal glands of the South American digger wasp genus *Trachypus* (Hymenoptera, Crabronidae). *Physiological Entomology* 35(2):196-200.
- Kikuchi Y, Hayatsu M, Hosokawa T, Nagayama A, Tago K & Fukatsu T (2012) Symbiontmediated insecticide resistance. *Proceedings of the National Academy of Sciences* 109(22):8618-8622.
- Leroy PD, Sabri A, Heuskin S, Thonart P, Lognay G, Verheggen FJ, Francis F, Brostaux Y, Felton GW & Haubruge E (2011) Microorganisms from aphid honeydew attract and enhance the efficacy of natural enemies. *Nature communications* 2:348.
- Lundgren JG, Lehman RM & Chee-Sanford J (2007) Bacterial communities within digestive tracts of ground beetles (Coleoptera: Carabidae). *Annals of the Entomological Society of America* 100(2):275-282.
- Lyons DB (2010) Emerald ash borer. in *Frontline Technical Note 110.* (Natural Resources Canada, Great Lakes Forestry Centre, Sault Ste. Marie, Ontario), p 4.
- Lyons DB, Caister C, de Groot P, Hamilton B, Marchant K, Scarr T & Turgeon J (2007) *Guide pour les enquêtes de dépistage de l'agrile du frêne.* Ressources Naturelles Canada, Sault-Ste Marie, Ontario. 58 p
- MacFarlane DW & Meyer SP (2005) Characteristics and distribution of potential ash tree hosts for emerald ash borer. *Forest Ecology and Management* 213(1):15-24.
- Maleki-Ravasan N, Oshaghi MA, Hajikhani S, Saeidi Z, Akhavan AA, Gerami-Shoar M, Shirazi MH, Yakhchali B, Rassi Y & Afshar D (2014) Aerobic microbial community of insectary population of *Phlebotomus papatasi*. *Journal of Arthropod-Borne Diseases* 8(1):69.
- Marshall J, Storer A, Fraser I & Mastro V (2011) A predictive model for detection of *Agrilus planipennis* (Col., Buprestidae) larvae in girdled ash (*Fraxinus* spp.). *Journal of Applied Entomology* 135(1-2):91-97.
- McKenney DW, Pedlar JH, Yemshanov D, Barry Lyons D, Campbell KL & Lawrence K (2012) Estimates of the potential cost of emerald ash borer (*Agrilus planipennis* Fairmaire) in Canadian municipalities. *Arboriculture and Urban Forestry* 38(3):81.

- Medina R, Nachappa P & Tamborindeguy C (2011) Differences in bacterial diversity of hostassociated populations of *Phylloxera notabilis* Pergande (Hemiptera: Phylloxeridae) in pecan and water hickory. *Journal of Evolutionary Biology* 24(4):761-771.
- Montagna M, Chouaia B, Mazza G, Prosdocimi EM, Crotti E, Mereghetti V, Vacchini V, Giorgi A, De Biase A & Longo S (2015a) Effects of the diet on the microbiota of the red palm weevil (Coleoptera: Dryophthoridae). *PloS One* 10(1):e0117439.
- Montagna M, Gómez-Zurita J, Giorgi A, Epis S, Lozzia G & Bandi C (2015b) Metamicrobiomics in herbivore beetles of the genus *Cryptocephalus* (Chrysomelidae): toward the understanding of ecological determinants in insect symbiosis. *Insect Science* 22(3):340-352.
- Morales-Jiménez J, de León AV-P, García-Domínguez A, Martínez-Romero E, Zúñiga G & Hernández-Rodríguez C (2013) Nitrogen-fixing and uricolytic bacteria associated with the gut of *Dendroctonus rhizophagus* and *Dendroctonus valens* (Curculionidae: Scolytinae). *Microbial Ecology* 66(1):200-210.
- Morales-Jiménez J, Zúñiga G, Ramírez-Saad HC & Hernández-Rodríguez C (2012) Gutassociated bacteria throughout the life cycle of the bark beetle *Dendroctonus rhizophagus* Thomas and Bright (Curculionidae: Scolytinae) and their cellulolytic activities. *Microbial Ecology* 64(1):268-278.
- Morales-Jiménez J, Zúñiga G, Villa-Tanaca L & Hernández-Rodríguez C (2009) Bacterial community and nitrogen fixation in the red turpentine beetle, *Dendroctonus valens* LeConte (Coleoptera: Curculionidae: Scolytinae). *Microbial Ecology* 58(4):879-891.
- Moran NA, McCutcheon JP & Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics* 42:165-190.
- Mueller UG (2012) Symbiont recruitment versus ant-symbiont co-evolution in the attine antmicrobe symbiosis. *Current Opinion in Microbiology* 15(3):269-277.
- Muratoglu H, Sezen K & Demirbag Z (2011) Determination and pathogenicity of the bacterial flora associated with the spruce bark beetle, *Ips typographus* (L.)(Coleoptera: Curculionidae: Scolytinae). *Turkish Journal of Biology* 35(1):9-20.
- Musolin DL (2007) Insects in a warmer world: ecological, physiological and life-history responses of true bugs (Heteroptera) to climate change. *Global Change Biology* 13(8):1565-1585.
- Natural Resources Canada (2016) *Emerald ash borer*. <u>http://www.nrcan.gc.ca/forests/fire-insects-disturbances/top-insects/13377</u> (Consulté le 24-01-2016)
- Neuwirth E (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2.
- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH & Wagner H (2015). vegan: Community Ecology Package. R package version 2.3-1.
- Oliver KM, Campos J, Moran NA & Hunter MS (2008) Population dynamics of defensive symbionts in aphids. *Proceedings of the Royal Society of London B: Biological Sciences* 275(1632):293-299.

- Oliver KM, Russell JA, Moran NA & Hunter MS (2003) Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the National Academy of Sciences* 100(4):1803-1807.
- Popa V, Déziel E, Lavallée R, Bauce E & Guertin C (2012) The complex symbiotic relationships of bark beetles with microorganisms: a potential practical approach for biological control in forestry. *Pest Management Science* 68(7):963-975.
- Prado SS, Hung KY, Daugherty MP & Almeida RP (2010) Indirect effects of temperature on stink bug fitness, via maintenance of gut-associated symbionts. *Applied and Environmental Microbiology* 76(4):1261-1266.
- Priya NG, Ojha A, Kajla MK, Raj A & Rajagopal R (2012) Host plant induced variation in gut bacteria of *Helicoverpa armigera*. *PLoS One* 7(1):e30768.
- R Core Development Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rajagopal R (2009) Beneficial interactions between insects and gut bacteria. *Indian Journal of Microbiology* 49(2):114-119.
- Rani A, Sharma A, Rajagopal R, Adak T & Bhatnagar RK (2009a) Bacterial diversity analysis of larvae and adult midgut microflora using culture-dependent and culture-independent methods in lab-reared and field-collected *Anopheles stephensi*-an Asian malarial vector. *BMC Microbiology* 9(1):1.
- Rani A, Sharma A, Rajagopal R, Adak T & Bhatnagar RK (2009b) Bacterial diversity analysis of larvae and adult midgut microflora using culture-dependent and culture-independent methods in lab-reared and field-collected *Anopheles stephensi* an Asian malarial vector. *BMC Microbiology* 9(1):1.
- Rodriguez-Saona CR, Miller JR, Poland TM, Kuhn TM, Otis GW, Turk T & Ward DL (2007) Behaviors of adult *Agrilus planipennis* (Coleoptera: Buprestidae). *Great Lakes Entomologist* 40(1-2):1.
- Roe AD, James PM, Rice AV, Cooke JE & Sperling FA (2011) Spatial community structure of mountain pine beetle fungal symbionts across a latitudinal gradient. *Microbial Ecology* 62(2):347-360.
- Ryall KL, Fidgen JG & Turgeon JJ (2011) Detectability of the emerald ash borer (Coleoptera: Buprestidae) in asymptomatic urban trees by using branch samples. *Environmental Entomology* 40(3):679-688.
- Sabree ZL, Kambhampati S & Moran NA (2009) Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. *Proceedings of the National Academy of Sciences* 106(46):19521-19526.
- Scarborough CL, Ferrari J & Godfray H (2005) Aphid protected from pathogen by endosymbiont. *Science* 310(5755):1781-1781.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH & Robinson CJ (2009) Introducing mothur: open-source, platform-

independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* 75(23):7537-7541.

- Schmid RB, Lehman RM, Brözel VS & Lundgren JG (2014) An indigenous gut bacterium, *Enterococcus faecalis* (Lactobacillales: Enterococcaceae), increases seed consumption by *Harpalus pensylvanicus* (Coleoptera: Carabidae). *Florida Entomologist* 97(2):575-584.
- Scott JJ, Oh D-C, Yuceer MC, Klepzig KD, Clardy J & Currie CR (2008) Bacterial protection of beetle-fungus mutualism. *Science* 322(5898):63-63.
- Seipke RF, Barke J, Heavens D, Yu DW & Hutchings MI (2013) Analysis of the bacterial communities associated with two ant–plant symbioses. *Microbiology Open* 2(2):276-283.
- Sharon G, Segal D, Ringo JM, Hefetz A, Zilber-Rosenberg I & Rosenberg E (2010) Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences* 107(46):20051-20056.
- Shi W, Guo Y, Xu C, Tan S, Miao J, Feng Y, Zhao H, Leger RJS & Fang W (2014) Unveiling the mechanism by which microsporidian parasites prevent locust swarm behavior. *Proceedings of the National Academy of Sciences* 111(4):1343-1348.
- Shigenobu S, Watanabe H, Hattori M, Sakaki Y & Ishikawa H (2000) Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* 407(6800):81-86.
- Shin SC, Kim S-H, You H, Kim B, Kim AC, Lee K-A, Yoon J-H, Ryu J-H & Lee W-J (2011) *Drosophila* microbiome modulates host developmental and metabolic homeostasis via insulin signaling. *Science* 334(6056):670-674.
- Siegert NW, McCullough DG, Liebhold AM & Telewski FW (2014) Dendrochronological reconstruction of the epicentre and early spread of emerald ash borer in North America. *Diversity and Distributions* 20(7):847-858.
- Silk PJ, Ryall K, Mayo P, Lemay MA, Grant G, Crook D, Cossé A, Fraser I, Sweeney JD & Lyons DB (2011) Evidence for a volatile pheromone in *Agrilus planipennis* Fairmaire (Coleoptera: Buprestidae) that increases attraction to a host foliar volatile. *Environmental Entomology* 40(4):904-916.
- Six DL (2013) The bark beetle holobiont: why microbes matter. *Journal of Chemical Ecology* 39(7):989-1002.
- Stewart EJ (2012) Growing unculturable bacteria. *Journal of Bacteriology* 194(16):4151-4160.
- Storelli G, Defaye A, Erkosar B, Hols P, Royet J & Leulier F (2011) *Lactobacillus plantarum* promotes *Drosophila* systemic growth by modulating hormonal signals through TOR-dependent nutrient sensing. *Cell Metabolism* 14(3):403-414.
- Suen G, Scott JJ, Aylward FO, Adams SM, Tringe SG, Pinto-Tomás AA, Foster CE, Pauly M, Weimer PJ & Barry KW (2010) An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genetics* 6(9):e1001129.

- Teixeira L, Ferreira Á & Ashburner M (2008) The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLoS Biology* 6(12):e1000002.
- Tsuchida T, Koga R, Shibao H, Matsumoto T & Fukatsu T (2002) Diversity and geographic distribution of secondary endosymbiotic bacteria in natural populations of the pea aphid, *Acyrthosiphon pisum*. *Molecular Ecology* 11(10):2123-2135.
- Vasanthakumar A, Handelsman J, Schloss PD, Bauer LS & Raffa KF (2008) Gut microbiota of an invasive subcortical beetle, *Agrilus planipennis* Fairmaire, across various life stages. *Environmental Entomology* 37(5):1344-1353.
- Vavrek MJ (2011) fossil: palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica* 14(1).
- Wang Q, Garrity GM, Tiedje JM & Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* 73(16):5261-5267.
- Wang X-Y, Yang Z-Q, Gould JR, Zhang Y-N, Liu G-J & Liu E (2010) The biology and ecology of the emerald ash borer, *Agrilus planipennis*, in China. *Journal of Insect Science* 10.
- Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M & Venables B (2015). gplots: Various R Programming Tools for Plotting Data. R package version 2.17.0.
- Weiss B & Aksoy S (2011) Microbiome influences on insect host vector competence. *Trends in Parasitology* 27(11):514-522.
- Welch E, Macias J & Bextine B (2015) Geographic patterns in the bacterial microbiome of the glassy-winged sharpshooter, *Homalodisca vitripennis* (Hemiptera: Cicadellidae). *Symbiosis* 66(1):1-12.
- Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, Khouri H, Tallon LJ, Zaborsky JM, Dunbar HE & Tran PL (2006) Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biology* 4(6):e188.
- Xiang H, Wei G-F, Jia S, Huang J, Miao X-X, Zhou Z, Zhao L-P & Huang Y-P (2006) Microbial communities in the larval midgut of laboratory and field populations of cotton bollworm (*Helicoverpa armigera*). *Canadian Journal of Microbiology* 52(11):1085-1092.
- Zilber-Rosenberg I & Rosenberg E (2008) Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiology Reviews* 32(5):723-735.
- Zouache K, Raharimalala FN, Raquin V, Tran-Van V, Raveloson LHR, Ravelonandro P & Mavingui P (2011) Bacterial diversity of field-caught mosquitoes, *Aedes albopictus* and *Aedes aegypti,* from different geographic regions of Madagascar. *FEMS Microbiology Ecology* 75(3):377-389.
## **ANNEXE I : MATÉRIEL SUPPLÉMENTAIRE (ARTICLE)**

### BACTERIAL COMMUNITY DIVERSITY ASSOCIATED WITH DIFFERENT POPULATIONS OF AGRILUS PLANIPENNIS FAIRMAIRE (COLEOPTERA : BUPRESTIDEA)

### DIVERSITÉ DE LA COMMUNAUTÉ BACTÉRIENNE ASSOCIÉE À DIFFÉRENTES POPULATIONS D'AGRILUS PLANIPENNIS FAIRMAIRE (COLEOPTERA : BUPRESTIDEA)

#### Auteurs

Amélie Bergeron<sup>1</sup>, Robert Lavallée<sup>2</sup>, Philippe Constant<sup>1</sup>, Claude Guertin<sup>1</sup>

#### Affiliation

<sup>1</sup> INRS- Institut Armand-Frappier, Laval, QC, Canada, H7V 1B7

<sup>2</sup> Natural Resources Canada, Canadian Forest Service, Quebec, QC, Canada, G1V 4C7



Figure 1S : Rarefaction curves of observed OTUs for each replicate. SSM = Sault-Ste. Marie; MC = Middlesex County; SBLG = Saint-Basile-le-Grand; GAT = Gatineau and MTL = Montreal.



Figure 2S. Relative abundance at the genus level of OTUs identified that significantly influence the separation of samples in two ecotypes. Relative abundance of named genus is significantly different between populations of Ecotype 1 and Ecotype 2 (*t*-Test, p > 0.1). Non-abundant OTUs (<1% of relative abundance) are grouped in the category "others".

# ANNEXE II : PROCÉDURE UPARSE

La procédure suivante permet d'analyser les séquences obtenues par séquençage à haut débit Illumina Miseq Paired-end. Elle permet également de générer des courbes de raréfaction. Cette procédure inclut le protocole UPARSE ainsi que quelques commandes implémentées dans QIIME V. 1.8.0 et des commandes Linus. Une courte explication, identifiée par un #, précède chacune des lignes de commandes. Le script suivant doit être exécuté à l'aide d'un terminal dans le système d'exploitation Linux.

## 

#Créer un dossier où tous les fichiers de sortie seront enregistrés (Uparse\_analyse).

#Ce dossier doit contenir les fichiers suivants :

#Dossier contenant les séquences brutes; les séquences doivent être extraites dans ce fichier (clique de gauche sur la souris, "extract here").

#Le fichier binaire Usearch61 64bits (Usearch64);

#Ce pipeline UPARSE;

#Le fichier gold.fa (base de données pour chimère).

#Les fichiers RDP pour l'assignation taxonomique (rdp\_16s.fa; rdp\_16s.tt; rdp\_16s\_short.tc).

#Ouvrir le dossier d'analyse dans le terminal.

cd ~/Desktop/Amélie/Uparse\_analyse/

#### 

# La commande echo est une commande Linux permettant d'écrire ce qui suit la commande dans le terminal. Lorsqu'un script est lancé, cela permet de savoir quelle étape de la procédure est en cours. # La commande count\_seqs.py est intégrée dans QIIME et permet de déterminer le nombre de séquences totales au départ pour chacun des échantillons.

# Le compte est inscrit dans le fichier count\_seqs.txt. Le signe >> signifie que l'information sera inscrite à la suite du fichier sans effacer l'information déjà présente.

# À cette étape de l'analyse, deux fichiers sont inscrits pour chaque échantillon puisqu'un fichier contient les séquences sens (R1) et un fichier contient les séquences anti-sens (R2).

#### echo Count sequences

#### count\_seqs.py -i

Sequences\_brutes/1e\_R1.fastq,Sequences\_brutes/1e\_R2.fastq,Sequences\_brutes/1i\_R1.fastq, Sequences brutes/1i R2.fastq,Sequences brutes/2e R1.fastq,Sequences brutes/2e R2.fastq, Sequences brutes/2i R1.fastq,Sequences brutes/2i R2.fastq,Sequences brutes/3i R1.fastq,S equences\_brutes/3i\_R2.fastq,Sequences\_brutes/3e\_R1.fastq,Sequences\_brutes/3e\_R2.fastq,S equences brutes/4i R1.fastq,Sequences brutes/4i R2.fastq,Sequences brutes/4e R1.fastq,S equences brutes/4e R2.fastq.Sequences brutes/5i R1.fastq.Sequences brutes/5i R2.fastq.S equences brutes/5e R1.fastq,Sequences brutes/5e R2.fastq,Sequences brutes/6i R1.fastq,S equences\_brutes/6i\_R2.fastq,Sequences\_brutes/6e\_R1.fastq,Sequences\_brutes/6e\_R2.fastq,S equences\_brutes/7i\_R1.fastq,Sequences\_brutes/7i\_R2.fastq,Sequences\_brutes/7e\_R1.fastq,S equences brutes/7e R2.fastq,Sequences brutes/8i R1.fastq,Sequences brutes/8i R2.fastq,S equences brutes/8e R1.fastq,Sequences brutes/8e R2.fastq,Sequences brutes/9i R1.fastq,S equences brutes/9i R2.fastq,Sequences brutes/9e R1.fastq,Sequences brutes/9e R2.fastq,S equences brutes/10i R1.fastq.Sequences brutes/10i R2.fastq.Sequences brutes/10e R1.fast g,Sequences brutes/10e R2.fastg,Sequences brutes/11i R1.fastg,Sequences brutes/11i R2.f astq,Sequences brutes/11e R1.fastq,Sequences brutes/11e R2.fastq,Sequences brutes/12i R1.fastq,Sequences brutes/12i R2.fastq,Sequences brutes/12e R1.fastq,Sequences brutes/1 2e\_R2.fastq,Sequences\_brutes/13i\_R1.fastq,Sequences\_brutes/13i\_R2.fastq,Sequences\_brute s/13e R1.fastq,Sequences brutes/13e R2.fastq,Sequences brutes/14i R1.fastq,Sequences b rutes/14i R2.fastq.Sequences brutes/14e R1.fastq.Sequences brutes/14e R2.fastq.Sequence s brutes/15i R1.fastq,Sequences brutes/15i R2.fastq,Sequences brutes/15e R1.fastq,Seque nces brutes/15e R2.fastq,Sequences brutes/16i R1.fastq,Sequences brutes/16i R2.fastq,Se quences brutes/16e R1.fastq,Sequences brutes/16e R2.fastq,Sequences brutes/17i R1.fastq ,Sequences brutes/17i R2.fastq,Sequences brutes/17e R1.fastq,Sequences brutes/17e R2.f astq,Sequences brutes/18i R1.fastq,Sequences brutes/18i R2.fastq,Sequences brutes/18e R1.fastq,Sequences brutes/18e R2.fastq,Sequences brutes/19i R1.fastq,Sequences brutes/1 9i R2.fastq,Sequences brutes/19e R1.fastq,Sequences brutes/19e R2.fastq,Sequences brut es/20i R1.fastq,Sequences brutes/20i R2.fastq,Sequences brutes/20e R1.fastq,Sequences b rutes/20e R2.fastq,Sequences brutes/21i R1.fastq,Sequences brutes/21i R2.fastq,Sequence s brutes/21e R1.fastq,Sequences brutes/21e R2.fastq,Sequences brutes/22i R1.fastq,Seque nces brutes/22i R2.fastq,Sequences brutes/22e R1.fastq,Sequences brutes/22e R2.fastq,Se quences\_brutes/23i\_R1.fastq,Sequences\_brutes/23i\_R2.fastq,Sequences\_brutes/23e\_R1.fastq, Sequences brutes/23e R2.fastg, Sequences brutes/24i R1.fastg, Sequences brutes/24i R2.fa sta.Sequences brutes/24e R1.fastq,Sequences brutes/24e\_R2.fastq,Sequences\_brutes/25i\_R 1.fastq,Sequences brutes/25i R2.fastq,Sequences brutes/25e R1.fastq,Sequences brutes/25 e R2.fastq,Sequences brutes/26i R1.fastq,Sequences brutes/26i R2.fastq,Sequences brutes /26e R1.fastq,Sequences brutes/26e R2.fastq >> count seqs.txt

# La commande mkdir est une commande Linux permettant de créer un dossier. Le dossier MergeSeqs contiendra les séquences assemblées.

# L'assemblage des séquences se fait avec Usearch64, grâce à la commande -fastq\_mergepairs. Les séquences assemblées sont enregistrées dans le fichier MergeSegs.

echo merge paired-end

mkdir MergeSeqs

./usearch64 -fastq\_mergepairs Sequences\_brutes/1i\_R1.fastq -reverse Sequences\_brutes/1i\_R2.fastq -fastqout MergeSeqs/1i\_merged.fastq -fastq\_minovlen 20

./usearch64 -fastq\_mergepairs Sequences\_brutes/1e\_R1.fastq -reverse Sequences\_brutes/1e\_R2.fastq -fastqout MergeSeqs/1e\_merged.fastq -fastq\_minovlen 20

#Répéter pour tous les échantillons

# L'information est ajoutée à la suite du fichier count\_seqs.txt.

echo Compte des séquences assemblées

#### count\_seqs.py -i

MergeSegs/1e merged.fastq,MergeSegs/1i merged.fastq,MergeSegs/2e merged.fastq,MergeS eqs/2i merged.fastq,MergeSeqs/3i merged.fastq,MergeSeqs/3e merged.fastq,MergeSeqs/4i merged.fastq,MergeSeqs/4e\_merged.fastq,MergeSeqs/5i\_merged.fastq,MergeSeqs/5e\_merged .fastg,MergeSegs/6i merged.fastg,MergeSegs/6e merged.fastg,MergeSegs/7i merged.fastg,M ergeSegs/7e merged.fastg.MergeSegs/8i merged.fastg.MergeSegs/8e merged.fastg.MergeSe qs/9i merged.fastq,MergeSeqs/9e merged.fastq,MergeSeqs/10i merged.fastq,MergeSeqs/10e merged.fastg.MergeSegs/11i merged.fastg.MergeSegs/11e merged.fastg.MergeSegs/12i me rged.fastq,MergeSegs/12e merged.fastq,MergeSegs/13i merged.fastq,MergeSegs/13e merge d.fastq,MergeSegs/14i merged.fastq,MergeSegs/14e merged.fastq,MergeSegs/15i merged.fas tg,MergeSegs/15e merged.fastg,MergeSegs/16i merged.fastg,MergeSegs/16e merged.fastg, MergeSegs/17i merged.fastq,MergeSegs/17e merged.fastq,MergeSegs/18i merged.fastq,Mer geSegs/18e merged.fastq,MergeSegs/19i merged.fastq,MergeSegs/19e merged.fastq,MergeS eqs/20i merged.fastq,MergeSeqs/20e merged.fastq,MergeSeqs/21i merged.fastq,MergeSeqs/ 21e merged.fastq,MergeSeqs/22i\_merged.fastq,MergeSeqs/22e\_merged.fastq,MergeSeqs/23i merged.fastg.MergeSegs/23e merged.fastg.MergeSegs/24i merged.fastg.MergeSegs/24e m erged.fastq,MergeSeqs/25i\_merged.fastq,MergeSeqs/25e\_merged.fastq,MergeSeqs/26i\_merge d.fastq,MergeSegs/26e merged.fastq >> count segs.txt

# Puisqu'il n'y a pas de patron distinctif entre les communautés bactériennes associées aux mâles et aux femelles ainsi qu'entre l'endo et l'ectomicrobiome, les échantillons sont regroupés par répliquas par site d'échantillonnage.

# La commande cat est une commande Linux permettant de jumeler plusieurs fichiers ensemble.

cat MergeSeqs/1e\_merged.fastq MergeSeqs/1i\_merged.fastq MergeSeqs/4i\_merged.fastq MergeSeqs/4e\_merged.fastq > MTL\_1.fastq

cat MergeSeqs/2e\_merged.fastq MergeSeqs/2i\_merged.fastq MergeSeqs/5i\_merged.fastq MergeSeqs/5e\_merged.fastq > MTL\_2.fastq

#Répéter pour tous les échantillons.

#### 

# Le compte des séquences est ajouté à la suite du fichier count\_seqs.txt.

echo Compte des séquences regroupées

count\_seqs.py -i MTL\_1.fastq,MTL\_2.fastq,MTL\_3.fastq,SSM\_1.fastq,SSM\_2.fastq,SSM\_3.fastq,ABCA\_1.fastq,A BCA\_2.fastq,ABCA\_3.fastq,SBLG\_1.fastq,SBLG\_2.fastq,SBLG\_3.fastq,GAT\_1.fastq,GAT\_2.fas tq,GAT\_3.fastq >> count\_seqs.txt

#### 

# Dans ce cas-ci, les séquences assemblées de tous les échantillons sont jumelées dans un seul fichier (merged.fastq), afin d'obtenir de l'information statistique pour l'ensemble des séquences.

# La commande -fastq\_stats implantée dans Usearch génère un fichier de sortie (stats.log) contenant la taille des séquences, le Q score, le Q score en fonction de la taille des séquences, etc. Cela permet de mieux choisir les paramètres pour le contrôle de qualité.

echo Statistiques

cat MTL\_1.fastq,

MTL\_2.fastq,MTL\_3.fastq,SSM\_1.fastq,SSM\_2.fastq,SSM\_3.fastq,ABCA\_1.fastq,ABCA\_2.fastq,ABCA\_3.fastq,SBLG\_1.fastq,SBLG\_2.fastq,SBLG\_3.fastq,GAT\_1.fastq,GAT\_2.fastq,GAT\_3.fastq > totalseq.fastq

./usearch64 -fastq\_stats totalseq.fastq -log stats.log

# Les séquences qui ont passé le contrôle de qualité sont enregistrées dans un nouveau dossier (filter).

# La commande –fastq\_filter implantée dans Usearch filtre les séquences selon les paramètres indiqués. Dans ce cas-ci, le seul paramètre indiqué est le –fastq\_maxee de 1. Donc, toutes les séquences ayant un « expected error » plus grand que 1 sont éliminées.

echo Quality control

mkdir filter

./usearch64 -fastq\_filter MTL\_1.fastq -fastaout MTL\_1\_filtered.fa -fastq\_maxee 1

./usearch64 -fastq\_filter MTL\_2.fastq -fastaout MTL\_2\_filtered.fa -fastq\_maxee 1

#Répéter pour tous les échantillons.

#### echo compte seqs CQ

count\_seqs.py -i

MTL\_1\_filtered.fa,MTL\_2\_filtered.fa,MTL\_3\_filtered.fa,SSM\_1\_filtered.fa,SSM\_2\_filtered.fa,SS M\_3\_filtered.fa,ABCA\_1\_filtered.fa,ABCA\_2\_filtered.fa,ABCA\_3\_filtered.fa,SBLG\_1\_filtered.fa,S BLG\_2\_filtered.fa,SBLG\_3\_filtered.fa,GAT\_1\_filtered.fa,GAT\_2\_filtered.fa,GAT\_3\_filtered.fa>> count\_seqs.txt

#### 

# La commande sed est une commande Linux permettant de faire des modifications à l'intérieur d'un fichier.

# Puisque les séquences obtenues du séquençage ont déjà été démultiplexées et le barcode retiré, une identification est ajoutée à chacune des séquences afin de retracer à quel échantillon elles appartiennent.

# Dans ce cas-ci, le nom de l'échantillon est ajouté devant chacune des séquences et un nouveau fichier est généré.

echo Demultiplex illumina reads

#### mkdir barcodelabel

sed "-es/^>\(.\*\)/>\1; barcodelabel=MTL\_1; /" < MTL\_1\_filtered.fa > barcodelabel/MTL\_1.fa sed "-es/^>\(.\*\)/>\1; barcodelabel=MTL\_2; /" < MTL\_2\_filtered.fa > barcodelabel/MTL\_2.fa #Répéter pour tous les échantillons.

#### 

# Les séquences de chacun des échantillons sont jumelées dans un seul fichier (filter.fa) afin de générer la table des OTUs.

echo jumelage sequences

cat barcodelabel/MTL\_1.fa barcodelabel/MTL\_2.fa barcodelabel/MTL\_3.fa barcodelabel/SSM\_1.fa barcodelabel/SSM\_2.fa barcodelabel/SSM\_3.fa barcodelabel/ABCA\_1.fa barcodelabel/ABCA\_2.fa barcodelabel/ABCA\_3.fa barcodelabel/SBLG\_1.fa barcodelabel/SBLG\_2.fa barcodelabel/SBLG\_3.fa barcodelabel/GAT\_1.fa barcodelabel/GAT\_2.fa barcodelabel/GAT\_3.fa > filter.fa

#### 

# La commande –derep\_fullenght implantée dans Usearch génère un fichier contenant des séquences uniques ainsi qu'une annotation indiquant le nombre de fois que la séquence est retrouvée dans le fichier d'entrée.

# Les séquences doivent être identiques sur toute la longueur.

echo Dereplication

./usearch64 -derep\_fullength filter.fa -fastaout derep.fasta -sizeout

#### 

# La commande –sortbysize implantée dans Usearch permet de trier les séquences selon leur abondance et de retirer les singletons grâce à l'argument –minsize 2.

echo Filtre abondance et singleton

./usearch64 -sortbysize derep.fasta -fastaout seqs\_sorted\_size.fasta -minsize 2

# La commande –cluster\_otus implantée dans Usearch permet de faire le regroupement des séquences avec l'algorithme UPARSE-OTU.

# Par défault, le regroupement est effectué avec un pourcentage de similitude de 97%.

# L'argument –otus spécifie le fichier de sortie où les séquences représentatives seront inscrites.

# L'argument – relabel indique le préfixe pour chacun des OTU suivis de numéro (OTU\_1, OTU\_2, OTU\_3, etc.)

# Les arguments –sizein et –sizeout permet de prendre en compte le nombre de copies de chacune des séquences et les inscrire dans le fichier de sortie.

# L'argument –uparseout spécifie un fichier texte documentant comment les séquences d'entrée ont été classifiées.

#### echo Cluster OTU

./usearch64 -cluster\_otus seqs\_sorted\_size.fasta -otus otus.fasta -uparseout uparseout.up - relabel OTU\_ -sizein -sizeout

#### 

# La commande –uchime\_ref implantée dans Usearch permet de retirer les chimères avec une base de données de références (gold.fa).

# L'argument –strand plus signifie que seulement le brin sens sera considéré (même direction que la base de données).

# L'argument –nonchimeras spécifie le fichier de sortie contenant les séquences non chimériques. Par contre, lorsque cet argument est utilisé, un message d'erreur apparait. Pour résoudre ce problème, il faut utiliser l'argument –uchimeout. Dans ce fichier de sortie, les séquences sont identifiées chimériques ou non chimériques, mais ne sont pas retirées.

# Les commandes grep et cut (commandes de Linux) sont utilisées afin de générer un fichier texte (labels.txt) contenant seulement les séquences non chimériques.

# La commande –fastx\_getseqs implantée dans Usearch permet de générer un fichier fasta dont seulement les séquences retrouvées dans le fichier labels sont copiées.

echo chimeres

./usearch64 -uchime\_ref otus.fasta -db gold.fa -uchimeout otus1.fa -strand plus

grep "N\$" otus1.fa | cut -f2 > labels.txt

./usearch64 -fastx\_getseqs otus.fasta -labels labels.txt -fastaout otus2.fa

#### 

#Afin de retirer les OTU représentant des archaea, il faut d'abord effectuer une assignation taxonomique avec la base de données RDP des séquences du fichier fasta où les chimères ont été retirées (otus2.fa).

# Par la suite, un fichier contenant le numéro des OTU identifié comme étant des archaea est généré.

#Ouvrir le fichier otus2.fa et enregistrer de nouveau sous le nom otus3.fa.

# Retirer manuellement les otu identifiés comme étant archaea dans le fichier otus3.fa.

./usearch64 -utax otus2.fa -db rdp\_16s.fa -taxconfs rdp\_16s\_short.tc -utaxout tax.txt -tt rdp\_16s.tt

grep -i "Archaea" tax.txt > archaea.txt

#### 

# Les séquences doivent être associées aux OTU précédemment déterminés (séquences représentatives).

# Le fichier d'entrée (filter.fa) contient toutes les séquences avant le retrait des chimères et des singletons.

# La table d'OTU de référence est la dernière table d'OTU générée (chimères retirées).

# Les séquences sont associées à 97% avec les séquences représentatives pour produire une « map reads » (readmap.uc).

echo map reads

./usearch64 -usearch\_global filter.fa -db otus3.fa -strand plus -id 0.97 -uc readmap.uc

# Cette commande permet de savoir combien de séquences n'ont pas été associées aux séquences représentatives. Certaines séquences ne sont pas associées aux séquences représentatives (à 97% de similitude) puisqu'elles sont soit chimériques ou singletons.

grep -c "^N" readmap.uc

#### 

# Cette commande permet de convertir la map .uc (mapping reads) en table des OTU.

# Un script python est utilisé. Celui-ci est situé dans le dossier usr/bin.

echo table des OTU

python /usr/bin/drive5\_py/uc2otutab.py readmap.uc > tabfile.txt

#### 

# La standardisation consiste en le retrait des OTU représentés à moins de 0,005% et en l'égalisation du nombre de séquences par échantillon.

# L'outil biom convert permet de transformer un fichier texte en format biom.

# La commande filter\_otus\_from\_otu\_table.py est intégrée dans QIIME. Cette commande permet de retirer toutes les séquences représentées à moins de 0,005%.

# Afin de déterminer le nombre de séquences par échantillon, un fichier sommaire est généré avec la commande biom summary\_table. Dans ce fichier, le nombre de séquences minimales dans un échantillon est inscrit (ex. 243 859). Le nombre de séquences par échantillon doit être égalisé à cette valeur minimale pour tous les échantillons.

# Le fichier sommaire est généré une deuxième fois, après l'égalisation, afin de confirmer que le nombre de séquences par échantillon est égal pour tous.

# La table des OTU en format biom doit ensuite être reconvertie en fichier texte afin de pouvoir l'ouvrir.

biom convert -i tabfile.txt -o tabfile.biom --table-type="OTU table" --to-hdf5

filter\_otus\_from\_otu\_table.py -i tabfile.biom -o Otu\_table\_filtered.biom --min\_count\_fraction 0.00005

biom summarize\_table -i Otu\_table\_filtered.biom -o Otu\_table\_filtered\_summary.biom

single\_rarefaction.py -i Otu\_table\_filtered.biom -o Otu\_table\_filtered\_egaliser.biom -d 243859

biom summarize\_table -i Otu\_table\_filtered\_egaliser.biom -o Otu\_table\_filtered\_egaliser\_summary.biom

biom convert -i Otu\_table\_filtered\_egaliser.biom -o Otu\_table\_filtered\_egaliser.txt --to-tsv

#### 

# L'assignation taxonomique est effectuée sur les séquences du fichier fasta. Par contre, les séquences qui ont été retirées de la table des OTUs à l'étape de la standardisation n'ont pas été retirées dans le fichier fasta.

#Pour ce faire, la commande filter\_fasta.py intégrée dans QIIME peut être exécutée. Le fichier d'entrée est le fichier fasta contenant les séquences avant standardisation, le fichier de sortie est le fichier fasta contenant les séquences après la standardisation et la table des OTU est également inscrite afin de préciser quelles séquences doivent être retenues.

filter\_fasta.py -f otu3.fa -o fastafinal.fa -b Otu\_table\_filtered\_egaliser.biom

#### 

# L'assignation taxonomique est effectuée avec la base de données RDP.

./usearch64 -utax fastafinal.fa -db rdp\_16s.fa -taxconfs rdp\_16s\_short.tc -utaxout taxfinal.txt -tt rdp\_16s.tt

#### 

# Le fichier final pour la taxonomie est tax.final.txt

# Le fichier final pour la table des OTU est Otu\_table\_filtered\_egaliser.txt

# Afin de générer les courbes de raréfaction, il faut d'abord aligner les séquences (<u>http:</u>//giime.org/scripts/align\_seqs.html; http://giime.org/scripts/make\_phylogeny.html)

# Les courbes de raréfaction sont générées avec QIIME grâce à la commande core\_diversity\_analyses.py. -i est la table des OTU finale, -o est le fichier de sortie (celui-ci sera créé en lançant la commande), -e est le nombre de séquence par échantillon (59725 séquences), -m est le mapping file, (celui-ci est généré par nous même). Voir la documentation dans le cartable Projet Agrile- Analyses bio-informatiques ou sur le site internet de QIIME. La commande Une commande validate\_mapping\_file.py permet de valider le mapping file créé (<u>http:</u>//giime.org/scripts/core\_diversity\_analyses.html).

align\_seqs.py -i fastafinal.fa

make\_phylogeny.py -i pynast\_aligned/fastafinal\_aligned.fasta

validate\_mapping\_file.py -m MappingFile.txt -b -p -j Description

core\_diversity\_analyses.py -i Otu\_table\_filtered\_egaliser.biom -o alphadiversity/ -m MappingFile.txt -e 59725 -t pynast\_aligned/fastafinal\_aligned.tre --suppress\_taxa\_summary

# ANNEXE III : CRÉATION D'UN FICHIER CSV

Afin d'effectuer les analyses de diversité avec le logiciel R, la table des OTU doit être en format csv. Pour ce faire, les étapes suivantes doivent être effectuées :

- Récupérer la table des OTU nommée « Otu\_table\_filtered\_egaliser.txt » avec le logiciel Excel de la suite Microsoft Office;
- 2. Apporter les modifications suivantes au fichier :
  - a. Les OTU peuvent être renuméroté si désiré;
  - b. L'élément « ; size=XXX » est retiré;
  - c. Les points sont remplacés par des virgules;
  - d. #Constructed from biom table est retiré;
  - e. OTU ID est remplacé par « group »
  - f. Le nom des échantillons correspond aux lignes alors que les OTU correspondent aux colonnes;
- 3. Enregistrer le fichier en csv (séparateur : point virgule);
- 4. Ouvrir le fichier avec le logiciel WordPad et remplacer les points virgules par des virgules;
- 5. Le fichier csv est prêt à être utilisé avec le logiciel R pour les analyses de diversité.

# ANNEXE IV : SCRIPT R

Les scripts qui suivent sont utilisés pour l'exécution des analyses de diversité suivantes : estimateur ACE, analyse de dispersion multivariée (β-diversité), dendogramme de similarité, analyse en composante principale, espèces indicatrices (carte de densité). La table des OTU utilisée pour faire ces analyses est le fichier csv présenté à l'annexe 3.

#### 

#This code is for calculation of ACE index# Download librarieslibrary(fossil)library(vegan)

# Specification of file location
setwd("C : \\RBook")

# Import the data from CSV files env <- read.csv("OTU\_nongroupe.csv", row.names=1)</pre>

# Select the wanted column for the analysis # Calculation ACE index (species richness estimator) - sample 1 ace <- ACE(env\$SSM\_1, taxa.row = TRUE) ace

# Répéter pour tous les échantillons

#This code is for β-diversity analysis# Download librarieslibrary(vegan)

# Specification of file location
setwd("C : \\RBook")

# Import the data from CSV files

env <- read.csv("OTU\_nongroupe\_inverser.csv", row.names=1)</pre>

# Emphasis on the distribution of the proportion of each species within replicated experimental plots and between treatments

# Hellinger transformation of dataset (owing to several "zero" values)

# In this case, we will compare the proportion of OTUs in the samples.

spe.dbln <- decostand(env, "hel")

# We calculate Euclidean distance between samples on hellinger-transformed dataset.

dis <- dist(spe.dbln, "euc")

# Definition of the groups

```
groups <- factor(c(rep(1,3), rep(2,3), rep(3,3), rep(4,3), rep(5,3)), labels = c("SSM","MC","SBLG","GAT","MTL"))
```

# Calculate multivariate dispersions

```
mod <- betadisper(dis, groups, type = "centroid")
.</pre>
```

mod

## S3 method for class 'betadisper' :
boxplot(mod, ylab = "Distance to centroid")
## S3 method for class 'betadisper' :
scores(mod, display = c("sites", "centroids"), choices = c(1,2))

## Perform test
anova(mod)
## Permutation test for F
permutest(mod, pairwise = TRUE)

# Download libraries
library(ade4)
library (vegan) # should be loaded after ade4 to avoid some conflicts !
library (gclus)
library(cluster)
library(RColorBrewer)
library(labdsv)
library(ape)
library(clustsig)
# Specification of the location of the data files
setwd("C : \\RBook\\")

# Import the data from CSV files

env <- read.csv("OTU\_nongroupe\_inverser.csv", row.names=1)</pre>

# STEP 1 : We need to identify the best model fitting to our dataset.

# Ribotyping data : compute chord distance matrix of standardized relative abundance of ribotypes (the same weight is given to each ribotype; the argument "nor" normalize site vector to 1 followed by single linkage agglomerative clustering.

# We first standardize the dataset (average for each variable becomes 0 and the variance becomes 1.

spe.dbln <- decostand(env, "hellinger")</pre>

# We then calculate a distance matrix for the dataset and use the function hclust to calculate a dendrogram using the generated matix.

spe.ch <- vegdist(spe.dbln, "euc")</pre>

spe.ch.single <- hclust(spe.ch, method="single")</pre>

# Plot a dendrogram using the default options
windows(title="Biolog - Standardized - Single linkage",12,8)
plot(spe.ch.single)

# Compute complete-linkage agglomerative clustering spe.ch.complete <- hclust(spe.ch, method="complete") windows(title="Biolog - Standardized - Complete linkage",12,8) plot(spe.ch.complete)

# Compute UPGMA agglomerative clustering spe.ch.UPGMA <- hclust(spe.ch, method="average") windows(title="Biolog - Standardized - UPGMA",12,8) plot(spe.ch.UPGMA)

# Compute Ward's minimum variance clustering spe.ch.ward <- hclust(spe.ch, method="ward") windows(title="Biolog - Standardized - Ward",12,8) plot(spe.ch.ward)

# The next command will produce data more comparable to the other model (the scale of the distances is adjusted by the square root).

windows(title="Biolog - Standardized - sqrt(Ward)",12,8)
spe.ch.ward\$height <- sqrt(spe.ch.ward\$height)
plot(spe.ch.ward)</pre>

# Cophenetic correlation. This will allow us to determine which model is the best in explaining our dataset (the ramifications represented by the models are compared to the real distance among the different samples). Correlations are then calculated between modelled and real distances. The higher is the correlation score, the better is the model(THIS IS AN ASSUMPTION as it is not possible to calculate a significance level to the correlation scores). In the utilized in the following command, Pearson correlations are specified since the argument "Method" of the command "cor" is not specified. This imply that normal distribution is assumed for the individual variables.

# Single linkage clustering
spe.ch.single.coph <- cophenetic(spe.ch.single)
cor(spe.ch, spe.ch.single.coph)</pre>

# Complete linkage clustering
spe.ch.comp.coph <- cophenetic(spe.ch.complete)
cor(spe.ch, spe.ch.comp.coph)</pre>

# Average clustering
spe.ch.UPGMA.coph <- cophenetic(spe.ch.UPGMA)
cor(spe.ch, spe.ch.UPGMA.coph)</pre>

# Ward clustering
spe.ch.ward.coph <- cophenetic(spe.ch.ward)
cor(spe.ch, spe.ch.ward.coph)</pre>

# These correlations could also being computed using Spearman correlation. This method does not assume normal distribution of the variables.

cor(spe.ch, spe.ch.single.coph, method="spearman")
cor(spe.ch, spe.ch.comp.coph, method="spearman")
cor(spe.ch, spe.ch.UPGMA.coph, method="spearman")

cor(spe.ch, spe.ch.ward.coph, method="spearman")

# Shepard-like diagrams. This will give us an overview of the calculated correlations.

windows(title="Cophenetic correlation",8,9)

par(mfrow=c(2,2))

```
plot(spe.ch, spe.ch.single.coph, xlab="Standardized distance",
```

ylab="Cophenetic distance", asp=1, xlim=c(0,12), ylim=c(0,12),

main=c("Single linkage",paste("Cophenetic correlation ",

```
round(cor(spe.ch, spe.ch.single.coph),3))))
```

```
abline(0,1)
```

```
lines(lowess(spe.ch, spe.ch.single.coph), col="red")
```

plot(spe.ch, spe.ch.comp.coph, xlab="Standardized distance",

```
ylab="Cophenetic distance", asp=1, xlim=c(0,12), ylim=c(0,12),
```

```
main=c("Complete linkage", paste("Cophenetic correlation ",
```

```
round(cor(spe.ch, spe.ch.comp.coph),3))))
```

```
abline(0,1)
```

```
lines(lowess(spe.ch, spe.ch.comp.coph), col="red")
```

plot(spe.ch, spe.ch.UPGMA.coph, xlab="Standardized distance",

```
ylab="Cophenetic distance", asp=1, xlim=c(0,12), ylim=c(0,12),
```

```
main=c("UPGMA", paste("Cophenetic correlation ",
```

```
round(cor(spe.ch, spe.ch.UPGMA.coph),3))))
```

abline(0,1)

```
lines(lowess(spe.ch, spe.ch.UPGMA.coph), col="red")
```

plot(spe.ch, spe.ch.ward.coph, xlab="Standardized distance",

```
ylab="Cophenetic distance", asp=1, xlim=c(0,12),
```

```
ylim=c(0,max(spe.ch.ward$height)),
```

main=c("Ward clustering", paste("Cophenetic correlation ", round(cor(spe.ch, spe.ch.ward.coph),3)))) abline(0,1) lines(lowess(spe.ch, spe.ch.ward.coph), col="red")

# STEP 2 : We need to identify how many clusters we should define in the selected dendrogram.

# Graphs of fusion level values. Fusion level value represents the distance between a pair of samples bridged by a node. The dendrogram possess a maximum of N cluster where N represent the number of sample. The graph of fusion level will show the weight of each cluster in term of represented distance. At some point, addition of new cluster will not bring ecologically-relevant differences among site (topology impossible to interprete).

windows(title="Fusion levels",12,8)

par(mfrow=c(2,2))

# Plot the fusion level values of the single linkage clustering summary(spe.ch.single) # List of available results

plot(spe.ch.single\$height, nrow(env) : 2, type="S",

main="Fusion levels - Standardized - Single",

ylab="k (number of clusters)", xlab="h (node height)", col="grey")

text(spe.ch.single\$height, nrow(env) : 2, nrow(env) : 2, col="red", cex=0.8)

# Plot the fusion level values of the complete linkage clustering

plot(spe.ch.complete\$height, nrow(env): 2, type="S",

main="Fusion levels - Standardized - Complete",

ylab="k (number of clusters)", xlab="h (node height)", col="grey")

```
text(spe.ch.complete$height, nrow(env) : 2, nrow(env) : 2, col="red", cex=0.8)
```

```
# Plot the fusion level values of the UPGMA clustering
plot(spe.ch.UPGMA$height, nrow(env) : 2, type="S",
    main="Fusion levels - Standardized - UPGMA",
    ylab="k (number of clusters)", xlab="h (node height)", col="grey")
text(spe.ch.UPGMA$height, nrow(env) : 2, nrow(env) : 2, col="red", cex=0.8)
```

# Plot the fusion level values of the Ward clustering plot(spe.ch.ward\$height, nrow(env) : 2, type="S",

```
main="Fusion levels - Standardized - Ward",
ylab="k (number of clusters)", xlab="h (node height)", col="grey")
text(spe.ch.ward$height, nrow(env) : 2, nrow(env) : 2, col="red", cex=0.8)
```

```
# Optimal number of clusters according to Mantel statistic (Pearson)
# Function to compute a binary distance matrix from groups
grpdist <- function(X)
{
require(cluster)
gr <- as.data.frame(as.factor(X))
distgr <- daisy(gr, "gower")
distar
}
# Run based on the Single clustering
kt <- data.frame(k=1 : nrow(env), r=0)</pre>
for (i in 2 : (nrow(env)-1)) {
       gr <- cutree(spe.ch.UPGMA, i)
       distgr <- grpdist(gr)
       mt <- cor(spe.ch, distgr, method="pearson")
       kt[i,2] <- mt
}
kt
k.best <- which.max(kt$r)
# The plot is produced by function plot.silhouette cluster
windows(title="Optimal number of clusters - Mantel")
plot(kt$k, kt$r, type="h", main="Mantel-optimal number of clusters - UPGMA",
       xlab="k (number of groups)", ylab="Pearson's correlation")
axis(1, k.best, paste("optimum", k.best, sep="\n"), col="red", font=2,
       col.axis="red")
points(k.best, max(kt$r), pch=16, col="red", cex=1.5)
```

# Silhouette plot of the final partition. This will help us to figure out if grouped samples really belong to the identified clusters (they must be all positive or negative in the cluster - samples with different

magnitude are considered not so well classified. Again, this is not a statistical test (no significance level is calculated).

# Choose the number of clusters (we select according to the Mantel statistic AND OUR OWN INTERPRETATION OF THE CLUSTERS).

k <- 6

# Silhouette plot

cutg <- cutree(spe.ch.UPGMA, k=k)

sil <- silhouette(cutg, spe.ch)</pre>

silo <- sortSilhouette(sil)</pre>

```
rownames(silo) <- row.names(env)[attr(silo,"iOrd")]
```

windows(title="Silhouette plot - UPGMA - K=6")

plot(silo, main="Silhouette plot - Srandardized - UPGMA",

cex.names=0.8, col=cutg+1, nmax.lab=100)

# STEP 3 : We combine clustering and ordination results.

#Cut the dendrogram to yield 6 groups

gr <- cutree(spe.ch.UPGMA, k=6)

grl <- levels(factor(gr))

```
spe.chwo <- reorder.hclust(spe.ch.UPGMA, spe.ch)</pre>
```

```
plot(spe.chwo, hang=-1, xlab="4 groups", sub="", ylab="Sites", main="UPGMA", labels=cutree(spe.chwo, k=6))
```

rect.hclust(spe.chwo, k=6)

source("hcoplot.R")

hcoplot(spe.ch.UPGMA, spe.ch, k=6)

# STEP 4 : Number of significant clusters

sim <- simprof(data=spe.dbln, num.expected=1000, num.simulated=999, method.cluster="average", method.distance="euclidean",

method.transform="identity", alpha=0.05, sample.orientation="row", const=0, silent=TRUE, increment=100, undef.zero=TRUE, warn.braycurtis=TRUE)

sim

# Additional functions (Script R trouvé sur Internet, doivent être enregistré dans le RBook) source("evplot.R") source("hcoplot.R")

# Specification of file location
setwd("C : \\RBook")

# Import the data from CSV files env <- read.csv("OTU\_nongroupe\_inverser.csv", row.names=1)</pre>

# A reminder of the content of the env dataset summary(env)# Descriptive statistics

# PCA on the full dataset (correlation matrix : scale=TRUE)
spe.h <- decostand(env, "hellinger")
env.pca <- rda(spe.h)</pre>

env.pca summary(env.pca) # Default scaling 2 summary(env.pca, scaling=1)

# Examine and plot partial results from PCA output
# Eigenvalues
(ev <- env.pca\$CA\$eig)</pre>

# Apply Kaiser-Guttman criterion to select axes
ev[ev > mean(ev)]

```
# Plot eigenvalues and % of variance for each axis
source("evplot.R")
evplot(ev)
```

```
# Two PCA biplots : scaling 1 and scaling 2
# Plots using biplot.rda
windows(title="PCA biplots - environment - biplot.rda", 12, 6)
par(mfrow=c(1,2))
biplot(env.pca, scaling=1, main="PCA - scaling 1")
biplot(env.pca, main="PCA - scaling 2") # Default scaling = 2
```

```
# Plots using cleanplot.pca
# A rectangular graphic window is needed for the two plots
windows(title="PCA biplots - environment - cleanplot.pca", 12, 6)
source("cleanplot.pca.R")
cleanplot.pca(env.pca, point=TRUE) # with points for sites and arrowheads
cleanplot.pca(env.pca)  # with site labels only (vegan's standard)
cleanplot.pca(env.pca, ahead=0) # ... and without arrowheads
```

# This code is utilized for Indicator species analysis# Load required librarieslibrary(indicspecies)

# Specification of file location
setwd("C : \\RBook")

# Import the data from CSV files
spe <- read.csv("OTU\_redefinitiongroupe.csv", row.names=1)</pre>

# Defining the classification of samples (ici, selon la séparation du dendrogramme)
groups <- c(rep(1,8), rep(2,7))
groups</pre>

indval <- multipatt(spe, groups, control = how(nperm=999))
summary(indval)</pre>

spe.pa <- as.data.frame(ifelse(spe>0,1,0))
phi <- multipatt(spe.pa, groups, func="r.g", control=how(nperm=999))</pre>

#This code is for heatmap# Download librarieslibrary(gplots)library(RColorBrewer)

# Specification of file location
setwd("C : \\RBook")

# A) Reading in data and transform it into matrix format
# La table utilisée doit contenir seulement les OTU que l'on veut voir sur la carte.
data <- read.csv("OTU\_especesindicatrices.csv", comment.char="#")</li>
rnames <- data[,1] # assign labels in column 1 to "rnames"</li>
mat\_data <- data.matrix(data[,2 : ncol(data)]) # transform column 2-5 into a matrix</li>
rownames(mat data) <- rnames # assign row names</li>

# B) Customizing and plotting the heat map
# creates a own color palette from red to green
my\_palette <- colorRampPalette(c("blue", "yellow", "orange", "purple", "red"))(n = 499)</li>

```
# (optional) defines the color breaks manually for a "skewed" color transition
col_breaks = c(seq(0,100,length=100), # for blue
seq(100,1000,length=100), # for yellow
seq(1000,5000,length=100), # for yellow
seq(5000,10000,length=100), # for yellow
seq(10000,16000,length=100)) # for red
```

heatmap.2(as.matrix(mat\_data),Rowv = NA,
# block sepration
sepcolor="white",
sepwidth=c(0.05,0.05),
main = "Correlation", # heat map title
notecol="black", # change font color of cell labels to black
density.info="none", # turns off density plot inside color legend
trace="none", # turns off trace lines inside the heat map

margins =c(12,9), # widens margins around plot col=my\_palette, # use on color palette defined earlier breaks=col\_breaks, # enable color transition at specified limits Colv="NA") # turn off column clustering

## ANNEXE V :

DENDROGRAMME DE SIMILARITÉ REPRÉSENTANT LES ÉCHANTILLONS MÂLES ET FEMELLES.

